Registries for Evaluating Patient Outcomes:

A User's Guide

Fourth Edition

Registries for Evaluating Patient Outcomes: A User's Guide

Fourth Edition

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 290-2014-00004-C

Senior Editors:
Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1

Michelle B. Leavy, M.P.H.
Head, Healthcare Research & Policy
OM1

Nancy A. Dreyer, Ph.D., M.P.H., FISPE
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions

AHRQ Publication No. 19(20)-EHC020
September 2020

The Effective Health Care Program of the Agency for Healthcare Research and Quality (AHRQ) conducts and supports research focused on the outcomes, effectiveness, comparative clinical effectiveness, and appropriateness of pharmaceuticals, devices, and healthcare services. More information on the Effective Health Care Program can be found at www.effectivehealthcare.ahrq.gov. This report was produced under contract to the Agency for Healthcare Research and Quality, Rockville, MD by L&M Policy Research, LLC, and partners OM1 and IQVIA under Contract No. 290-2014-00004-C. The AHRQ Task Order Officer for this project was Elise Berliner, Ph.D.

The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services. All statements in this report, including its findings and conclusions, are solely those of the authors. The authors participated as individuals and not necessarily as representatives of their organizations. No statement in this document should be construed as an official position of any organization represented here.

Copyright Information:
©2020 United States Government, as represented by the Secretary of the Department of Health and Human Services, by assignment. All rights reserved. The Agency for Healthcare Research and Quality (AHRQ) permits members of the public to reproduce, redistribute, publicly display, and incorporate this work into other materials provided that it must be reproduced without any changes to the work or portions thereof, except as permitted as fair use under the U.S. Copyright Act. This work contains certain tables and figures noted herein that are subject to copyright by third parties. These tables and figures may not be reproduced, redistributed, or incorporated into other materials independent of this work without permission of the third-party copyright owner(s). This work may not be reproduced, reprinted, or redistributed for a fee, nor may the work be sold for profit or incorporated into a profit-making venture without the express written consent of AHRQ. This work is subject to the restrictions of Section 1140 of the Social Security Act, 42 U.S.C. § 1320b-10. When parts of this work are used or quoted, the following citation should be used:

Citation: Gliklich RE, Leavy MB, Dreyer NA (sr eds). Registries for Evaluating Patient Outcomes: A User's Guide. 4th ed. (Prepared by L&M Policy Research, LLC, under Contract No. 290-2014-00004-C with partners OM1 and IQVIA) AHRQ Publication No. 19(20)-EHC020. Rockville, MD: Agency for Healthcare Research and Quality; September 2020. Posted final reports are located on the Effective Health Care Program search page. DOI: https://doi.org/10.23970/AHRQEPCREGISTRIES4.

Acknowledgments: Numerous individuals contributed to the coordination and formatting of this document. In addition to the authors and reviewers, we wish to acknowledge the following individuals for their significant time and dedication: Devyn Andrews of OM1, Monica Sarmiento of L&M Policy Research, and Elise Berliner and Chris Heidenrich of the Agency for Healthcare Research and Quality.

Preface

This project was performed under a contract from the Agency for Healthcare Research and Quality (AHRQ) with the purpose of updating *Registries for Evaluating Patient Outcomes: A User's Guide*. First published in 2007, the User's Guide, with translations available in Chinese and Korean, serves as a reference for planning, developing, maintaining, and evaluating registries designed to collect data about patient outcomes. The second (2010) and third (2014) editions incorporated updates to existing topics and included new chapters on methodological and technological advances in registry science. The first addendum to the Third Edition, *21st Century Patient Registries*, was published in March 2018 and addressed approaches to engaging patient partners in registry design and operations. The second addendum to the Third Edition, *Tools and Technologies for Registry Interoperability*, was published in December 2019 and addressed concepts related to incorporating other data sources into patient registries.

The purpose of this revised fourth edition is to incorporate information on new methodological and technological advances into the existing chapters and to consolidate and organize the content into a format that emphasizes the key principles of registry design, operations, and analysis. Two new chapters were added (Registry Governance and Selecting and Defining Outcome Measures for Registries), and existing chapters were updated as part of this effort. Like the prior editions of the User's Guide, this fourth edition was created with support from a large group of stakeholders representing academia, industry, government, patient organizations, and technology organizations. At the outset, we solicited feedback on chapter topics and outlines from AHRQ, academics, and other experts in the field. We then reached out to topic experts inviting participation in writing or reviewing the final topics selected. Once the authorship groups were established, many meetings were held to update the chapters prior to sending them for constructive feedback and editorial review to the assigned reviewer group for each paper. The collaborative efforts of contributors, reviewers, and editors resulted in a draft document that was posted for public comment on the Effective Health Care website in April 2019. This document incorporates much of the feedback received. Like previous editions, the contributors and reviewers participated as individual experts and not necessarily as representatives of their organizations. We are grateful to all those who contributed in writing, reviewing and editing this document.

To begin the discussion of registries, we would like to clarify some distinctions between registries and clinical trials. Although this subject is discussed further in Chapter 1, we offer here the following distinctions from a high-level perspective. A clinical trial is an experiment in which an active intervention intended to change a human subject's outcome is implemented, generally through a randomization procedure that takes decision making away from the practitioner. The research protocol describes inclusion and exclusion criteria that are used to select the patients who will participate as human subjects, focusing the experiment on a homogeneous group. Human subjects and clinical researchers agree to adhere to a strict schedule of visits and to conduct protocol-specific tests and measurements.

In contrast, registries use an observational study design that does not specify treatments or require any therapies intended to change patient outcomes (except insofar as specific treatments or therapies may be inclusion criteria). Inclusion and exclusion criteria are kept to a minimum in an effort to study a broad range of patients in order to make the results more generalizable.

Patients are typically observed as they present for care, and the data collected generally reflect whatever tests and measurements a provider customarily uses.

Patient registries represent a useful tool for a number of purposes. Their ideal use and their role in evidence development, design, operations, and evaluation resemble but differ from clinical trials in a number of substantive ways, and therefore they should not be evaluated with the same constructs. This user's guide presents what the contributors and reviewers consider good registry practices. Many registries today may not meet even the basic practices described. On the whole, registry science is in an active state of development. This fourth edition of the user's guide is an important step in developing the field.

This book is divided into four sections: Creating Registries; Legal and Ethical Considerations for Registries; Operating Registries; and Evaluating Registries. The first three sections provide basic information on key areas of registry development and operations, highlighting the spectrum of practices in each of these areas and their potential strengths and weaknesses.

Section I, "Creating Registries," contains six chapters. "Patient Registries" defines and characterizes types of registries, their purposes, and uses, and describes their place within the scope of this document. "Planning a Registry" focuses on the recommended steps in planning a registry, from determining if a registry is the right option to describing goals and objectives. "Registry Design" examines the specifics of designing a registry once the goals and objectives are known. "Selecting and Defining Outcome Measures for Registries" describes considerations related to identifying the most relevant outcomes to measure and selecting appropriate definitions for these outcomes. "Data Elements for Registries" provides a scientific and practical approach to selecting data elements. "Data Sources for Registries" describes how existing data sources (administrative, pharmacy, other registries, etc.) may be used to enhance the value of patient registries.

Section II, "Legal and Ethical Considerations for Registries," contains three chapters. "Principles of Registry Ethics, Data Ownership, and Privacy" reviews several key legal and ethical issues that should be considered in creating or operating a registry. "Informed Consent for Registries" discusses how the requirements of informed consent for patient registries differ from those for clinical trials and offers suggestions for creating informed consent documents that address the unique aspects of registries. "Registry Governance" describes considerations that should inform plans for registry governance and provides examples of potential governance structures.

Section III, "Operating Registries," provides a practical guide to the day-to-day operational issues and decisions for producing and interpreting high-quality registries. "Recruiting and Retaining Participants in the Registry" describes strategies for recruiting and retaining providers and patients. "Obtaining Data and Quality Assurance" reviews key areas related to obtaining, cleaning, and storing data and quality assurance for registries. "Adverse Event Detection, Processing, and Reporting" examines relevant practical and regulatory issues. "Analysis, Interpretation, and Reporting of Registry Data To Evaluate Outcomes" addresses key considerations in analyzing and interpreting registry data.

Section VI is "Evaluating Registries." This final chapter on "Assessing Quality" summarizes key points from the earlier chapters in a manner that can be used to review the structure, data, or

interpretations of patient registries. It describes good registry practice in terms of “essential elements” and “further indicators of quality.” This information might be used by a person developing a registry, or by a reviewer or user of registry data or interpretations derived from registries.

Richard E. Gliklich
Michelle B. Leavy
Nancy A. Dreyer
Editors

Contents

Executive Summary **1**

Section I. Creating Registries **7**

Chapter 1. Patient Registries **8**

1. Introduction 8
2. Current Uses for Patient Registries 10
3. Taxonomy for Patient Registries 16
4. Patient Registries and Policy Purposes 18
5. Global Registries 21
6. Future of Registries 21
7. Summary 21

References for Chapter 1 22

Chapter 2. Planning a Registry **27**

1. Introduction 27
2. Steps in Planning a Registry 27
3. Anticipating and Preparing for Change 42
4. Special Considerations 43
5. Resources for Registries 45
6. Summary 45

References for Chapter 2 45

Case Examples for Chapter 2 50

Chapter 3. Registry Design **60**

1. Introduction 60
2. Research Questions Appropriate for Registries 60
3. Translating Clinical Questions Into Measurable Exposures and Outcomes 63
4. Finding the Necessary Data 64
5. Resources and Efficiency 65
6. Study Designs for Registries 65
7. Choosing Patients for Study 68
8. Registry Size and Duration 73
9. Internal and External Validity 75
10. Special Considerations 79
11. Summary 84

References for Chapter 3 85

Case Examples for Chapter 3 90

Chapter 4. Selecting and Defining Outcome Measures for Registries **98**

1. Introduction 98
2. Outcome Measures Framework 98
3. Types of Outcome Measures 101
4. Selecting Patient-Reported Outcomes 103
5. Standardized Outcome Measures 104
6. Conclusions 106

References for Chapter 4 107

Case Examples for Chapter 4 110

Chapter 5. Data Elements for Registries **114**

1. Introduction 114
2. Identifying Domains 114
3. Selecting Data Elements 115
4. Registry Data Map 125
5. Pilot Testing 126
6. Summary 127
References for Chapter 5 127
Case Examples for Chapter 5 128
Chapter 6. Data Sources for Registries 132
1. Introduction 132
2. Intended Uses for Data Elements 132
3. Types of Data 133
4. Data Sources 137
5. Other Considerations for Secondary Data Sources 148
6. Summary 150
References for Chapter 6 150
Case Examples for Chapter 6 153

Section II. Legal and Ethical Considerations for Registries 159
Chapter 7. Principles of Registry Ethics, Data Ownership, and Privacy 160
1. Introduction 160
2. Ethical Concerns Relating to Health Information Registries 168
3. Applicable Regulations 181
4. Registry Transparency, Oversight, and Data Ownership 203
5. Conclusions 207
References for Chapter 7 209
Chapter 8. Informed Consent 221
1. Introduction 221
2. Distinguishing Research and Other Activities 221
3. Growth of Database Research 226
4. Regulatory Consent Requirements 228
5. Approaches to Registry Consents 240
6. Consent Guidance 244
References for Chapter 8 247
Case Examples for Chapter 8 251
Chapter 9. Registry Governance 258
1. Introduction 258
2. Governance Roles 258
3. Governance Structure and Functions 259
4. Special Applications 264
5. Key Principles for Successful Governance 266
References for Chapter 9 267
Case Examples for Chapter 9 269

Section III. Operating Registries 273
Chapter 10. Recruiting and Retaining Participants in the Registry 274
1. Introduction 274
2. Recruitment 275

3. Retention 281
4. Pitfalls in Recruitment and Retention 283
References for Chapter 10 284
Case Examples for Chapter 10 286
Chapter 11. Obtaining Data and Quality Assurance 290
1. Introduction 290
2. Obtaining Data 290
3. Quality Assurance 304
4. Resource Considerations 311
References for Chapter 11 313
Chapter 12. Adverse Event Detection, Processing, and Reporting 315
1. Introduction 315
2. Identifying and Reporting Adverse Drug Events 315
3. Collecting AE Data in a Registry 318
4. AE Reporting by the Registry 319
5. Coding 321
6. Adverse Event Management 322
7. Adverse Event Required Reporting for Registry Sponsors 323
8. Special Case: Risk Evaluation and Mitigation Strategies (REMS) 325
9. Reporting Breaches of Confidentiality or Other Risks 326
References for Chapter 12 326
Chapter 13. Analysis, Interpretation, and Reporting of Registry Data To Evaluate Outcomes 329
1. Introduction 329
2. Research Questions and Registry Purposes 330
3. Patient Population 331
4. Data Quality 335
5. Data Analysis 340
6. Interpretation and Reporting of Registry Data 353
7. Summary 354
References for Chapter 13 354
Case Examples for Chapter 13 361

Section IV. Evaluating Registries 366
Chapter 14. Assessing Quality 368
1. Introduction 368
2. Defining Quality 369
3. Measuring Quality 369
4. Quality Domains 371

Contributors and Reviewers 385
Contributors 386
Reviewers 393
Case Example Contributors 400

Appendixes 409
Appendix A. An Illustration of Sample Size Calculations 410
Appendix B. Copyright Law 415

Tables

Table 3-1. Considerations for study design 60
Table 3-2. Overview of registry purposes 61
Table 3-3. Examples of research questions and key exposures and outcomes 63
Table 5-1. Standard terminologies 116
Table 5-2. Examples of potential registry data elements 119
Table 6-1. Key data sources—strengths and limitations 138
Table 7-1. Summary of Privacy Rule and Common Rule requirements 161
Table 9-1. Key points to consider when engaging patients in governance 264
Table 10-1. Hospital recruitment 276
Table 10-2. Physician recruitment 277
Table 10-3. Patient recruitment 279
Table 11-1. Data activities performed during registry coordination 312
Table 12-1. Overview of serious adverse event reporting requirements for marketed products 323
Table 13-1. Impact of loss to followup on incidence rates per 1,000 in a hypothetical study of 1,000 patients in a registry 347
Table 14-1. Overview of registry purposes 371
Table 14-2. Research quality—good practice for establishing and operating registries 372
Table 14-3. Evidence quality—indicators of good evidence quality for registries 378

Figures

Figure 1-1. Deciding when to develop a registry: The "value of information" exercise 20
Figure 4-1. Outcome Measures Framework 100
Figure 9-1. Committee functions comprising governance 260
Figure 12-1. Overview of reporting requirements for registries related to drug/biologic therapies 317
Figure 13-1. Patient populations 331

Executive Summary

Defining Patient Registries

This User's Guide is intended to support the design, implementation, analysis, interpretation, and quality evaluation of registries created to increase understanding of patient outcomes. For the purposes of this guide, a patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more stated scientific, clinical, or policy purposes. A registry database is a file (or files) derived from the registry. Although registries can serve many purposes, this guide focuses on registries created for one or more of the following purposes: to describe the natural history of disease, to determine clinical effectiveness or cost-effectiveness of healthcare products and services, to measure or monitor safety and harm, and/or to measure quality of care.

Registries are classified according to how their populations are defined. For example, product registries include patients who have been exposed to biopharmaceutical products or medical devices. Health services registries consist of patients who have had a common procedure, clinical encounter, or hospitalization. Disease or condition registries are defined by patients having the same diagnosis, such as cystic fibrosis or heart failure.

Planning a Registry

There are several key steps in planning a patient registry, including articulating its purpose, determining whether it is an appropriate means of addressing the research question, identifying stakeholders, defining the scope and target population, assessing feasibility, and securing funding. The registry team and advisors should be selected based on their expertise and experience. The plan for registry governance and oversight should clearly address such issues as overall direction and operations, scientific content, ethics, safety, data access, publications, and change management. It is also helpful to plan for the entire lifespan of a registry, including how and when the registry will end and any plans for transition at that time. Special consideration should be given to the unique challenges of planning specific types of registries, such as rare disease registries or quality improvement registries.

Registry Design

A patient registry should be designed with respect to its major purpose, with the understanding that different levels of rigor may be required for registries designed to address focused analytical questions to support decision making, in contrast to registries intended primarily for descriptive purposes. The key points to consider in designing a registry include formulating a research question; choosing a study design; translating questions of clinical interest into measurable exposures and outcomes; choosing patients for study, including deciding whether a comparison group is needed; determining where data can be found; and deciding how many patients need to be studied and for how long. Once these key design issues have been settled, the registry design should be reviewed to evaluate potential sources of bias (systematic error); these should be addressed to the extent that is practical and achievable. The information value of a registry is

enhanced by its ability to provide an assessment of the potential for bias and to quantify how this bias could affect the study results.

The specific research questions of interest will guide the registry's design, including the choice of exposures and outcomes to be studied and the definition of the target population (the population to which the findings are meant to apply). The registry population should be designed to approximate the characteristics of the target population as much as possible. The number of study subjects to be recruited and the length of observation (followup) should be planned in accordance with the overall purpose of the registry. The desired study size (in terms of subjects or person-years of observation) is determined by specifying the magnitude of an expected, clinically meaningful effect or the desired precision of effect estimates. Study size determinants are also affected by practicality, cost, and whether the registry is intended to support regulatory decision making. Depending on the purpose of the registry, internal, external, or historical comparison groups strengthen the understanding of whether the observed effects are indeed real and in fact different from what would have occurred under other circumstances. Registry study designs often restrict eligibility for entry to individuals with certain characteristics (e.g., age) to ensure that the registry will have subgroups with sufficient numbers of patients for analysis. Or the registry may use some form of sampling—random selection, systematic sampling, or a haphazard, nonrandom approach—to achieve this end.

Special consideration should be given to the unique challenges of designing registries for specific purposes, such as product safety surveillance, rare diseases, medical devices, and quality improvement.

Selecting and Defining Outcome Measures for Registries

The selection and definition of patient outcomes of interest is a critical step in designing a patient registry. The outcomes of interest, together with the exposures(s) of interest, drive many of the decisions regarding the study duration, the necessary data elements, and the source(s) of the data. Outcomes should be selected primarily based on the research questions of interest, with consideration given to the feasibility of capturing the desired outcomes within the study scope and budget. It is also important to consider the perspectives of multiple stakeholders when determining which outcomes are most relevant. Tools such as the Outcomes Measures Framework can be helpful to guide the selection and definition of outcome measures for use within registries. In addition, the use of standardized outcome measures or other data standards, when available, is essential so that registries can maximally contribute to evolving medical knowledge. Standard terminologies—and to a greater degree, higher level groupings into core datasets for specific conditions—not only improve efficiency in establishing registries but also promote more effective sharing, combining, or linking of datasets from different sources. Furthermore, the use of well-defined standards for data elements and data structure ensures that the meaning of information captured in different systems is the same. This is critical to maximize the value of registries as tools in learning health systems and a national research infrastructure.

Data Elements

The selection of data elements requires balancing such factors as their importance for the integrity of the registry and for the analysis of primary outcomes, their reliability, their contribution to the overall burden for respondents, and the incremental costs associated with their collection. Selection begins with identifying relevant domains. Specific data elements are then selected with consideration for established clinical data standards, common data definitions, and whether patient identifiers will be used. It is important to determine which elements are absolutely necessary and which are desirable but not essential. In choosing measurement scales for the assessment of patient-reported outcomes, it is preferable to use scales that have been appropriately validated, when such tools exist. Once data elements have been selected, a data map should be created, and the data collection tools should be pilot tested. Testing allows assessment of respondent burden, the accuracy and completeness of questions, and potential areas of missing data. Inter-rater agreement for data collection instruments can also be assessed, especially in registries that rely on chart abstraction. Overall, the choice of data elements should be guided by parsimony, validity, and a focus on achieving the registry's purpose.

Data Sources

A single registry may integrate data from various sources. The form, structure, availability, and timeliness of the required data are important considerations. Data sources can be classified as primary or secondary. Primary data are collected by the registry for its direct purposes. Secondary data have been collected by a secondary source for purposes other than the registry, and may not be uniformly structured or validated with the same rigor as the registry's primary data. Sufficient identifiers are necessary to guarantee an accurate match between data from secondary sources and registry patients. Furthermore, it is advisable to obtain a solid understanding of the original purpose of the secondary data, because the way those data were collected and verified or validated will help shape or limit their use in a registry. Common secondary sources of data linked to registries include medical records systems, institutional or organizational databases, administrative health insurance claims data, death and birth records, census databases, and related existing registry databases.

Ethics, Data Ownership, and Privacy

Critical ethical and legal considerations should guide the development and use of patient registries. The Common Rule is the uniform set of regulations on the ethical conduct of human subjects research, issued by the Federal agencies that fund such research. Institutions that conduct research agree to comply with the Common Rule for federally funded research, and may opt to apply that rule to all human subjects activities conducted within their facilities or by their employees and agents, regardless of the source of funding. The Privacy Rule, promulgated under the Health Insurance Portability and Accountability Act of 1996 (HIPAA), establishes Federal protections for the privacy of individually identifiable health information created and maintained by health plans, healthcare clearinghouses, and most healthcare providers (collectively, "covered entities"). The purpose of a registry, the type of entity that creates or maintains the registry, the types of entities that contribute data to the registry, and the extent to which registry data are individually identifiable affect how the regulatory requirements apply. Other important concerns

include transparency of activities, oversight, and data ownership. This chapter of the User's Guide focuses solely on U.S. law. Health information is also legally protected in European and some other countries by distinctly different rules.

Informed Consent for Registries

The requirement of informed consent often raises different issues for patient registries versus clinical trials. For example, registries may be used for public health or quality improvement activities, which may not constitute "human subjects research." Also, registries may integrate data from multiple electronic sources (e.g., claims data, electronic health records) and may be linked to biobanks. Institutional review boards may approve waivers or alterations of informed consent (e.g., electronic consent, oral consent) for some registries, depending on the purpose and risk to participants. Established registries that undergo a change in scope (e.g., changes in data sharing policies, changes to the protocol, extension of the followup period) may need to ask patients to "re-consent." When planning informed consent procedures, registry developers should consider several factors, including documentation and format, consent revisions and re-consent, the applicability of regulatory requirements, withdrawal of participants from the study, and the physical and electronic security of patient data and biological specimens. In addition, registry developers may need to consider the individual authorization requirements of the HIPAA Privacy Rule, where applicable.

Registry Governance

Registries function in a dynamic environment and are often shaped by the complex relationships among individual health, public health policy, economics, geography, and culture. Complexity within registries stems from the topics being studied, stakeholders with different agendas, and the legal and political climates for such research, among other factors. Governance is an important tool to help registries manage complexities such as these across the registry lifecycle, from the initial planning phase through the dissemination of results. Registry governance refers to a formalized structure or plan for managing the registry and guiding decision making related to registry funding, operations, and dissemination of information. Registry governance can take many forms depending on the scope of the registry, the number of stakeholders, and the purpose of the registry, but some principles for successful governance apply across all governance models. In particular, all aspects of governance should be codified in a written format that can be reviewed, shared, and refined over time, and transparency regarding any perceived or actual conflicts of interest is important for effective governance. Expectations of each research partner should be clearly delineated, pragmatic, and transparent. Lastly, policies and procedures should be developed to support stakeholder engagement and transparency.

Patient and Provider Recruitment and Management

Recruitment and retention of patients as registry participants, and of providers as registry sites, are essential to the success of a registry. Recruitment typically occurs at several levels, including facilities (hospitals, physicians' practices, and pharmacies), providers, and patients. The motivating factors for participation at each level and the factors necessary to achieve retention differ according to the registry. Factors that motivate participation include the perceived

relevance, importance, or scientific credibility of the registry, as well as a favorable balance of any incentives for participation versus the risks and burdens thereof. Because patient and provider recruitment and retention can affect how well a registry represents the target population, well-planned strategies for enrollment and retention are critical. Goals for recruitment, retention, and followup should be explicitly laid out in the registry planning phase, and deviations during the conduct of the registry should be continuously evaluated for their risk of introducing bias.

Obtaining Data and Quality Assurance

The integrated system for obtaining, cleaning, storing, monitoring, reviewing, and reporting on registry data determines the utility of those data for meeting the registry's goals. A broad range of procedures and systems are available for obtaining or collecting data. Some are more suitable than others for particular purposes. Critical factors in the ultimate quality of the data include how data elements are structured and defined, how personnel are trained, and how data problems (e.g., missing, out of range, or logically inconsistent values) are handled. Registries may also be required to conform to guidelines or to the standards of specific end users of the data (e.g., 21 Code of Federal Regulations, Part 11). Quality assurance aims to affirm that the data were, in fact, collected in accordance with established procedures and that they meet the requisite standards of quality to accomplish the registry's intended purposes and the intended use of the data. Requirements for quality assurance should be defined during the registry's inception and creation. Because certain requirements may have significant cost implications, a risk-based approach to developing a quality assurance plan is recommended. It should be based on identifying the most important or likely sources of error or potential lapses in procedures that may affect the quality of the registry in the context of its intended purpose.

Adverse Event Detection, Processing, and Reporting

The U.S. Food and Drug Administration defines an adverse event (AE) as any untoward medical occurrence in a patient administered a pharmaceutical product, whether or not related to or considered to have a causal relationship with the treatment. AEs are categorized according to the seriousness and, for drugs, the expectedness of the event. Although AE reporting for all marketed products is dependent on the principle of "becoming aware," collection of AE data falls into two categories: those events that are intentionally solicited (meaning data that are part of the uniform collection of information in the registry) and those that are unsolicited (meaning that the AE is volunteered or noted in an unsolicited manner). The determination of whether the registry should use a case report form to collect AEs should be based on the scientific importance of the information for evaluating the specified outcomes of interest. Regardless of whether or not AEs constitute a primary objective of the registry, it is important for any registry that has direct patient interaction to develop a plan for detecting, processing, and reporting AEs. If the registry receives sponsorship, in whole or in part, from a regulated industry (drugs or devices), the sponsor has mandated reporting requirements including stringent timelines, and the registry should establish the process for detecting and reporting AEs and should provide training to registry personnel on how to identify AEs and to whom they should be reported. Sponsors of registries designed specifically to meet requirements for surveillance of drug or device safety are encouraged to hold discussions with health authorities about the most appropriate process for reporting serious AEs.

Analysis, Interpretation, and Reporting of Registry Data

Analysis and interpretation of registry data begin with answering a series of core questions: Who was studied, and how were they chosen for study? How were the data collected, edited, and verified, and how were missing data handled? How were the analyses performed? Four populations are of interest in describing who was studied: the target population, the accessible population, the intended population, and the population actually studied (the "actual population"). The representativeness of the actual population to the target population is referred to as generalizability.

Analysis of registry outcomes first requires an analysis of recruitment and retention, of the completeness of data collection, and of data quality. Considerations include an evaluation of losses to followup; completeness for most, if not all, important covariates; and an understanding of how missing data were handled and reported. Analysis of a registry should provide information on the characteristics of the patient population, the exposures of interest, and the endpoints. Descriptive registry studies focus on describing frequency and patterns of various elements in a patient population, whereas analytical studies concentrate on associations between patients or treatment characteristics and health outcomes of interest. A statistical analysis plan describes the analytical plans and statistical techniques that will be used to evaluate the primary and secondary objectives specified in the study plan. Interpretation of registry data should be provided so that the conclusions can be understood in the appropriate context and any lessons from the registry can be applied to the target population and used to improve patient care and outcomes.

Evaluating Registries

Although registries can provide useful information, there are levels of rigor that enhance validity and make the information from some registries more useful for guiding decisions. The term "quality" can be applied to registries to describe the confidence that the design, conduct, and analysis of the registry can be shown to protect against bias and errors in inference—that is, erroneous conclusions drawn from the registry. Although there are limitations to any assessment of quality, a quality component analysis is used both to evaluate high-level factors that may affect results and to differentiate between research quality (which pertains to the scientific process) and evidence quality (which pertains to the data/ findings emanating from the research process). Quality components are classified as either "essential elements of good practice," which can be viewed as a checklist that should be considered for all patient registries, or as "potential enhancements to good practice," which may strengthen the value of the information in particular circumstances. The results of such an evaluation should be considered in the context of the disease area(s), the type of registry, and the purpose of the registry, and should also take into account feasibility and affordability.

Section I

Creating Registries

Chapter 1. Patient Registries

1. Introduction

The purpose of this document is to serve as a guide for the design and use of patient registries for scientific, clinical, and health policy purposes. Properly designed and executed, patient registries can provide a real-world view of clinical practice, patient outcomes, safety, and comparative effectiveness. This User's Guide primarily focuses on practical design and operational issues, evaluation principles, and best practices. Where topics are well covered in other materials, references and/or links are provided. The goal of this document is to provide stakeholders in both the public and private sectors with information they can use to guide the design and implementation of patient registries, the analysis and interpretation of data from patient registries, and the evaluation of the quality of a registry or one of its components. Where useful, case examples have been incorporated to illustrate particular points or challenges.

The term *registry*[1] is defined both as the act of recording or registering and as the record or entry itself. Therefore, "registries" can refer to both programs that collect and store data and the records that are so created.

The term *patient registry* is generally used to distinguish registries focused on health information from other record sets, but there is no consistent definition in current use. E.M. Brooke, in a 1974 publication of the World Health Organization, further delineated registries in health information systems as "a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a predetermined purpose."[2]

The National Committee on Vital and Health Statistics[3] describes registries used for a broad range of purposes in public health and medicine as "an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes [them] to the occurrence of a health-related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects."

Other terms also used to refer to patient registries include clinical registries, clinical data registries, disease registries, and outcomes registries.[4,5]

This User's Guide focuses on patient registries that are used for evaluating patient outcomes, defined as follows:

- A patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more stated scientific, clinical, or policy purposes.
- The patient registry database describes a file (or files) derived from the registry.

It is not intended to address several other types of or uses for registries (although many of the principles may be applicable), such as geographically based population registries (not based on a disease, condition, or exposure); registries created for public health reporting without tracking

outcomes (e.g., vaccine registries); or listing registries that are used solely to identify patients with particular diseases in clinical practices but are not used for evaluating outcomes.

Based on these definitions, the User's Guide focuses on patient registries in which the following are true (although exceptions may apply):

- The data are collected in a naturalistic manner, such that the management of patients is determined by the caregiver and patient together and not by the registry protocol.
- The registry is designed to fulfill specific purposes, and these purposes are typically defined before collecting or aggregating the data. In other words, the data collection or aggregation is purpose driven rather than the purpose being data driven (meaning limited to or derived from what is already available in an existing dataset). However, registries may add or change purposes over time.
- The registry captures data elements with specific and consistent data definitions.
- The data are collected in a uniform manner for every patient or can be transformed into uniform data for every patient. This consideration refers to both the types of data and the frequency of their collection.
- The data collected include data derived from and reflective of the clinical status of the patient (e.g., history, examination, laboratory test, or patient-reported data). Registries include the types of data that clinicians would use for the diagnosis and management of patients.
- It is common for at least one element of registry data collection to be active, meaning that some data are collected or derived specifically for the purpose of the registry (usually collected from the patient or clinician) rather than inferred from sources that are collected for another purpose (administrative, billing, pharmacy databases, etc.). Also, registry data collection may be a specific, but not the exclusive reason data are being collected, such as might occur with registries that incorporate data from electronic health records. This definition also does not exclude the incorporation of other data sources. Registries can be enriched by linkage with extant databases (e.g., to determine deaths and other outcomes or to assess pharmacy use or resource utilization), as discussed in Chapter 6.

Data from patient registries are generally used for studies that address the purpose for which the registry was created. In some respects, such as the collection of detailed clinical and longitudinal followup data, studies derived from the patient registries described in this User's Guide resemble traditional observational cohort studies. Beyond traditional cohort studies, however, some registry-based studies may be more flexible in that the scope and focus of the data collection activity of the registry may be adapted over time to address additional needs. For example, new studies, such as cluster-randomized studies or case-control studies, may be nested within an ongoing registry, and the database derived from the registry may be used to support secondary studies, such as studies that link the registry database with other data sources to explore new questions.

2. Current Uses for Patient Registries

A patient registry can be a powerful tool to observe the course of disease; to understand variations in treatment and outcomes; to examine factors that influence prognosis and quality of life; to describe care patterns, including appropriateness of care and disparities in the delivery of care; to assess effectiveness; to monitor safety and harm; and to measure quality of care. Through functionalities such as feedback of data, registries are also being used to study quality improvement.[6]

Different stakeholders perceive and may benefit from the value of registries in different ways. For example, for a clinician, registries can collect data about disease presentation and outcomes on large numbers of patients rapidly, thereby producing a real-world picture of disease, current treatment practices, and outcomes. For a physician organization, a registry might provide data that can be used to assess the number of real-world procedures performed using specific new technique or technology, to examine the degree to which clinicians are managing a disease in accordance with evidence-based guidelines, to evaluate the improvement in quality of life of patients following therapeutic management, to focus attention on specific aspects of a particular disease that might otherwise be overlooked, or to provide data for clinicians to compare themselves with their peers.[7] For patients and patient advocacy organizations, a registry may increase understanding of the natural history of a disease, contribute to the development of treatment guidelines, or facilitate research on treatment.[8,9] From a payer's perspective, registries can provide detailed information from large numbers of patients on how procedures, devices, or pharmaceuticals are actually used and on their effectiveness in different populations. This information may be useful for determining coverage policies or informing or supporting value-based care programs.[10] For a drug or device manufacturer, a registry-based study might demonstrate the performance of a product in the real world, meet a postmarketing commitment or requirement,[11] develop hypotheses, or identify patient populations that will be useful for product development, clinical trials design, and patient recruitment. The U.S. Food and Drug Administration (FDA) has noted that "through the creation of registries, a sponsor can evaluate safety signals identified from spontaneous case reports, literature reports, or other sources, and evaluate the factors that affect the risk of adverse outcomes such as dose, timing of exposure, or patient characteristics."[12]

Registries may also generate real-world data and real-world evidence to inform regulatory decision making. The 21st Century Cures Act requires the FDA to develop a program "to evaluate the use of real-world evidence (1) to help support the approval of a new indication for a drug approved under section 505(c); and (2) to help to support or satisfy post-approval study requirements."[13] The FDA defines real-world data (RWD) and real-world evidence (RWE) as "data relating to patient health status and/or the delivery of healthcare routinely collected from a variety of sources. Real-World Evidence (RWE) is the clinical evidence about the usage and potential benefits or risks, of a medical product derived from analysis of RWD. Examples of RWD include data derived from electronic health records (EHRs); medical claims and billing data; data from product and disease registries; patient-generated data, including from in-home-use settings; and data gathered from other sources that can inform on health status, such as mobile devices.[14] Some registries have already been used as a source of RWD to support regulatory decision making.[15-18]

The use of patient registries varies by condition, with cancer and cardiovascular disease having a large number of registries and conditions such as developmental delay or dementia, far fewer. Overall, the use of patient registries appears to be active and growing. For example, a review of ClinicalTrials.gov focusing on cancer reveals over 650 entries for patient registries with a wide range of purposes. Of these, four have more than 100,000 participants, and 56 have more than 10,000. In some cases, the drivers for these registries have been Federal stakeholders. Since 2005, the FDA Center for Devices and Radiological Health has called for some 265 post-approval studies, many of which use new or existing registries to study the real-world effectiveness of specific devices in community practice.[19]

2.1 Evaluating Patient Outcomes

Studies from patient registries and randomized controlled trials (RCTs) have important and complementary roles in evaluating patient outcomes.[20] Ideally, patient registries collect data in a comprehensive manner (with few excluded patients) and therefore produce outcome results that may be generalizable to a wide range of patients. They also evaluate care as it is actually provided, because care is not assigned, determined, or even recommended by a protocol. As a result, the outcomes reported may be more representative of what is achieved in real-world practice. Patient registries also offer the ability to evaluate patient outcomes when clinical trials are not practical (e.g., very rare diseases), and they may be the only option when clinical trials are not ethically acceptable. They are a powerful tool when RCTs are difficult to conduct, such as in surgery or when very long-term outcomes are desired.

RCTs are controlled experiments designed to test hypotheses that can ultimately be applied to real-world care. Because RCTs are often conducted under strict constraints, with detailed inclusion and exclusion criteria (and the need for subjects who are willing to be randomized), they are sometimes limited in their generalizability. If RCTs are not generalizable to the populations to which the information will be applied, they may not be sufficiently informative for decision making. Conversely, patient registries that observe real-world clinical practice may collect the information needed to assess patient outcomes in a generalizable way, but interpreting this information correctly requires analytic methodology geared to address the potential sources of bias that challenge observational studies (see Chapter 13). Interpreting patient registry data also requires checks of internal validity and sometimes the use of external data sources to validate key assumptions (such as comparing the key characteristics of registry participants with external sources in order to demonstrate the comparability of registry participants with the ultimate reference population). Patient registries, RCTs, other study designs, and other data sources should all be considered tools in the toolbox for evidence development, each with its own advantages and limitations.[21]

2.2 Hierarchies of Evidence

One question that arises in a discussion of this type is where to place studies derived from patient registries within the hierarchies of evidence that are frequently used in developing guidelines or decision making. While the definition of patient registry used in this User's Guide is intentionally broad, the parameters of quality described in Chapter 14 are intended to help the user evaluate and identify registries that are sufficiently rigorous observational studies for use as evidence in decision making. Many registries are, or include, high-quality studies of cohorts

designed to address a specific problem and hypothesis. Still, even the most rigorously conducted registries, like prospective observational studies, are traditionally placed in a subordinate position to RCTs in some commonly used hierarchies, although equal to RCTs in others.[22-24] Debate continues in the evidence community regarding these traditional methods of grading levels of evidence, their underlying assumptions, their shortcomings in assessing certain types of evidence (e.g., benefit vs. harm), and their interscale consistency in evaluating the same evidence.[21,25,26]

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group has proposed a more robust approach that addresses some of the decision-making issues described in this User's Guide. As noted by the GRADE collaborators:

> [R]andomised trials are not always feasible and, in some instances, observational studies may provide better evidence, as is generally the case for rare adverse effects. Moreover, the results of randomised trials may not always be applicable—for example, if the participants are highly selected and motivated relative to the population of interest. It is therefore essential to consider study quality, the consistency of results across studies, and the directness of the evidence, as well as the appropriateness of the study design.[27]

AHRQ has also developed a guidance system for grading the strength of evidence that recommends a careful assessment of the potential value of observational studies. The guidance, which is designed to support the systematic reviews conducted by the Evidence-based Practice Center (EPC) program, is conceptually similar to the GRADE system.[28] When using the AHRQ approach, reviewers typically give evidence from observational studies a low starting grade and evidence from RCTs a high starting grade. These initial grades can then be raised or lowered depending on the strength of the five required evidence domains (study limitations, directness, consistency, precision, and reporting bias).[29] For example, the reviewers may find that observational studies are particularly relevant for some systematic review questions. The report notes:

> EPCs may act on the judgment that, for certain outcomes such as harms, observational studies have less risk of bias than do RCTs or that the available RCTs have a substantial risk of bias. In such instances, the EPC may move up the initial grade for strength of evidence based on observational studies to moderate or move down the initial rating based on RCTs to moderate.[28]

Reviewers may also raise or lower evidence grades based on a secondary set of domains (dose-response association, existence of confounding that would diminish an observed effect, and strength of association). These secondary domains supplement the required domains and are used when relevant to the systematic review question. The report explains that the secondary domains "may increase strength of evidence and are especially relevant for observational studies where one may begin with a lower overall strength of evidence grade based on study limitations."[28]

As the methods for grading evidence for different purposes continue to evolve, this User's Guide can serve as a guide to help such evaluators understand study quality and identify well-designed registries. Beyond the evidence hierarchy debate, users of evidence understand the value of registries for providing complementary information that can extend the results of clinical trials to populations not studied in those trials, for demonstrating the real-world effects of treatments

outside of the research setting (potentially in large subsets of affected patients), and for providing long-term followup when such data are not available from clinical trials.

2.3 Defining Patient Outcomes

The focus of this User's Guide is the use of registries to evaluate patient outcomes. An outcome may be thought of as an end result of a particular healthcare practice or intervention. According to AHRQ, end results include effects that people experience and care about.[30] The National Cancer Institute further clarifies that "final" endpoints are those that matter to decision-makers: patients, providers, private payers, government agencies, accrediting organizations, or society.[31] Examples of these outcomes include biomedical outcomes, such as survival and disease-free survival, health-related quality of life, satisfaction with care, and economic burden.[32] Although final endpoints are ultimately what matter, it is sometimes more practical when creating registries to collect intermediate outcomes (such as whether processes or guidelines were followed) and clinical outcomes (such as whether a tumor regressed or recurred) that predict success in improving final endpoints. Clinical outcomes are particularly important in chronic conditions such as asthma and diabetes, where the treatment intent is management.

In *Crossing the Quality Chasm*,[33] the Institute of Medicine (now the National Academy of Medicine) describes the six guiding aims of healthcare as providing care that is safe, effective, efficient, patient-centered, timely, and equitable. (The last three aims focus on the delivery and quality of care.) While these aims are not outcomes per se, they generally describe the dimensions of results that matter to decision-makers in the use of a healthcare product or service: Is it safe? Does it produce greater benefit than harm? Is it clinically effective? Does it produce the desired effect in real-world practice? Does the right patient receive the right therapy or service at the right time? Is it cost effective or efficient? Does it produce the desired effect at a reasonable cost relative to other potential expenditures? Is it patient oriented, timely, and equitable? Most of the patient outcomes that registries evaluate reflect one or more of the six guiding aims. For example, a patient presenting with an ischemic stroke to an emergency room has a finite window of opportunity to receive a thrombolytic drug, and the patient outcome, whether the patient achieves full recovery, is dependent not only on the product's dissolving the clot but also on the timeliness of its delivery.[34,35]

When defining patient outcomes, as well as other data elements, registries are advised to consider using existing core or minimum sets of outcome measures and common data elements whenever possible. Core or minimum sets of outcome measures specify and define the critical outcomes of interest in a disease or condition area; these recommendations generally are developed through consensus-based processes. Common data elements (CDEs) are standardized terms that can be used in multiple research settings. Core sets of outcome measures and CDEs can be found in the scientific literature. Some initiatives, such as the National Library of Medicine's CDE Resource Portal[36] and the COMET Initiative,[37] compile information on existing CDEs and core outcome sets. Use of core sets of outcome measures and CDEs is intended to facilitate comparing, aggregating, and linking data across registries and other research efforts. AHRQ has funded the development of some minimum sets of outcome measures specific to registries through the Outcome Measures Framework (OMF) project.[38,39] The OMF is a common model for measuring patient and clinician relevant outcomes across different conditions for registries and clinical practice.[23] Through a series of collaborative stakeholder and patient

registry working groups, minimum measure sets are or have been developed and published for several conditions.

2.4 Purposes of Registries

As discussed throughout this User's Guide, registries should be designed and evaluated with respect to their intended purpose(s). Registry purposes can be broadly described in terms of patient outcomes. While there are a number of potential purposes for registries, this handbook primarily discusses four major purposes: (1) describing the natural history of disease, (2) determining clinical and/or cost-effectiveness, (3) assessing safety or harm, and (4) measuring or improving quality of care. Other purposes of patient registries mentioned but not discussed in detail in this User's Guide are public health surveillance and disease control. An extensive body of literature from the last half century of experience with cancer and other disease surveillance registries is available.

2.4.1 Describing Natural History of Disease

Registries may be established to evaluate the natural history of a disease, meaning its characteristics, management, and outcomes with and/or without treatment. The natural history may be variable across different groups or geographic regions, and it often changes over time. In many cases, the natural histories of diseases are not well described. Furthermore, the natural histories of diseases may change after the introduction of certain therapies. As an example, patients with rare diseases, such as the lysosomal storage diseases, who did not previously survive to their 20s, may now be entering their fourth and fifth decades of life, and this uncharted natural history is being first described through a registry.[40] An ancillary purpose of some registries tracking the natural history of a disease is to facilitate identification of participants for potential recruitment into clinical trials.[41]

2.4.2 Determining Effectiveness

Registries may be developed to determine clinical effectiveness or cost-effectiveness in real-world clinical practice. Multiple studies have demonstrated disparities between the results of clinical trials and results in actual clinical practice.[42,43] Furthermore, efficacy in a clinical trial for a well-defined population may not be generalizable to other populations or subgroups of interest. As an example, many important heart failure trials have focused on a predominantly white male population with a mean age of approximately 60 years, whereas actual heart failure patients are older, more diverse, and have a higher mortality rate than the patients in these trials.[44] Similarly, underrepresentation of older patients has been reported in clinical trials of 15 different types of cancer (e.g., studies with only 25 percent of patients age 65 years and over, while the expected rate is greater than 60 percent).[45] Data from registries have been used to fill these gaps for decision-makers. For example, the FDA used the American Academy of Ophthalmology's intraocular lens registry to expand the label for intraocular lenses to younger patients.[46] Registries may also be particularly useful for tracking effectiveness outcomes for a longer period than is typically feasible with clinical trials. For example, some growth hormone registries have tracked children well into adulthood.

In addition to clinical effectiveness, registries can be used to assess cost-effectiveness. Registries can be designed to collect cost data and effectiveness data for use in modeling cost-effectiveness.[47] Cost-effectiveness is a means to describe the comparative value of a healthcare product or service in terms of its ability to achieve a desired outcome for a given unit of resources.[48] A cost-effectiveness analysis examines the incremental benefit of a particular intervention and the costs associated with achieving that benefit. Cost-effectiveness studies compare costs with clinical outcomes measured in units such as life expectancy or disease-free periods. Cost-utility studies compare costs with outcomes adjusted for quality of life (utility), such as quality-adjusted life years (QALYs). Utilities allow comparisons to be made across conditions because the measurement is not disease specific.[49] It should be noted that for both clinical effectiveness and cost-effectiveness, differences between treatments are indirect and must be inferred from data analysis, simulation modeling, or some mixture.

With improvement in methodologies, including better methods for managing bias and better understanding of the limitations, there is increasing interest and investment from many stakeholders in registries for comparative effectiveness research (CER) and patient-centered outcomes research (PCOR). The Patient-Centered Outcomes Research Institute (PCORI) has defined PCOR as research that "assesses the benefits and harms of preventive, diagnostic, therapeutic, palliative, or health delivery system interventions to inform decision making, highlighting comparisons and outcomes that matter to people; is inclusive of an individual's preferences, autonomy and needs, focusing on outcomes that people notice and care about such as survival, function, symptoms, and health-related quality of life; incorporates a wide variety of settings and diversity of participants to address individual differences and barriers to implementation and dissemination; and investigates (or may investigate) optimizing outcomes while addressing burden to individuals, availability of services, technology, and personnel, and other stakeholder perspectives."[50] PCORI has recognized registries as an important potential source of data to support PCOR, in large part because of their ability to provide information on 'real-world' settings and broad patient populations. PCORI included minimum standards for the use of registries for PCOR in the Methodology Report.[51] While some registries are designed explicitly to examine questions of comparative effectiveness or patient-centered outcomes research, many others are designed for different objectives yet still collect data that are useful for these analyses. Registries that were not explicitly designed for CER or PCOR may need to be augmented or linked to other data sources—for example, to obtain long-term outcomes data in the case of an in-hospital registry using linkage to claims data to evaluate blood pressure medications.[52]

2.4.3 Measuring or Monitoring Safety and Harm

Registries may be created to assess safety versus harm. Safety here refers to the concept of being free from danger or hazard. One goal of registries in this context may be to quantify risk or to attribute it properly. Broadly speaking, patient registries can serve as an active surveillance system for the occurrence of unexpected or harmful events for products and services. Such events may range from patient complaints about minor side effects to severe adverse events such as fatal drug reactions or patient falls in the hospital.

Patient registries offer multiple advantages for active surveillance. First, the current practice of spontaneous reporting of adverse events relies on a nonsystematic recognition of an adverse

event by a clinician and the clinician's active effort to make a report to manufacturers and health authorities. (In the United States, patients may also report adverse events directly to the FDA.) Second, these events are generally reported without a denominator (i.e., the exposed or treated population), and therefore an incidence rate is difficult to determine. Because patient registries can provide systematic data on adverse events and the incidence of these events, they are being used with increasing frequency in the areas of healthcare products and services. For example, the FDA recently established the National Evaluation System for health Technology (NEST) to integrate data from patient registries, electronic health records, and administrative claims data to efficiently generate evidence of medical device safety and effectiveness.[53]

2.4.4 Measuring Quality

Registries may be created to measure quality of care. The Institute of Medicine defined quality as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge." Quality-focused registries are being used increasingly to assess differences between providers or patient populations based on performance measures that compare treatments provided or outcomes achieved with "gold standards" (e.g., evidence-based guidelines) or comparative benchmarks for specific health outcomes (e.g., risk-adjusted survival or infection rates). Such programs may be used to identify disparities in access to care, demonstrate opportunities for improvement, establish differentials for payment by third parties, or provide transparency through public reporting. There are multiple examples of such differences in treatment and outcomes of patients in a range of disease areas.[54-58] As healthcare transitions from quality measurement focused on adherence with evidence-based practice standards to longitudinal, standardized, patient-level outcomes assessment, registries will play an important role.

2.4.5 Multiple Purposes

Many registries will be developed to serve more than one of these purposes. Registries developed for one purpose may also be modified to serve additional purposes as the research, practice, or policy environment changes. While registries often serve more than one purpose, their original or primary purpose generally guides their design and, as a result, more care is needed in evaluating results for secondary or additional purposes.

3. Taxonomy for Patient Registries

Even limited to the definitions described above, the breadth of studies that might be described as patient registries is large. Patients in a registry are typically selected based on a particular disease, condition (e.g., a risk factor), or exposure. This User's Guide uses these common selection criteria to develop a taxonomy or classification based on how the populations for registries are defined. Three general categories with multiple subcategories and combinations account for the majority of registries that are developed for evaluating patient outcomes. These categories include observational studies in which the patient has had an exposure to a product or service, or has a particular disease or condition, or various combinations thereof.

3.1 Product Registries

In the case of a product registry, the patient is exposed to a healthcare product, such as a drug or a device, or other potential risk factor (e.g., environmental or personal exposure). The exposure may be brief, as in a single dose of a pharmaceutical product, or extended, as in an implanted device or chronic usage of a medication.

Device registries may include all, or a subset, of patients who receive the device. A registry for all patients who receive an implantable cardioverter defibrillator, a registry of patients with hip prostheses, or a registry of patients who wear contact lenses are all examples of device registries. Biopharmaceutical product registries similarly have several archetypes, which may include all, or subsets, of patients who receive the biopharmaceutical product. For example, the British Society for Rheumatology established a national registry of patients on biologic therapy.[59] Again, the duration of exposure may range from a single event to a lifetime of use. Eligibility for the registry includes the requirement that the patient received the product or class of products (e.g., COX-2 inhibitors). In some cases, public health authorities mandate such registries to ensure safe use of medications. Examples include registries for thalidomide, clozapine, and isotretinoin.

Pregnancy registries represent a separate class of biopharmaceutical product registries that focus on possible exposures during pregnancy and the neonatal consequences. The FDA has a specific guidance focused on pregnancy exposure registries.[60]

3.2 Health Services Registries

In the context of evaluating patient outcomes, another type of exposure that can be used to define registries is exposure to a healthcare service. Healthcare services that may be used to define inclusion in a registry include individual clinical encounters, such as office visits or hospitalizations, procedures, or full episodes of care. Examples include registries enrolling patients undergoing a procedure (e.g., carotid endarterectomy, appendectomy, or primary coronary intervention) or admitted to a hospital for a particular diagnosis (e.g., community-acquired pneumonia). In these registries, one purpose of the registry is to evaluate the healthcare service with respect to the outcomes. Healthcare service registries are sometimes used to evaluate the processes and outcomes of care for quality measurement purposes (e.g., Get With The Guidelines® of the American Heart Association, National Surgical Quality Improvement Program of the Department of Veterans Affairs and the American College of Surgeons).

3.3 Disease or Condition Registries

Disease or condition registries use the state of a particular disease or condition as the inclusion criterion. In disease or condition registries, the patient may always have the disease (e.g., a rare disease such as cystic fibrosis or Pompe disease, or a chronic illness such as heart failure, diabetes, or end-stage renal disease) or may have the disease or condition for a more limited period of time (e.g., infectious diseases, some cancers, obesity). These registries typically enroll the patient at the time of a routine healthcare service, although patients also can be enrolled through voluntary self-identification processes that do not depend on utilization of healthcare services (such as Internet recruiting of volunteers). In other disease registries, the patient has an

underlying disease or condition, such as atherosclerotic disease, but is enrolled only at the time of an acute event or exacerbation, such as hospitalization for a myocardial infarction or ischemic stroke.

3.4 Combinations

Complicating this classification approach is the reality that these categories can be overlapping in many registries. For example, a patient with ischemic heart disease may have an acute myocardial infarction and undergo a primary coronary intervention with placement of a drug-eluting stent and postintervention management with clopidogrel. This patient could be enrolled in an ischemic heart disease registry tracking all patients with this disease over time, a myocardial infarction registry that is collecting data on patients who present to hospitals with acute myocardial infarction (cross-sectional data collection), a primary coronary intervention registry that includes management with and without devices, a coronary artery stent registry limited to ischemic heart disease patients, or a clopidogrel product registry that includes patients undergoing primary coronary interventions.

3.5 Duration of Observation

The duration of the observational period for a registry is also a useful descriptor. Observation periods may be limited to a single episode of care (e.g., a hospital discharge registry for diverticulitis), or they may extend for as long as the lifetime of patients with a chronic disease (e.g., cystic fibrosis or Pompe disease) or patients receiving a novel therapy (e.g., gene therapy). The period of observation or followup depends on the outcomes of interest.

3.6 From Registry Purpose to Design

As will be discussed extensively in this document, the purpose of the registry defines the registry focus (e.g., product vs. disease) and therefore the registry type. A registry created for the purpose of evaluating outcomes of patients receiving a particular coronary artery stent might be designed as a single product registry if, for example, the purpose is to systematically collect adverse event information on the first 10,000 patients receiving the product. However, the registry might alternatively be designed as a healthcare service registry for primary coronary intervention if the purpose is to collect comparative effectiveness or safety data on other treatments or products within the same registry.

4. Patient Registries and Policy Purposes

In addition to the growth of patient registries for scientific and clinical purposes, registries are receiving increased attention for their potential role in policymaking or decision making. As stated earlier, registries may offer a view of real-world healthcare that is typically inaccessible from clinical trials or other data sources and may provide information on the generalizability of the data from clinical trials to populations not studied in those trials.

The utility of registry data for decision making is related to three factors: the stakeholders, the primary scientific question, and the context. The stakeholders are those associated with the

disease or procedure that may be affected from a patient, provider, payer, regulator, or other perspective. The primary scientific question for a registry may relate to effectiveness, safety, or practice patterns. The context includes the scientific context (e.g., previous randomized trials and modeling efforts that help to more precisely define the primary scientific question), as well as the political, regulatory, funding, and other issues that provide the practical parameters around which the registry is developed. In identifying the value of information from registries, it is essential to look at the data with specific reference to the purpose and focus of the registry.

From a policy perspective, there are several scenarios in which the decision to develop a registry may arise. One possible scenario is as follows. An item or service is considered for use. Stakeholders in the decision collaboratively define "adequate data in support of the decision at hand." Here, "adequate data" refers to information of sufficient relevance and quality to permit an informed decision. An evidence development strategy is selected from one of many potential strategies (RCT, practical clinical trial, registry, etc.) based on the quality of the evidence provided by each design, as well as the burden of data collection and the cost that is imposed. This tradeoff of the quality of evidence versus cost of data collection for each possible design is termed the "value of information" exercise (Figure 1–1). Registries should be preferred in those circumstances where they provide sufficiently high-quality information for decision making at a sufficiently low cost (relative to other "acceptable" designs).

One set of policy determinations that may be informed by a patient registry centers on the area of payment for items or services. For example, the Centers for Medicare & Medicaid Services (CMS) issued Guidance on National Coverage Determinations With Data Collection as a Condition of Coverage in 2006. That original guidance document (which has undergone subsequent revisions, most recently in 2014) provided several examples of how data collected in a registry might be used in the context of coverage determinations. As described in the Guidance:

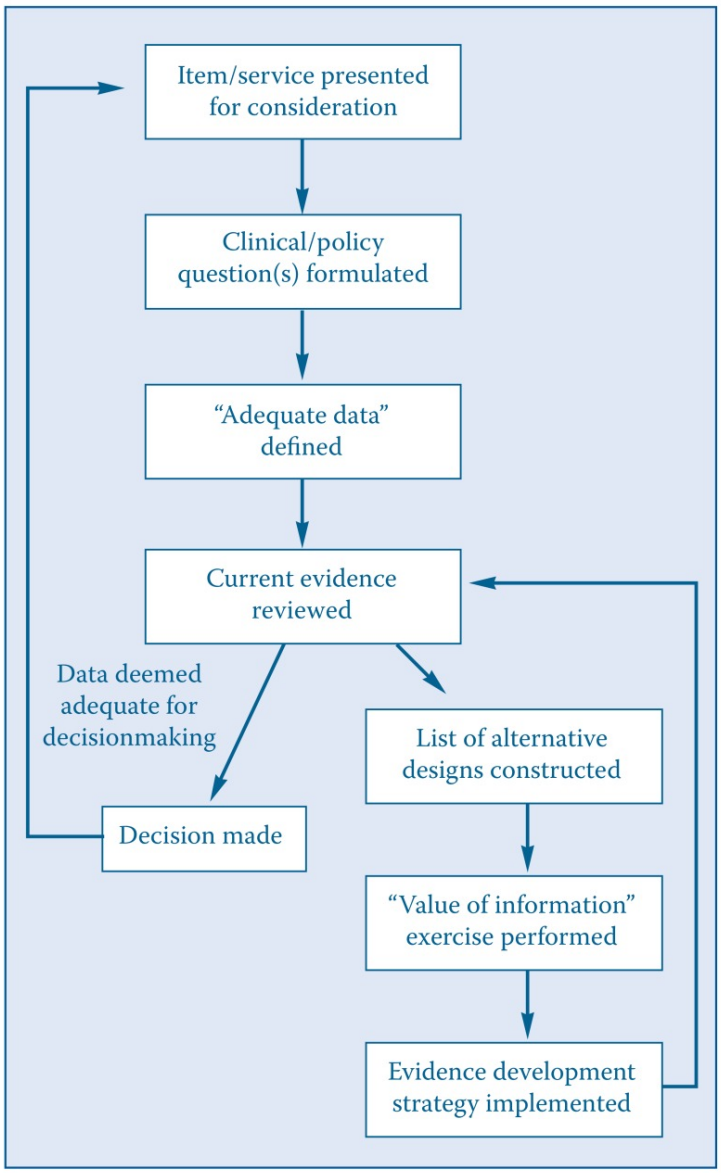
> [T]he purpose of CED [Coverage with Evidence Development] is to generate data on the utilization and impact of the item or service evaluated in the NCD [National Coverage Determination], so that Medicare can (a) document the appropriateness of use of that item or service in Medicare beneficiaries under current coverage; (b) consider future changes in coverage for the item or service; (c) generate clinical information that will improve the evidence base on which providers base their recommendations to Medicare beneficiaries regarding the item or service.[61]

The Guidance provided insight into when registry data may be useful to policymakers. These purposes range from demonstrating that a particular item or service was provided appropriately to patients meeting specific characteristics, to collecting new information that is not available from existing clinical trials. CED based on registries may be especially relevant when current data do not address relevant outcomes for beneficiaries, off-label or unanticipated uses, important patient subgroups, or operator experience or other qualifications. Registry-based studies may also be important when an existing treatment is being reconsidered. (An RCT may not be possible under such circumstances.) Registry-based studies are also being used increasingly in fulfillment of postmarketing commitments and requirements.

In many countries, policy determinations on payment rely on cost-effectiveness and cost-utility data and therefore can be informed by registries as well as clinical trials.[62] These data are used and reviewed in a variety of ways. In some countries, there may be a threshold above which a

payer is willing to pay for an improvement in patient outcomes.[63] In these scenarios—particularly for rare diseases, when it can be difficult to gather clinical effectiveness data together with quality-of-life data in a utility format—the establishment of disease-specific data registries has been recommended to facilitate the process of technology assessment and improving patient care.[63] In fact, the use of new or existing registries to assess health technology or risk-sharing arrangements is growing in such countries as the United Kingdom, France, Germany, and Australia, and in conditions ranging from bariatric surgery to stroke care.[64-67]

Figure 1-1. Deciding when to develop a registry: The "value of information" exercise



Consider the clinical question of carotid endarterectomy surgery for patients with a high degree of stenosis of the carotid artery. Randomized trials, using highly selected patients and surgeons, indicate a benefit of surgery over medical management in the prevention of stroke. However, that benefit may be exquisitely sensitive to the surgical complication rates; a relatively small increase

in the rate of surgical complications is enough to make medical management the preferred strategy instead. In addition, the studies of surgical performance in a variety of hospitals may suggest substantial variation in surgical mortality and morbidity for this procedure. In such a case, a registry to evaluate treatment outcomes, adjusted by hospital and surgeon, might be considered to support a policy decision as to when the procedure should be reimbursed (e.g., only when performed in medical centers resembling those in the various randomized trials, or only by surgeons or facilities with an acceptably low rate of complications).[68]

5. Global Registries

As many stakeholders have international interests in diseases, conditions, and healthcare products and services, it is not surprising that interest in global patient registries is growing. While some of the specific legal and regulatory discussions in this User's Guide are intended for and limited to the United States, most of the concepts and specifics are more broadly applicable to similar activities worldwide. Chapter 7 (ethics, data ownership, and privacy), Chapter 8 (informed consent), and Chapter 12 (adverse event detection, processing, and reporting) are perhaps the most limited in their applicability outside the United States. There may be additional considerations in data element selection and patient-reported outcome measure selection (Chapter 5) stemming from differences ranging from medical training to use of local remedies; the types of data sources that are available outside the United States (Chapter 6); the issues surrounding clinician and patient recruitment and retention in different health systems and cultures (Chapter 10); and specific data collection and management options and complexities (Chapter 11), ranging from available technologies to languages. It is also important to note that this User's Guide does not address the new European Union (EU) Clinical Trials Regulation scheduled to go into effect in 2019.

6. Future of Registries

With the growing availability of digitized healthcare data and specialized technologies such as natural language processing and machine learning, and the ability to link different data sources from different settings, registries are rapidly evolving. These technologies have greatly advanced and automated patient registry data collection and analyses and are enabling more timely and comprehensive real-world data for a number of the purposes described above. Coupled with standardized outcomes as described later in this document, this evolution of registries will support more timely feedback of critical information to healthcare stakeholders for decision making and will foster the development of learning healthcare systems.[69]

7. Summary

A patient registry is an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure and that serves stated scientific, clinical, or policy purpose(s). Studies derived from well-designed and well-performed patient registries can provide a real-world view of clinical practice, patient outcomes, safety, and clinical, comparative, and cost-effectiveness, and can serve a number of evidence development and decision-making

purposes. In the chapters that follow, this User's Guide presents practical design and operational issues, evaluation principles, and good registry practices.

References for Chapter 1

1. Jacobs V, Cutler MJ, Day JD, et al. Atrial fibrillation and dementia. Trends Cardiovasc Med. 2015;25(1):44-51. PMID: 25442732. DOI: 10.1016/j.tcm.2014.09.002.

2. Brooke EM. The current and future use of registers in health information systems. Geneva: World Health Organization; 1974. Report No.: Publication No. 8.

3. Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. BMJ. 2017;357:j2262. PMID: 28515234. DOI: 10.1136/bmj.j2262.

4. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. Am Heart J. 2009;157(6):971-82. PMID: 19464406. DOI: 10.1016/j.ahj.2009.03.023.

5. Hammill BG, Hernandez AF, Peterson ED, et al. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. Am Heart J. 2009;157(6):995-1000. PMID: 19464409. DOI: 10.1016/j.ahj.2009.04.002.

6. LaBresh KA, Gliklich R, Liljestrand J, et al. Using "get with the guidelines" to improve cardiovascular secondary prevention. Jt Comm J Qual Saf. 2003;29(10):539-50. PMID: 14567263.

7. Kennedy L, Craig AM. Global registries for measuring pharmacoeconomic and quality-of-life outcomes: focus on design and data collection, analysis and interpretation. Pharmacoeconomics. 2004;22(9):551-68. PMID: 15209525. DOI: 10.2165/00019053-200422090-00001.

8. Charrow J, Esplin JA, Gribble TJ, et al. Gaucher disease: recommendations on diagnosis, evaluation, and monitoring. Arch Intern Med. 1998;158(16):1754-60. PMID: 9738604.

9. Bushby K, Lynn S, Straub T, et al. Collaborating to bring new therapies to the patient--the TREAT-NMD model. Acta Myol. 2009;28(1):12-5. PMID: 19772190.

10. Dhruva SS, Phurrough SE, Salive ME, et al. CMS's landmark decision on CT colonography--examining the relevant data. N Engl J Med. 2009;360(26):2699-701. PMID: 19474419. DOI: 10.1056/NEJMp0904408.

11. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131(4):e29-322. PMID: 25520374. DOI: 10.1161/CIR.0000000000000152.

12. Sanchez JM, Ballinger B, Olgin JE, et al., editors. AF Detection and Ablation Outcomes: Answering Questions That Matter to Patients: Detecting Atrial Fibrillation using a Smart Watch – the mRhythm study. HRS Heart Rhythm 2017; 2017 May 11; Chicago, IL.

13. 45 CFR 164.514 (a)-(c).

14. U.S. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. https://www.fda.gov/media/120060/download. Accessed June 3, 2019.

15. Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using "Real-World" Data. JAMA. 2017;318(8):703-4. PMID: 28715550. DOI: 10.1001/jama.2017.9991.

16. Shuren J., Zuckerman B. How Creative FDA Regulation Led to First-in-the-World Approval of a Cutting-Edge Heart Valve. FDA Voice. U.S. Food and Drug Administration. June 14, 2017. https://blogs.fda.gov/fdavoice/index.php/tag/transcatheter-aortic-valve-replacement-tavr/. Accessed January 17, 2019.

17. Lutonix® 035 Drug Coated Balloon PTA Catheter. Summary of Safety and Effectiveness Data. U.S. Food and Drug Administration. February 7, 2017. https://www.accessdata.fda.gov/cdrh_docs/pdf13/P130024S009B.pdf. Accessed June 4, 2019.

18. Gliklich RE, Leavy MB. Assessing Real-World Data Quality: The Application of Patient Registry Quality Criteria to Real-World Data and Real-World Evidence. Ther Innov Regul Sci. 2019:2168479019837520. PMID: 30917687. DOI: 10.1177/2168479019837520.

19. U.S. Food and Drug Administration. FDA Post-Approval Studies. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMA/pma_pas.cfm?utm_campaign=Google2&utm_source=fdaSearch&utm_medium=website&utm_term=post-approval%20studies&utm_content=1. Accessed June 4, 2019.

20. Dreyer NA, Garner S. Registries for robust evidence. JAMA. 2009;302(7):790-1. PMID: 19690313. DOI: 10.1001/jama.2009.1092.

21. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25):1887-92. PMID: 10861325. DOI: 10.1056/NEJM200006223422507.

22. Guyatt GH, Sackett DL, Sinclair JC, et al. Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. JAMA. 1995;274(22):1800-4. PMID: 7500513.

23. Gliklich RE, Leavy MB, Karl J, et al. A framework for creating standardized outcome measures for patient registries. J Comp Eff Res. 2014;3(5):473-80. PMID: 25350799. DOI: 10.2217/cer.14.38.

24. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008;5(3):e67. PMID: 18336067. DOI: 10.1371/journal.pmed.0050067.

25. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res. 2004;4(1):38. PMID: 15615589. DOI: 10.1186/1472-6963-4-38.

26. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clin Med (Lond). 2008;8(6):579-88. PMID: 19149278. DOI: 10.7861/clinmedicine.8-6-579.

27. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328(7454):1490. PMID: 15205295. DOI: 10.1136/bmj.328.7454.1490.

28. Berkman ND, Lohr KN, Ansari M, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Methods for Effective Health Care. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.

29. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Methods for Effective Health Care. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.

30. Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. Science. 1998;282(5387):245-6. PMID: 9841388.

31. Lipscomb J, Snyder CF. The Outcomes of Cancer Outcomes Research: focusing on the National Cancer Institute's quality-of-care initiative. Med Care. 2002;40(6 Suppl):III3-10. PMID: 12064755.

32. Lipscomb J, Donaldson MS, Hiatt RA. Cancer outcomes research and the arenas of application. J Natl Cancer Inst Monogr. 2004(33):1-7. PMID: 15504917. DOI: 10.1093/jncimonographs/lgh038.

33. Hurtado MP, Swift EK, Corrigan JM. Crossing the quality chasm: a new health system for the 21st Century. Washington DC: National Academy Press, Institute of Medicine; 2001.

34. Schwamm LH, LaBresh KA, Pan WQ, et al. Get With The Guidelines – Stroke produces sustainable improvements in hospital-based acute stroke care. Stroke. 2006;37(2):624.

35. Schwamm LH, LaBresh KA, Pan WQ, et al. Get With The Guidelines – Stroke improves the rate of "defect-free" acute stroke care. Stroke. 2006;37(2):659.

36. NIH Common Data Elements (CDE) Repository. National Library of Medicine. National Institutes of Health. https://cde.nlm.nih.gov/. Accessed June 4, 2019.

37. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative. http://www.comet-initiative.org/. Accessed June 4, 2019.

38. Gliklich RE, Castro M, Leavy MB, et al. Harmonized outcome measures for use in asthma patient registries and clinical practice. J Allergy Clin Immunol. 2019. PMID: 30857981. DOI: 10.1016/j.jaci.2019.02.025.

39. Calkins H, Gliklich RE, Leavy MB, et al. Harmonized outcome measures for use in atrial fibrillation patient registries and clinical practice: Endorsed by the Heart Rhythm Society Board of Trustees. Heart Rhythm. 2019;16(1):e3-e16. PMID: 30449519. DOI: 10.1016/j.hrthm.2018.09.021.

40. Barranger JA, O'Rourke E. Lessons learned from the development of enzyme therapy for Gaucher disease. J Inherit Metab Dis. 2001;24 Suppl 2:89-96; discussion 87-8. PMID: 11758684.

41. Tan MH, Thomas M, MacEachern MP. Using registries to recruit subjects for clinical trials. Contemp Clin Trials. 2015;41:31-8. PMID: 25545027. DOI: 10.1016/j.cct.2014.12.012.

42. Wennberg DE, Lucas FL, Birkmeyer JD, et al. Variation in carotid endarterectomy mortality in the Medicare population: trial hospitals, volume, and patient characteristics. JAMA. 1998;279(16):1278-81. PMID: 9565008.

43. MacIntyre K, Capewell S, Stewart S, et al. Evidence of improving prognosis in heart failure: trends in case fatality in 66 547 patients hospitalized between 1986 and 1995. Circulation. 2000;102(10):1126-31. PMID: 10973841.

44. Konstam MA. Progress in heart failure Management? Lessons from the real world. Circulation. 2000;102(10):1076-8. PMID: 10973832.

45. Hutchins LF, Unger JM, Crowley JJ, et al. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med. 1999;341(27):2061-7. PMID: 10615079. DOI: 10.1056/NEJM199912303412706.

46. Opthalmic devices and clinical epidemiology. In: Brown SL, Bright RA, Tavris DR, editors. Medical device epidemiology and surveillance: John Wiley & Sons, Ltd.; 2007.

47. Lipscomb J, Yabroff KR, Brown ML, et al. Health care costing: data, methods, current applications. Med Care. 2009;47(7 Suppl 1):S1-6. PMID: 19536004. DOI: 10.1097/MLR.0b013e3181a7e401.

48. Eichler HG, Kong SX, Gerth WC, et al. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value Health. 2004;7(5):518-28. PMID: 15367247. DOI: 10.1111/j.1524-4733.2004.75003.x.

49. Palmer AJ. Health economics--what the nephrologist should know. Nephrol Dial Transplant. 2005;20(6):1038-41. PMID: 15840678. DOI: 10.1093/ndt/gfh824.

50. Patient-Centered Outcomes Research Institute. Patient-Centered Outcomes Research. https://www.pcori.org/research-results/patient-centered-outcomes-research. Accessed June 4, 2019.

51. Patient-Centered Outcomes Research Institute. PCORI Methodology Standards. April 2018. https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf. Accessed June 4, 2019.

52. Li Q, Glynn RJ, Dreyer NA, et al. Validity of claims-based definitions of left ventricular systolic dysfunction in Medicare patients. Pharmacoepidemiol Drug Saf. 2011;20(7):700-8. PMID: 21608070. DOI: 10.1002/pds.2146.

53. U.S. Food and Drug Administration. National Evaluation System for Health Technology (NEST). https://nestcc.org/. Accessed June 4, 2019.

54. Hodgson DC, Fuchs CS, Ayanian JZ. Impact of patient and provider characteristics on the treatment and outcomes of colorectal cancer. J Natl Cancer Inst. 2001;93(7):501-15. PMID: 11287444. DOI: 10.1093/jnci/93.7.501.

55. Reeves MJ, Fonarow GC, Zhao X, et al. Quality of care in women with ischemic stroke in the GWTG program. Stroke. 2009;40(4):1127-33. PMID: 19211482. DOI: 10.1161/STROKEAHA.108.543157.

56. Fonarow GC, Abraham WT, Albert NM, et al. Influence of a performance-improvement initiative on quality of care for patients hospitalized with heart failure: results of the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients With Heart Failure (OPTIMIZE-HF). Arch Intern Med. 2007;167(14):1493-502. PMID: 17646603. DOI: 10.1001/archinte.167.14.1493.

57. Greene FL, Gilkerson S, Tedder P, et al. The role of the hospital registry in achieving outcome benchmarks in cancer care. J Surg Oncol. 2009;99(8):497-9. PMID: 19466739. DOI: 10.1002/jso.21186.

58. Schweikert B, Hunger M, Meisinger C, et al. Quality of life several years after myocardial infarction: comparing the MONICA/KORA registry to the general population. European heart journal. 2009;30(4):436-43. PMID: 19019995. DOI: 10.1093/eurheartj/ehn509.

59. Griffiths I, Silman A, Symmons D, et al. BSR Biologics Registry. Rheumatology (Oxford). 2004;43(12):1463-4. PMID: 15328420. DOI: 10.1093/rheumatology/keh356.

60. U.S. Food and Drug Administration. Postapproval Pregnancy Safety Studies. DRAFT Guidance for Industry. May 2019. https://www.fda.gov/media/124746/download. Accessed June 4, 2019.

61. Porter ME. What is value in health care? N Engl J Med. 2010;363(26):2477-81. PMID: 21142528. DOI: 10.1056/NEJMp1011024.

62. Connock M, Burls A, Frew E, et al. The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review. Health Technol Assess. 2006;10(24):iii-iv, ix-136. PMID: 16796930.

63. Connock M, Juarez-Garcia A, Frew E, et al. A systematic review of the clinical effectiveness and cost-effectiveness of enzyme replacement therapies for Fabry's disease and mucopolysaccharidosis type 1. Health Technol Assess. 2006;10(20):iii-iv, ix-113. PMID: 16729919.

64. Chalkidou K, Tunis S, Lopert R, et al. Comparative effectiveness research and evidence-based health policy: experience from four countries. Milbank Q. 2009;87(2):339-67. PMID: 19523121. DOI: 10.1111/j.1468-0009.2009.00560.x.

65. Parikh RB, Kakad M, Bates DW. Integrating Predictive Analytics Into High-Value Care: The Dawn of Precision Delivery. JAMA. 2016;315(7):651-2. PMID: 26881365. DOI: 10.1001/jama.2015.19417.

66. Graves SE, Davidson D, Ingerson L, et al. The Australian Orthopaedic Association National Joint Replacement Registry. Med J Aust. 2004;180(5 Suppl):S31-4. PMID: 14984361.

67. Owen A, Spinks J, Meehan A, et al. A new model to evaluate the long-term cost effectiveness of orphan and highly specialised drugs following listing on the Australian Pharmaceutical Benefits Scheme: the Bosentan Patient Registry. J Med Econ. 2008;11(2):235-43. PMID: 19450082. DOI: 10.3111/13696990802034525.

68. Matchar DB, Oddone EZ, McCrory DC, et al. Influence of projected complication rates on estimated appropriate use rates for carotid endarterectomy. Appropriateness Project Investigators. Academic Medical Center Consortium. Health Serv Res. 1997;32(3):325-42. PMID: 9240284.

69. National Academies of Sciences, Engineering, and Medicine. Health and Medicine Division. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. September 2012. http://www.nationalacademies.org/hmd/Reports/2012/Best-Care-at-Lower-Cost-The-Path-to-Continuously-Learning-Health-Care-in-America.aspx. Accessed June 4, 2019.

Chapter 2. Planning a Registry

1. Introduction

There is tremendous variability in size, scope, and resource requirements for registries. Registries may be large or small in terms of numbers of patients or participating sites. They may target rare or common conditions and exposures. They may require the collection of limited or extensive amounts of data, operate for short or long periods of time, and be funded generously or operate with limited financial support. In addition, the scope and focus of a registry may be adapted over time to reach broader or different populations, assimilate additional data, focus on or expand to different geographical regions, or address new research questions. While this degree of flexibility confers enormous potential, registries require good planning to be successful.

When planning a registry, it is desirable to follow these initial steps: (1) articulate the purpose of the registry; (2) determine if a registry is an appropriate means to achieve the purpose; (3) identify key stakeholders; and (4) assess the feasibility and sustainability of a registry.

Once a decision is made to proceed, the next considerations in planning are to (5) build a registry team; (6) establish a governance and oversight plan; (7) define the scope and rigor needed; (8) define the dataset, patient outcomes, and target population; (9) develop a study plan or protocol; and (10) develop a project plan. Of course, the planning for a registry is often not a linear process. Many of the steps described in this chapter occur in parallel.

Registry planners should also recognize the importance of periodic critical evaluations of the registry by key stakeholders to ensure that the objectives are being met. This is particularly important for patient registries that collect data over many years. When registry objectives are no longer being met or when clinical or other changes affect the registry (e.g., changes in treatment practices, the introduction of a new therapy), the registry may need to be adapted, or the registry may stop collecting new data.

Useful resources for registry planners include the Guidelines for Good Pharmacoepidemiology Practice from the International Society of Pharmacoepidemiology;[1] the Updated Guidelines for Evaluating Public Health Surveillance Systems (especially the appendixes, which provide various checklists);[2] A Guide to the Project Management Body of Knowledge (PMBOK® Guide);[3] the Patient-Centered Outcomes Research Institute (PCORI) Methodology Standards for patient-centered outcomes research;[4] and the Good ReseArch for Comparative Effectiveness (GRACE) principles for comparative effectiveness research.[5]

2. Steps in Planning a Registry

2.1 Articulate the Registry's Purpose

One of the first steps in planning a registry is articulating its purpose. Having a clearly defined goal and/or purpose and supporting rationale makes it easier to evaluate whether a registry is the right approach for capturing the information of interest.[6,7] In addition, a clearly defined purpose

helps clarify the need for certain data. Conversely, having a clear sense of how the data may be used will help refine the stated purpose.

A registry may have a singular purpose or several purposes.[8] In either case, the overall purpose should be translated into specific objectives or questions to be addressed through the registry. Note, even if a registry is primarily derived from electronic health record (EHR) data, the purpose(s) of the registry will drive many of the decisions made in data acquisition, normalization, cohort identification and so forth. This process needs to consider the interests of those collaborating in the registry and the key audiences to be reached.[9] Clear objectives are essential to define the structure and process of data collection and to ensure that the registry effectively addresses the important questions through the appropriate outcomes analyses. Specific objectives also help the registry to avoid collecting large amounts of data of limited value. The time and resources needed to collect and process data from a registry can be substantial.[10] Attempts to be all inclusive may add cost but not value, resulting in overly burdensome data collection that can reduce quality and erode compliance. Thus, the identification of a core dataset is essential. The benefits of any data element included in the registry must outweigh the costs of including it.

Registry planners should establish specific objectives by considering what key questions the registry needs to answer. These questions will determine the type of registry (e.g., whether single focus or comparative), the data elements to be captured, and the types of analysis to be undertaken. Examples of key or driving questions are listed below:

- What is the natural course of a disease, and how does geographic location affect the course?
- Does a treatment lead to long-term benefits or harm, including delayed complications?
- How is disease progression affected by available therapies?
- What are significant predictors of poor outcomes?
- What is the safety profile of a specific therapy?
- Is a specific product or therapy teratogenic?
- How do clinical practices vary, and what are the best predictors of treatment practices?
- Are there disparities in the delivery and/or outcomes of care?
- What characteristics or practices enhance adherence?
- How do quality improvement programs affect patient outcomes?
- What process and outcomes metrics should be incorporated to track quality of patient care?
- Should a procedure or product be a covered benefit in a particular population?
- Was an intervention program or risk-management activity successful?
- What are the resources used/economic parameters of actual use in typical patients?

2.2 Determine if a Registry Is an Appropriate Means To Achieve the Purpose

Two key questions to consider are whether a registry (or other study) is needed to address the purpose and, if the answer is yes, whether prospective data collection through a registry is an appropriate means of accomplishing the scientific objectives. Every registry developer should consider the following questions early in the planning process:

- Do these data already exist?
- If so, are they of sufficient quality to answer the research question?
- Are they accessible, or does an entirely new data collection effort need to be initiated?

For example, could the necessary data be extracted from EHRs or administrative health insurance claims data? In such cases, registries might avoid re-collecting data that have already been collected elsewhere and are accessible. Thought should be given to adapting an existing registry and/or linking to other relevant data sources (including "piggybacking" onto other registries). Literature searches and searches of research databases, such as ClinicalTrials.gov, are helpful for identifying existing studies that may be useful sources of data. When the required data have not been sufficiently collected or are not accessible for the desired purpose, it is appropriate to consider creating a new registry.

The next step is to consider whether the purpose would be well served by a registry. When making this decision, it is important to fully define the specific research question(s) of interest and to consider the state of current knowledge and gaps in evidence. Other factors that may influence this decision include the breadth of the target population of interest, the length of an observational period needed to achieve the objective, the variety and complexity of treatments used, the approximate amount of funding available to address these objectives, and the urgency of decisions that will be made based on the resulting evidence. Registries may be the most appropriate choice for some research questions. For example, registries are particularly useful in situations where a comprehensive, flexible research design is needed,[11,12] or when the purpose is to discover how a product works in a wide variety of subgroups. (See Chapter 3 for a discussion of research questions appropriate for registries.) In some cases, a hybrid approach, such as a registry that incorporates data collected retrospectively as well as prospectively, will be required.

A research strategy, as opposed to a single study, may be necessary to address some research questions. For example, some research questions may require an interventional approach to address concerns about efficacy combined with an observational approach to examine long-term outcomes and quality of life in a broad patient population. When making a decision about study design, it is important to select the approach or combination of approaches best able to answer the specific research questions, from both scientific and practical standpoints. A careful evaluation of the possibilities for data collection and registry design, the degree of certainty required, and the timeframe in which this certainty is expected can help in selecting an appropriate study design.

Historically, there has been a lack of consensus standards for conducting and reporting methods and results for registries. Therefore, registries have been more variable in implementation and more difficult to assess for quality than randomized controlled trials. In recent years, advances in epidemiological and biostatistical methods have broadened the scope of questions that can be

addressed through observational studies such as registries. Stratification, propensity score matching, and risk adjustment are increasingly useful approaches for addressing confounding issues and for creating comparably homogeneous subgroups for analysis within registry datasets, and advances in bias analysis are being used to help interpret results from observational studies such as registries.[13-15] (See Chapters 3 and 13.) These techniques may allow registries to be used to support investigations of comparative safety and effectiveness. Following good registry practices, as described in this User's Guide, can strengthen scientific rigor (Chapter 14).

2.3 Identify Key Stakeholders

As a means of identifying potential stakeholders, it is important to consider to whom the research questions matter. It is useful to identify these stakeholders at an early stage of the registry planning process, as they may have important input into the type and scope of data to be collected, they may ultimately be users of the data, and/or they may have a key role in disseminating the results of the registry.

One or more parties could be considered stakeholders of the registry. These parties could be as specific as a regulatory agency that may be assessing its potential use in collecting relevant premarket data for regulatory submissions or in monitoring postmarketing studies or as broad as the general population, or simply those patients with the conditions of interest. Often, a stakeholder's input directly influences whether development of a registry can proceed, and it can have a strong influence on how a registry is conducted. A regulatory agency looking for management of a therapeutic product with a known toxicity profile may require a different registry design than a manufacturer with general questions about how a product is being used.

Typically, there are primary and secondary stakeholders for any registry. A primary stakeholder is usually responsible for creating and funding the registry. The party that requires the data, such as a regulatory authority, may also be considered a primary stakeholder. A secondary stakeholder is a party that would benefit from knowledge of the data or that would be impacted by the results but is not critical to establishing the registry. Treating clinicians and their patients could be considered secondary stakeholders. A partial list of possible stakeholders, both primary and secondary, follows:

- Public health or regulatory authorities
- Product manufacturers
- Healthcare service providers
- Payer or commissioning authorities
- Patients/caregivers
- Patient advocacy groups
- Treating clinician groups
- Academic institutions or consortia
- Professional societies
- Funding agencies

Although interactions with potential stakeholders will vary, the registry will be best supported by defined interactions and communications with these parties. Defining these interactions during the planning stage will ensure that adequate dialog occurs and appropriate input is received to support the overall value of the registry. Interactions throughout the entire duration of the registry can also assure stakeholders that the registry is aligned with the purposes and goals that were set out during the planning stages and that the registry complies with all required guidance documents, rules, and/or regulations.

Engagement with patient stakeholders is an increasing area of interest for some registries. The concept of patient-centered research has gained attention in recent years, most notably with the establishment of the Patient-Centered Outcomes Research Institute (PCORI) in 2010. PCORI is an independent, nonprofit organization that funds patient-centered comparative effectiveness research. As a condition of funding, all PCORI awardees must actively engage patients and other stakeholders in all phases of a research project. In addition to the work of PCORI, many other efforts are encouraging the active participation of patients, their caregivers, and patient advocates in clinical research and regulatory decision making. For instance, there has been considerable interest on the part of some regulatory agencies in adopting "adaptive approaches" for drug approval and reimbursement decisions and in factoring patient preferences in weighing benefits and risks of medical devices. These approaches hinge on the early and continuous engagement of patients and other key stakeholders throughout the life-span of drug development.[16,17] The U.S. Food and Drug Administration (FDA) has taken several steps to formalize the integration of patient perspectives into the regulatory process, including establishing a Patient Engagement Collaborative.[18] At the same time, an increasing number of well-organized, vocal patient advocacy groups are actively contributing to the research landscape by funding grants and facilitating collaborations across academic sites; serving as clinical trial recruitment partners, particularly for rare conditions; and enhancing capacity to develop biomarkers or other clinical screening and monitoring tests for therapeutic products.[19]

As a result of these initiatives and other efforts, investigators are increasingly integrating a patient-centered approach into their clinical research and seeking patient perspectives on how a registry can meet patients' needs and expectations, how investigators can best engage with patients, how to best collect the required data, and how the registry could provide additional value to patients beyond data collection. More information on engaging patient partners in the design and conduct of patient registries can be found in 21st Century Registries, an eBook addendum to the User's Guide.[20]

2.4 Assess Feasibility

A key element in determining the feasibility of developing a new registry relates to funding. Registries that meet the attributes described in this User's Guide will most likely require significant funding. The degree of expense incurred will be determined by the scope of the registry, the rigor of data collection, and any audits that may be required. Traditionally, the cost of the registry was driven largely by the number of sites, the number of patients, and the scope of data collected. When using data from secondary sources, the primary cost drivers are the number of health IT systems to be integrated; the amount of effort needed to clean, standardize, and normalize the data; the need to extract data from unstructured fields; and the need to include specialized data such as images. Funding will be affected by whether the registry adapts to new

issues over time and whether multiple funding sources participate. Funding needs should also be examined in terms of the projected life of the registry and/or its long-term sustainability.

There are many potential funding sources for registries. Funding sources are likely to want to share in planning and to provide input for the many choices that need to be made in the implementation plans. Funding sources may negotiate to receive access to deidentified data as a condition for their participation. Funding models for registries may vary significantly, and there is no preferred approach. Rather, the funding model for a registry should be dictated by the needs of the registry. Potential sources of funding include:

- *Foundations:* Nonprofit disease foundations may be interested in a registry to track the natural history of the disease of interest as well as the impact of therapeutic interventions. Registries may be used to track practice patterns and outcomes for quality improvement initiatives. Ongoing registries can sometimes serve the additional purpose of assisting in recruitment for clinical trials.[21]
- *Government:* Federal agencies, such as the National Institutes of Health (NIH), Centers for Disease Control and Prevention (CDC), Centers for Medicare & Medicaid Services (CMS), Agency for Healthcare Research and Quality (AHRQ), FDA, and State agencies, may be interested in a registry to provide seed funding for early development or to determine long-term outcomes of agents, devices, groups of drugs, or procedures. While the pharmaceutical industry or device manufacturers collect most long-term data on drug and device safety, many research questions arise that could potentially be suitable for government funding, ranging from clinical or comparative effectiveness to natural history of disease to the performance of healthcare providers based on accepted measures of quality of care. To determine if an agency might be interested in funding a registry, look for Requests for Proposals (RFPs) on its website. An RFP posting or direct communication with the appropriate agency staff may provide a great deal of specific information as to how a submission will be judged and what criteria would be needed in order for a proposal to be favorably ranked. Even if an RFP is not posted, contacting the appropriate agency staff may uncover potential interest in a registry to fill an unmet need.
- *Health plan providers:* Under certain circumstances, health plan providers may be interested in funding a registry, since practical clinical research is increasingly viewed as a useful tool for providing evidence for health coverage and healthcare decisions.[22]
- *Patient groups:* Patients may be able to contribute funding to focus on rare diseases or patient subgroups of interest for more common conditions. They may also contribute value in-kind.
- *Private funding:* Private philanthropic individuals or charitable foundations and trusts may have an interest in furthering research to better understand the effects of a particular intervention or sets of interventions on a disease process.
- *Product manufacturers:* Product manufacturers may be interested in studying the natural history of the disease for which they have (or are developing) a product; demonstrating the effectiveness and/or safety of existing products in real-world use through Risk Evaluation and Mitigation Strategy (REMS) programs as part of postmarketing commitments or requirements or through studies; evaluating the effectiveness of REMS

programs or risk minimization strategies; or assisting providers in evaluating or improving quality of care.

- *Professional societies:* Healthcare professional associations are increasingly participating in developing or partnering with registries for scientific and quality measurement or improvement purposes.
- *Professional society/pharmaceutical industry "hybrids":* Situations may exist in which a product manufacturer funds a registry designed and implemented by a professional society to gain insight into a set of research questions.
- *Multiple sponsors:* Registries may meet the goals of multiple stakeholders, and such stakeholders may have an interest in sharing the funding. Registries for isotretinoin and antiretrovirals in pregnancy are examples, as is INTERMACS™, a registry for patients who are receiving mechanical circulatory support device therapy to treat advanced heart failure.[23] While multiple sponsorship can decrease the costs for each funding source, their varied interests and needs almost always increase the complexity and overall cost of the registry.

A public-private partnership is a venture that is funded and operated through a collaboration between a public agency and a private-sector organization. While some public-private partnerships for registries currently exist (e.g., State-level immunization registries, bioterrorism surveillance efforts), there is great potential for growth in this approach. Both government and private sources have shown increasing interest in registries for improved safety monitoring, for comparative effectiveness goals, and for streamlining the costs of the drug development process.[24-29] Several legislative actions have stated or suggested the role of public-private partnerships for activities such as registry development.[30] There are many good reasons for multiple stakeholders, including government agencies, providers, and industry, to work together for certain purposes. Thus, it is anticipated that shared funding mechanisms are likely to become more common. Chapter 9 provides more detail on the use of public-private partnerships to support registries.

2.5 Build a Registry Team

Several different kinds of knowledge, expertise, and skills are needed to plan and implement a registry. In a small registry run by a single individual, consultants may be able to provide the critical levels of expertise needed to plan all components of the registry. In a large registry, a variety of individuals may work together as a team to contribute the necessary expertise. Depending on the size, scope, and purpose of the registry, few, some, or all of the individuals representing the components of expertise described below may be included at the time of the planning process. Whatever number of individuals is eventually assembled, it is important to build a group that can work together as a collegial team to accomplish the goals of the registry. Additionally, the team participants must understand the data sources. By understanding the goals and data sources, the registry team will enable the data to be used in the most appropriate context for the most appropriate interpretation. The different kinds of expertise and experience that are useful include the following:

- *Project management:* Project management will be needed to coordinate the components of the registry; to manage timelines, milestones, deliverables, and budgets; and to ensure

communication with sites, stakeholders, oversight committees, and funding sources. Ongoing oversight of the entire process will require a team approach.

- *Subject matter:* A registry must be designed so that it contains the appropriate data to meet its goals as well as the needs of its stakeholders. For example, experts in the diagnosis and treatment of the clinical disease to be studied who are also familiar with the potential toxicities of the treatment(s) to be studied are critical to the success of any registry collecting that data. The population under study must be clearly defined and ascertained before subjects are included in the registry for the data to have external validity. Clinical experts must be able to apply all of the latest published clinical, toxicity, and outcome data to components of the registry and determine which elements are necessary, desirable, or superfluous. Additionally, depending on the outcomes and registry purpose, it is often useful to have patient representatives or advocates.
- *Registry science:* Epidemiology and biostatistics expertise specific to the subtleties of patient registries and observational research is very important in the design, implementation, and analysis of registry data. Epidemiologists can provide the study design and can work in collaboration with biostatisticians to develop a mutual understanding of the research objectives and data needed. Health outcomes researchers and economics researchers can also lend valuable expertise to the registry team. These scientists should work with the subject matter experts to ensure that appropriate analytic methods are being used to address the clinical issues relevant to achieving the goals of the registry.
- *Data collection and database management:* The decision to include various data elements can be made in consultation with experts in this field to place "critical fields" in a prominent and logical position on the data form for both paper-based and electronic data collection tools and to determine the most appropriate source of data that will be extracted from other sources. (A final determination of what is usable and workable for data collection should be approved by all members of the team.) These experts may also need to write specific programs so that the data received from the registry are identified, grouped, and stored appropriately. They may generate reports for individuals who track registry participation, and they may provide data downloads periodically to registry analysts. This team will also be responsible for implementing and maintaining firewalls to protect the data according to accepted levels of security for similar collections of sensitive data. Registries that incorporate secondary data sources may also require support from clinical informaticists to clean and standardize these data.
- *Legal issues/patient privacy:* It is critical that either information that identifies individual patients be excluded or applicable legal requirements for the inclusion of patient identifiable information be met (e.g., obtaining informed consent or Health Insurance Portability and Accountability Act [HIPAA] authorization, where required). The complexities of this topic are dealt with in detail in Chapters 7 and 8. Legal and privacy expertise is needed to protect the patients and the owners of the database by ensuring that the registry complies with all Federal and State laws applicable to patient information.
- *Quality assurance:* As discussed in Chapter 11, quality assurance of procedures and data is another important component of registry success. Expertise in quality assurance will

help in planning a good registry. The goals for quality assurance should be established for each registry, and the efforts made and the results achieved should be described.

2.6 Establish a Governance and Oversight Plan

Governance refers to guidance and high-level decision making, including purpose, funding, execution, and dissemination of information. A goal of proper governance and oversight should be transparency to stakeholders in operations, decision making, and reporting of results. Registries fulfill governance roles in a variety of ways depending on the purpose and size of the registry. Chapter 9 provides more information on registry governance.

2.7 Consider the Scope and Rigor Needed

2.7.1 Scope of Data

The scope of a registry may be viewed in terms of size, setting, duration, geography, and financing. The purpose and objectives of the registry should frame the scope, but other factors (aside from feasibility) may ultimately shape it. For example, the scope may be affected by:

- Regulatory requirements, such as those imposed by the FDA as a condition of product marketing.
- Reimbursement decisions, such as national coverage decisions by CMS or "Prior Authorization" requirements used by health insurers in some situations.
- National research interests, such as those driven by NIH.
- Public health policy, such policies issued by the CDC.

The scope is also affected by the degree of uncertainty that is acceptable to the primary stakeholders, with that uncertainty being principally driven by the quantity, quality, and detail of the data collection balanced against its considered value and funding. Therefore, it is critical to understand the potential questions that may or may not be answerable because of the quantity and quality of the data. It should also be noted that the broader the audience of stakeholders, the broader the list of questions that may need to be included. This increased breadth can result in an increase in the number of patients who need to be enrolled and/or data points that need to be collected in order to meet the objective of the registry with an acceptable level of precision.

Some of the specific variables that can characterize the scope of a registry include:

- *Size:* This refers to the number and complexity of data points, the frequency of data collection, and the enrollment of investigators and patients. A registry with a large number of complex data points may allow for detailed and thoughtful analyses but may be so burdensome as to discourage investigator and patient enrollments. In turn, a small registry with few patients and data points may be easier to execute, but the data could lack depth and be less meaningful.[31] Size also determines the precision with which measures of risk or risk difference can be calculated.
- *Duration:* The planning of a registry must reflect both the length of time that the registry is expected to collect the data in order to achieve its purpose and the time needed to

perform analyses of the data collected. Some registries may be time-limited by commercial interests, such as when the product under study is approaching the end of its patent life.

- *Setting:* This refers to the specific setting through which the registry will recruit investigators and patients as well as collect data (e.g., hospital, doctor's office, pharmacy, home). Some registries operate within a single setting (e.g., doctor's offices), while others operate in multiple settings (e.g., enroll patients in a doctor's office and follow them through direct-to-patient outreach or collection from or linkage to data from other organizational settings).
- *Geography:* A registry that collects data in one country is very different in scope from a registry that collects data from multiple countries, in terms of setup, management, and analysis. A multinational registry poses challenges (e.g., language, cultural, time zone, regulatory) that must be taken into consideration in the planning process.[32]
- *Cost:* The scope of a registry will determine the cost of creating and managing the registry and analyzing the data. Budgetary constraints must be carefully considered before moving from conception to reality. Additionally, the value of the information is a major factor in the financial decision making. Certain choices in planning, such as building on existing infrastructure and/or linking to data sources relevant to the purposes of the registry, may increase the net return.
- *Richness of clinical data needed:* In some situations, the outcome may be relatively simple to characterize (e.g., death). In other cases, the focus of interest may be a complex set of symptoms and measurements (e.g., for Churg-Strauss Syndrome) or may require specialized diagnostic testing or tissue sampling (e.g., sentinel node in melanoma). Some outcomes may require assessment by an independent third party. Depending on the objectives of the registry, collection and storage of biological samples may be considered. (See 'Scientific Rigor' below.) The collection of biosamples and the resulting genomic data is a rapidly evolving field, and registry developers should consult both technical and legal experts regarding how to include biosamples and genomic data in a registry.

2.7.2 When Data Need To Be Available for Analysis

Meaningful data on disease progression or other long-term patient outcomes may not be available through a registry for many years, whereas safety data could be examined periodically over time. Therefore, the type of data on patient outcomes and when they will be available for analysis should be addressed from the perspective of the intended uses of the data in both the short term and long term. For industry-sponsored registries, if planning begins at an early stage, it may be possible to consider whether to align registry questions with those from the clinical trial setting (where appropriate) so that some data can carry over for more comprehensive longitudinal analyses.

2.7.3 Scientific Rigor

The content of the data to be collected should be driven by the scientific analyses that are planned for the registry, which, in turn, are determined by the specific objectives of the registry. A registry designed primarily for monitoring safety will contain different data elements from one

designed primarily for monitoring effectiveness. Similarly, the extent to which data need to be validated will depend on the purpose of the registry and the complexity of the clinical information being sought. For some outcomes, clinical diagnosis may be sufficient; for others, supporting documents from hospitalizations, imaging studies, referrals, or biopsies may be needed; and for others, formal adjudication by a committee may be required. Generally, registries that are undertaken for regulatory decision making will require increased attention toward diagnostic confirmation (i.e., enhanced scientific rigor).

2.8 Define the Core Dataset, Patient Outcomes, and Target Population

2.8.1 Core Dataset

Elements of data to be included must have potential value in the context of the current scientific and clinical climate and must be chosen by a team of experts, preferably with input from experts in biostatistics, epidemiology, and the clinical area of interest. Each data element should relate to the purpose and specific objectives of the registry. Ideally, each data element should address the central questions for which the registry was designed. It is useful to consider the generalizability of the information collected, as appropriate. For example, when seeking information on cost-effectiveness, it may be preferable to collect data on resource utilization rather than actual costs of this utilization, since the broader descriptor can be more easily generalized to other settings and cost structures. While a certain number of "speculative" fields may be desired to generate and explore hypotheses, these must be balanced against the burden of capturing superfluous data. A plan for quality assurance should be considered in tandem with developing the core dataset.

The core dataset variables ("need to know") define the information set needed to address the critical questions for which the registry was created. Determination of the core dataset variables should be guided by a clear statement of the primary research questions being posed and the analytic plans for addressing those questions. At a minimum, registry planners must account for these fields when calculating the resource needs and overall design of the registry. If additional noncore ("nice to know") variables, such as more descriptive or exploratory variables, are included, it is important that such data elements align with the goals of the registry and take into account the burden of data collection and entry at the site level or, in the case of secondary data sources, the resources required to obtain and process these data. A parsimonious use of "nice to know" variables is important for several reasons.

First, when data elements change, there is a cascade effect on all dependent components of the registry process and outputs. For example, the addition of new data elements may require changes to the data collection system, retraining of site personnel on data definitions and collection practices, adjustments to the registry protocol, and amendment submissions to institutional review boards. Such changes often require additional financial resources. Ideally, the registry would both limit the total number of data elements and include, at the outset, data elements that might change from "nice to know" to "need to know" during the course of the registry. In practice, this is a difficult balance to achieve, so most registries should plan adequate resources to be used for change management.

Second, a registry should avoid attempting to accomplish too many goals, or its burden will outweigh its usefulness to the clinical sites and researchers. Examples exist, however, of

registries that serve multiple purposes successfully without overburdening clinicians. (See Case Example 1.)

Third, even "need-to-know" variables can sometimes be difficult to collect reliably (e.g., use of illegal substances) or without substantial burden (e.g., unusual laboratory tests). Even with a limited core dataset, feasibility must still be considered. (See Chapter 5.)

Fourth, it is useful to consider what data are already available and/or collected and what additional data need to be collected. When determining additional data elements, it is imperative to consider whether the information desired is consistent with general practice or whether it might be more intensive or exceeding usual practice. For some purposes, collecting specific laboratory results or additional visits may be necessary, but could change how the registry is perceived by institutional review boards or ethics committees. The distinction between "interventional" and "observational" is straightforward in terms of random assignment to treatment, but some registries with requirements that exceed a threshold of usual practice may be subject to additional requirements more typical of "interventional" research. The determination that a registry should be considered "interventional" from a regulatory perspective can add significant burden and cost to the registry program, and, therefore, the tradeoffs must be carefully considered in planning schedules for registry visits and the collection of data and/or specimens. Registries should carefully consider the potential for different interpretations of "interventional" and "observational."

Finally, it is important to consider patient privacy, national and international regulatory requirements, and ethical considerations to ensure that the registry data requirements do not jeopardize patient privacy or put institutional/ethics reviews and approvals at risk.

2.8.2 Patient Outcomes

The outcomes of greatest importance should be identified early in the concept phase of the registry. Delineating these outcomes (e.g., primary or secondary endpoints) will force registry designers to establish priorities. Prioritization of interests in the planning phase will help focus the work of the registry and will guide study size requirements. (See Chapter 3.) Identifying the patient outcomes of the greatest importance will also help to guide the selection of the dataset. Avoiding the temptation to collect "nice to know" data that are likely of marginal value is of paramount importance, yet some registries do, in fact, need to collect large amounts of data to accomplish their purposes. Possessing adequate data in order to properly address potential confounders during analysis is one reason that extensive data collection is sometimes required.[33]

Methods to ascertain the principal outcomes should be clearly established. The diagnostic requirements, level of data detail, and level of data validation and/or adjudication should also be addressed. The issues of ascertainment noted here are important to consider because they will have a bearing on some attributes by which registries may be evaluated. These attributes include sensitivity (the extent to which the methods identify all outcomes of interest) and external validity (generalizability to similar populations), among others.

As discussed in Chapter 4, relying on established guidelines and standards to aid in defining outcomes of interest has many benefits and should be considered.

2.8.3 The Role of Patient-Reported Outcomes in Registries[i]

A patient outcome, as discussed above, refers to any outcome related to the patient, whether reported by the patient or described by a third party (e.g., by an imaging report, laboratory evaluation, or clinician assessment). As part of a shift toward patient-centered care, there has been an increasing recognition of the importance of measuring and reporting those aspects of health and well-being that are best reported by patients themselves, whether related to disease, treatment, or both. The FDA defines a patient-reported outcome (PRO) as a measurement based on a report that comes directly from the patient (i.e., the study subject) about the status of a patient's health condition, without amendment or interpretation of the patient's response by a clinician or anyone else.[34] Over the past 20 years, an expanding body of literature has demonstrated that PROs are associated with traditional outcomes, such as overall survival[35-40] and tumor response.[41] PROs themselves are also increasingly recognized as valid outcomes (e.g., quality of life [QOL], pain, breathlessness, physical functioning).[42-49]

Systematic collection of PROs in clinical trials, patient registries, and usual clinical care is feasible and efficient.[50-54] PROs are more reflective of underlying health status than physician reporting[55] and facilitate discussion of important symptoms and QOL with clinicians.[56] Additionally, they have been shown to serve as supporting documentation,[51] improve symptom management,[57] and potentially impact clinical decision making.[52,58] While widespread adoption of PROs as a key component in clinical research has not occurred, there is increasing recognition of their role in complementing traditional clinical and administrative data.[59] The FDA has identified PROs as the regulatory standard for supporting subjective endpoints, like symptoms, in drug approval and labeling and has provided clear instructions on PRO measurement in drug development trials.[34] In addition, FDA has underscored the value and use of PROs in assessing the effects of medical devices in its strategic priorities.[60]

While there are no formal guidelines for inclusion of PROs in registries, PROs contribute information across the spectrum of registry purposes described in Chapter 1, and inclusion of PROs should be considered during the planning and design phases. For example, registries intended to describe the natural history of a disease must include adequate information about symptom burden and related QOL trajectories, especially in the setting of rare diseases, inherited diseases with increasing life span (e.g., cystic fibrosis, sickle cell disease), and heterogeneous diseases (e.g., chronic obstructive pulmonary disease, breast cancer). In registries designed to study effectiveness, patient-reported symptoms can be indicators of adverse consequences of therapy (e.g., toxicity monitoring), targets for meaningful intervention (e.g., symptom control intervention), and means of understanding how patient perceptions of toxicities impact

[i] *Adapted from Abernethy A, Basch E, Kulig K, et al. Use of Patient-Reported Outcomes in Registries. In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

effectiveness (e.g., through adherence behavior). PROs can also serve an important purpose in registries designed for measuring or monitoring safety and measuring quality.

Validated tools should be used to capture PROs whenever possible. Chapters 4 and 5 provides more information on identifying and selecting tools for capturing PROs. Chapter 6 discusses patient-generated data.

2.8.4 Target Population

The target population is the population to which the findings of the registry are meant to apply. It must be defined for two basic reasons. First, the target population serves as the foundation for planning the registry. Second, it also represents a major constituency that will be impacted by the results of the registry.

One of the goals for registry data may be to enable generalization of conclusions from clinical research on narrowly defined populations to broader ones, and therefore the inclusion criteria for most (although not all) registries are relatively broad. As an example, screening criteria for a registry may allow inclusion of elderly patients, patients with multiple comorbidities, patients on multiple therapies, patients who switch treatments during the period of observation, or patients who are using products "off label." The definition of the target population will depend on many factors (e.g., scope and cost), but ultimately will be driven by the purpose of the registry.

As with defining patient outcomes, target population criteria and/or definitions should be consistent with established guidelines and standards within the therapeutic area. Achieving this goal increases the potential utility of the registry by leveraging other data sources (historical or concurrent) with different information on the same target population and enhancing statistical power if similar information is collected on the target population.

In establishing target population criteria, consideration should be given to the feasibility of access to that population. One should try to distinguish the ideal from the real. Some questions to consider in this regard are:

- How common is the exposure or disease of interest?
- Can eligible people be readily identified?
- Are other sources competing for data on the same patients?
- Is care centralized or dispersed (e.g., in a referral or tertiary care facility)?
- How mobile is the target population?

Ultimately, methods to ascertain members of the target population should be carefully considered (e.g., use of screening logs that identify all potential patients and indicate whether they participate and, if not, why not), as should the use of sources outside the registry (e.g., patient groups). Greater accessibility to the target population will reap benefits in terms of enhanced representativeness and statistical power.

Lastly, thought should be given to comparison (control) groups either internal or external to the registry. Again, much of this consideration will be driven by the purpose and specific objectives

of the registry. For example, natural history registries do not need controls, but controls are especially desirable for registries created to evaluate comparative effectiveness or safety.

2.9 Develop a Study Plan or Protocol

The study plan documents the objectives of the registry and describes how those objectives will be achieved. At a minimum, the study plan should include the registry objectives, the eligibility criteria for participants, and the data collection procedures. Ideally, a full study protocol will be developed to document the objectives, design, participant inclusion/exclusion criteria, outcomes of interest, data to be collected, data collection procedures, governance procedures, and plans for complying with ethical obligations and protecting patient privacy.

In addition to a study plan or protocol, registries may have statistical analysis plans. Chapters 13 and 14 discuss the importance of analysis plans.

2.10 Develop a Project Plan

Developing an overall project plan is critically important so that the registry team has a roadmap to guide their collective efforts. Depending on the complexity of the registry project, the project plan may include some or all of the following elements:

- Scope management plan to control the scope of the project. It should provide the approach to making changes to the scope through a clearly defined change-control system.
- Detailed timeline and schedule management plan to ensure that the project and its deliverables are completed on time.
- Cost management plan for keeping project costs within the budget. The cost management plan may provide estimates on cost of labor, purchases and acquisitions, compliance with regulatory requirements, et cetera. This plan should be aligned with the change-control system so that all changes to the scope will be reflected in the cost component of the registry project.
- Quality management plan to describe the procedures to be used to test project concepts, ideas, and decisions in the process of building a registry. Having a quality management plan in place can help in detecting design errors early, formulating necessary changes to the scope, and ensuring that the final product meets stakeholders' expectations. The quality management plan should include a clear process for management and disposition of change requests.
- Staffing management plan to determine what skills will be needed and when to meet the project goals. (See Chapter 2).
- Communication plan that includes who is responsible for communicating information and to whom it should be communicated. Considerations include different categories of information, frequency of communications, and methods of communication. The plan should also provide steps to escalate issues that cannot be resolved on a lower staff level.

- Procurement plan for external components or equipment and/or outsourced software development for the planned registry, if pertinent. Such a plan should describe how the procurement process would be managed within the organization. Decisions to procure products or services may have a direct impact on other components of the project plan, including the staffing plan and timeline.
- Risk management plan to identify and mitigate risks. Many project risks are predictable events, and therefore they can and should be assessed in the very early stages of registry planning. It is important to prioritize project risks by their potential impact on the specific objectives and to develop an adequate risk response plan for the most significant risks. Some predictable risks include—
    - Disagreement between stakeholders over the scope of specific tasks.
    - Inaccurate cost estimates.
    - Delays in the timeline.
    - Poor quality or missing data for key variables.

3. Anticipating and Preparing for Change[ii]

Most, if not all registries, should undergo periodic critical evaluation by key stakeholders to ensure that the objectives are being met. When registry objectives are no longer being met or when clinical or other changes affect the registry (e.g., changes in treatment practices, introduction of a new therapy), the registry may need to be adapted or stop collecting new data.

Many factors may drive the decision to modify an existing registry. For example, a registry may need to transition to a new technology platform to remain functional for its participants, or a registry that was designed to study the natural history of a disease for which there was no effective treatment may change its purpose when a new product or therapy becomes available in the market. Other scenarios in which modifications may be necessary include changes in funding sources and stakeholders or the recognition of potential new regulatory uses (e.g., premarket expansion of indications or safety surveillance) as well as the introduction of new regulatory requirements. (See Case Examples 3 and 4.)

While the considerations for modifying a registry are similar to those for the launch of a new registry, there are several distinguishing features. First, a registry modification is facilitated by an existing registry and the collective experience of conducting that registry. The existing registry can essentially serve as the starting point for creating a prototype of the revision. The planning and design of the registry modification should also benefit from lessons learned in operating the existing version of the registry. What has worked well, and what has been problematic? What challenges have been encountered at every level, from staff entering data at the participating sites to the analyst creating reports? Indeed, one or more of these issues may be contributing factors in

[ii] *Adapted from Cole JA, Elbert A, Haas J, et al. 'Modifying and Stopping Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

the decision to proceed with the changes. Even if this is not the case, the modification provides an opportunity to address these issues. Registry modifications also present unique challenges distinct from the development of a new registry. In particular, transferring data collected in an existing registry to the revised registry (i.e., data migration) can be a complex and resource-intensive process.

Despite these differences, the steps in the execution of a major registry transition are analogous to those involved in planning a new registry. Registries that identify the need for major changes should follow the planning steps outlined in this chapter and the considerations for design and implementation outlined in subsequent chapters of this document.

Registries, particularly those undertaken without a fixed stopping point, may also need to periodically evaluate whether the registry data collection should continue. (See Case Example 2.) Chapter 14 of the third edition of the User's Guide discusses considerations related to stopping registry data collection.

4. Special Considerations

4.1 Rare Disease Registries[iii]

When planning a rare disease registry, special consideration should be given to the role of stakeholders and the possibility for the registry to evolve over time. Stakeholders in a rare disease registry may include patient advocacy groups (often multiple), regulatory agencies (especially if the registry is being developed to support future drug development and approval or to fulfill postmarketing commitments or requirements), clinicians, scientists, industry, payers, and the individuals and families affected by the disease. To date, collaboration between stakeholder groups has been critical to the progress made in rare disease research and product development, the adoption of important public policy changes in the United States and worldwide, and the promotion of patient access to treatments as they become available.[61]

Avoiding multiple competing registries in rare diseases is particularly important given the limited number of patients available to be enrolled. This magnifies the role of collaboration. The importance of patient registries in rare diseases and the need to support many organizations has also brought umbrella patient organizations (e.g., National Organization of Rare Disorders [NORD], the Genetic Alliance, EURORDIS) in as stakeholders, as these groups are charged with advising and supporting the development of registries. As with all registries, a single rare disease registry need not fulfill all goals for all potential stakeholders. However, early communication and collaboration with stakeholders can contribute to the development of a registry that provides an infrastructure to support different needs in an efficient way and eliminate barriers to scientific progress.

[iii] *Adapted from Dale DC, Groft SC, Harrison MJ, et al. 'Rare Disease Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

In developing plans for funding and oversight, registry planners should also consider the possibility for registry modifications. As understanding of the disease improves or as new treatments are developed, the scope of a rare disease registry may evolve over time, maturing from an outreach/community-building effort or a means for a basic understanding of patient and disease characteristics, to a supportive mechanism for research funding and attracting healthcare providers. (See Case Example 4.)

More information on rare disease registries can be found in Chapter 20 of the third edition of the User's Guide.

4.2 Quality Improvement Registries[iv]

Similar to rare disease registries, early engagement with stakeholders is important for the success of quality improvement (QI) registries. In QI registries, the care provider needs to be engaged and active, as the program is not simply supporting a feedback function or providing a descriptive or analytic function, but is often focused on patient and/or provider behavior change. In many QI registries, these active providers are termed "champions" and are vital for success, particularly early in development.[62] Once a registry matures, other incentives may drive participation (e.g., recognition, competition, financial rewards, regulatory requirements), but the role of the champion in the early phases cannot be overstated.

A second major difference between planning a QI registry and planning other types of registries is the funding model. QI registries use a wide variety of funding models. For example, a regional or national registry may be funded entirely by fees paid by participating providers or hospitals. Alternately, the registry may supplement participation fees with funding from professional associations, specialty societies, industry, foundations, or government agencies. Some QI registries may not charge a participation fee and may receive all of their funding from other organizations. Local QI registries that operate within a single institution may receive all of their funding from the institution or from research grants. The funding model used by a QI registry largely depends on the goals of the registry and the stakeholders in the specific disease area.

Lastly, change management is an important consideration in planning a QI registry. QI registries need to be nimble in order to adapt to two continual sources of change. First, new evidence comes forward that changes the way care should be managed, and the registry owner must make changes so that the registry is both current and relevant. Second, providers participating in registries manage what they measure, and over time, measures can be rotated in or out of the panel so that attention is focused where it is most critical to overcome a continuing treatment gap or performance deficiency. From a planning standpoint, QI registries should expect ongoing changes to the registry and plan for the resources required to support the changes.

[iv] *Adapted from Asher AL, Gliklich RE, Hernandez AF, et al. 'Quality Improvement Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

More information on quality improvement registries can be found in Chapter 22 of the third edition of the User's Guide.

5. Resources for Registries

In recent years, registry networks have formed to generate and share knowledge related to registry planning, design, and operations. A registry network is a formal community of organizations operating or using information from patient registries to measure and improve patient health outcomes. Registry networks may be general in nature or focused on specific domains. These networks provide a supportive infrastructure that organizes participants to undertake activities related to the specific goals of the network. Although registry networks may be established for different purposes, at a fundamental level they are strategically collaborative groups where organizations and individuals come together to advance their work, generate and share knowledge, and solve shared challenges. In addition to their knowledge sharing activities, registry networks may build a common infrastructure, create new knowledge, and provide a place to learn about the registries and other registry participants and contributors. Additionally, some registry networks provide access to technical infrastructure such as access to combined datasets or metadata, often facilitated by member registry adoption and use of data standards that the network has developed.

More information on registry networks can be found in *21st Century Registries*,[20] the eBook addendum to the third edition of the User's Guide.

6. Summary

In summary, planning a patient registry involves several key steps, including articulating its purpose, determining whether it is an appropriate means of addressing the research question, identifying stakeholders, defining the scope and target population, assessing feasibility, and securing funding. A registry team and advisors must be assembled to develop, coordinate, and guide the registry; these individuals should be selected based on their expertise and experience. Governance and oversight for the registry should also be addressed during the planning phase.

While registries differ tremendously in size, scope, and resource requirements, the basic elements of planning described here are relevant for most, if not all registries, and can help to support the launch and operation of a successful registry.

References for Chapter 2

1. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008;17(2):200-8. PMID: 17868186. DOI: 10.1002/pds.1471.

2. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001;50(RR-13):1-35; quiz CE1-7. PMID: 18634202.

3. L&M Policy Research, LLC, Quintiles Outcome. Registry of Patient Registries Outcome Measures Framework: Literature Review Findings and Implications. OMF Literature Review Report. (Prepared under Contract No. 290-2014-00004-C.) AHRQ Publication No. 16-EHC036-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2016. www.effectivehealthcare.ahrq.gov/reports/final/cfm.

4. Patient-Centered Outcomes Research Institute. PCORI Methodology Standards. April 2018. https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf. Accessed June 4, 2019.

5. Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. Am J Manag Care. 2010;16(6):467-71. PMID: 20560690.

6. Dreyer NA, Garner S. Registries for robust evidence. JAMA. 2009;302(7):790-1. PMID: 19690313. DOI: 10.1001/jama.2009.1092.

7. Solomon DJ, Henry RC, Hogan JG, et al. Evaluation and implementation of public health registries. Public Health Rep. 1991;106(2):142-50. PMID: 1902306.

8. Glaser SL, Clarke CA, Gomez SL, et al. Cancer surveillance research: a vital subdiscipline of cancer epidemiology. Cancer Causes Control. 2005;16(9):1009-19. PMID: 16184466. DOI: 10.1007/s10552-005-4501-2.

9. Kennedy L, Craig AM. Global registries for measuring pharmacoeconomic and quality-of-life outcomes: focus on design and data collection, analysis and interpretation. Pharmacoeconomics. 2004;22(9):551-68. PMID: 15209525. DOI: 10.2165/00019053-200422090-00001.

10. Bookman MA. Using tumor registry resources in analyzing concordance with guidelines and outcomes. Oncology (Williston Park). 2000;14(11A):104-7. PMID: 11195403.

11. Avorn J. In defense of pharmacoepidemiology--embracing the yin and yang of drug research. N Engl J Med. 2007;357(22):2219-21. PMID: 18046025. DOI: 10.1056/NEJMp0706892.

12. Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. PLoS Med. 2008;5(3):e67. PMID: 18336067. DOI: 10.1371/journal.pmed.0050067.

13. Alter DA, Venkatesh V, Chong A, et al. Evaluating the performance of the Global Registry of Acute Coronary Events risk-adjustment index across socioeconomic strata among patients discharged from the hospital after acute myocardial infarction. Am Heart J. 2006;151(2):323-31. PMID: 16442894. DOI: 10.1016/j.ahj.2005.07.013.

14. Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155(2):176-84. PMID: 11790682. DOI: 10.1093/aje/155.2.176.

15. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data: Springer; 2009.

16. Eichler HG, Baird LG, Barker R, et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. Clin Pharmacol Ther. 2015;97(3):234-46. PMID: 25669457. DOI: 10.1002/cpt.59.

17. Adaptive Pathways to Patients: Report on the Initial Experience of the Pilot Project. December 15, 2014. https://www.ema.europa.eu/en/documents/report/adaptive-pathways-patients-report-initial-experience-pilot-project_en.pdf. Accessed June 5, 2019.

18. 45 CFR 46.408(c). Requirements for permission by parents or guardians and for assent by children.

19. Dreyer NA, Blackburn SC, Mt-Isa S, et al. Direct-to-Patient Research: Piloting a New Approach to Understanding Drug Safety During Pregnancy. JMIR Public Health Surveill. 2015;1(2):e22. PMID: 27227140. DOI: 10.2196/publichealth.4939.

20. Gliklich RE, Dreyer NA, Leavy MB, Christian JB (eds). 21st Century Patient Registries. EBook addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. (Prepared by L&M Policy Research, LLC under Contract No. 290-2014- 00004-C.) AHRQ Publication No. 17(18)-EHC013-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018. www.effectivehealthcare.ahrq.gov.

21. Andersen MR, Schroeder T, Gaul M, et al. Using a population-based cancer registry for recruitment of newly diagnosed patients with ovarian cancer. Am J Clin Oncol. 2005;28(1):17-20. PMID: 15685029.

22. Tunis SR. A clinical research strategy to support shared decision making. Health Aff (Millwood). 2005;24(1):180-4. PMID: 15647229. DOI: 10.1377/hlthaff.24.1.180.

23. 45 CFR pts. 160, 164 (2010). HIPAA Security Rule.

24. Ray WA, Stein CM. Reform of drug regulation--beyond an independent drug-safety board. N Engl J Med. 2006;354(2):194-201. PMID: 16407517. DOI: 10.1056/NEJMsb053432.

25. Strom BL. How the US drug safety system should be changed. JAMA. 2006;295(17):2072-5. PMID: 16670415. DOI: 10.1001/jama.295.17.2072.

26. Okie S. Safety in numbers--monitoring risk in approved drugs. N Engl J Med. 2005;352(12):1173-6. PMID: 15788493. DOI: 10.1056/NEJMp058029.

27. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clin Med (Lond). 2008;8(6):579-88. PMID: 19149278. DOI: 10.7861/clinmedicine.8-6-579.

28. Lyratzopoulos G, Patrick H, Campbell B. Registers needed for new interventional procedures. Lancet. 2008;371(9626):1734-6. PMID: 18502284. DOI: 10.1016/S0140-6736(08)60742-4.

29. eCQI Resource Center. https://ecqi.healthit.gov/. Accessed June 10, 2019.

30. 42 U.S.C. § 299c-3(c). Dissemination of information.

31. Woodward M. Sample size determination. Epidemiology: study design and data analysis. Boca Raton (FL): Chapman & Hall/CRC Press; 2005.

32. 42 U.S.C. § 241(d). Protection of privacy of individuals who are research subjects.

33. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006;354(4):353-65. PMID: 16436767. DOI: 10.1056/NEJMoa051379.

34. Zubenko GS, Sommer BR, Cohen BM. On the Marketing and Use of Pharmacogenetic Tests for Psychiatric Treatment. JAMA Psychiatry. 2018;75(8):769-70. PMID: 29799933. DOI: 10.1001/jamapsychiatry.2018.0834.

35. Norekval TM, Fridlund B, Rokne B, et al. Patient-reported outcomes as predictors of 10-year survival in women after acute myocardial infarction. Health Qual Life Outcomes. 2010;8:140. PMID: 21108810. DOI: 10.1186/1477-7525-8-140.

36. Wang XS, Shi Q, Lu C, et al. Prognostic value of symptom burden for overall survival in patients receiving chemotherapy for advanced nonsmall cell lung cancer. Cancer. 2010;116(1):137-45. PMID: 19852033. DOI: 10.1002/cncr.24703.

37. Singh JA, Nelson DB, Fink HA, et al. Health-related quality of life predicts future health care utilization and mortality in veterans with self-reported physician-diagnosed arthritis: the veterans arthritis quality of life study. Semin Arthritis Rheum. 2005;34(5):755-65. PMID: 15846592. DOI: 10.1016/j.semarthrit.2004.08.001.

38. Cunningham WE, Crystal S, Bozzette S, et al. The association of health-related quality of life with survival among persons with HIV infection in the United States. J Gen Intern Med. 2005;20(1):21-7. PMID: 15693923. DOI: 10.1111/j.1525-1497.2005.30402.x.

39. Mapes DL, Lopes AA, Satayathum S, et al. Health-related quality of life as a predictor of mortality and hospitalization: the Dialysis Outcomes and Practice Patterns Study (DOPPS). Kidney Int. 2003;64(1):339-49. PMID: 12787427. DOI: 10.1046/j.1523-1755.2003.00072.x.

40. Gotay CC, Kawamoto CT, Bottomley A, et al. The prognostic significance of patient-reported outcomes in cancer clinical trials. J Clin Oncol. 2008;26(8):1355-63. PMID: 18227528. DOI: 10.1200/JCO.2007.13.3439.

41. Victorson D, Soni M, Cella D. Metaanalysis of the correlation between radiographic tumor response and patient-reported outcomes. Cancer. 2006;106(3):494-504. PMID: 16353212. DOI: 10.1002/cncr.21637.

42. Castellano D, del Muro XG, Perez-Gracia JL, et al. Patient-reported outcomes in a phase III, randomized study of sunitinib versus interferon-{alpha} as first-line systemic therapy for patients with metastatic renal cell carcinoma in a European population. Ann Oncol. 2009;20(11):1803-12. PMID: 19549706. DOI: 10.1093/annonc/mdp067.

43. Rolfson O, Karrholm J, Dahlberg LE, et al. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011;93(7):867-75. PMID: 21705555. DOI: 10.1302/0301-620X.93B7.25737.

44. Thornburg CD, Calatroni A, Panepinto JA. Differences in health-related quality of life in children with sickle cell disease receiving hydroxyurea. J Pediatr Hematol Oncol. 2011;33(4):251-4. PMID: 21516020. DOI: 10.1097/MPH.0b013e3182114c54.

45. Molnar-Varga M, Molnar MZ, Szeifert L, et al. Health-related quality of life and clinical outcomes in kidney transplant recipients. Am J Kidney Dis. 2011;58(3):444-52. PMID: 21658828. DOI: 10.1053/j.ajkd.2011.03.028.

46. Pronzato P, Cortesi E, van der Rijt CC, et al. Epoetin alfa improves anemia and anemia-related, patient-reported outcomes in patients with breast cancer receiving myelotoxic chemotherapy: results of a European, multicenter, randomized, controlled trial. Oncologist. 2010;15(9):935-43. PMID: 20798194. DOI: 10.1634/theoncologist.2009-0279.

47. Abernethy AP, McDonald CF, Frith PA, et al. Effect of palliative oxygen versus room air in relief of breathlessness in patients with refractory dyspnoea: a double-blind, randomised controlled trial. Lancet. 2010;376(9743):784-93. PMID: 20816546. DOI: 10.1016/S0140-6736(10)61115-4.

48. Deshpande AD, Sefko JA, Jeffe DB, et al. The association between chronic disease burden and quality of life among breast cancer survivors in Missouri. Breast Cancer Res Treat. 2011;129(3):877-86. PMID: 21519836. DOI: 10.1007/s10549-011-1525-z.

49. Sprangers MA. Disregarding clinical trial-based patient-reported outcomes is unwarranted: Five advances to substantiate the scientific stringency of quality-of-life measurement. Acta Oncol. 2010;49(2):155-63. PMID: 20059312. DOI: 10.3109/02841860903440288.

50. Basch E, Iasonos A, Barz A, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol. 2007;25(34):5374-80. PMID: 18048818. DOI: 10.1200/JCO.2007.11.2243.

51. Abernethy AP, Herndon JE, 2nd, Wheeler JL, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. J Pain Symptom Manage. 2009;37(6):1027-38. PMID: 19394793. DOI: 10.1016/j.jpainsymman.2008.07.011.

52. Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. J Clin Oncol. 2005;23(15):3552-61. PMID: 15908666. DOI: 10.1200/JCO.2005.04.275.

53. Snyder CF, Jensen R, Courtin SO, et al. PatientViewpoint: a website for patient-reported outcomes assessment. Qual Life Res. 2009;18(7):793-800. PMID: 19544089. DOI: 10.1007/s11136-009-9497-8.

54. Dudgeon DJ, Knott C, Chapman C, et al. Development, implementation, and process evaluation of a regional palliative care quality improvement project. J Pain Symptom Manage. 2009;38(4):483-95. PMID: 19699607. DOI: 10.1016/j.jpainsymman.2008.12.006.

55. Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst. 2009;101(23):1624-32. PMID: 19920223. DOI: 10.1093/jnci/djp386.

56. Berry DL, Blumenstein BA, Halpenny B, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. J Clin Oncol. 2011;29(8):1029-35. PMID: 21282548. DOI: 10.1200/JCO.2010.30.3909.

57. Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol. 2004;22(4):714-24. PMID: 14966096. DOI: 10.1200/JCO.2004.06.078.

58. Detmar SB, Muller MJ, Schornagel JH, et al. Role of health-related quality of life in palliative chemotherapy treatment decisions. J Clin Oncol. 2002;20(4):1056-62. PMID: 11844830. DOI: 10.1200/JCO.2002.20.4.1056.

59. New York City Health Department. The NYC Macroscope. http://www1.nyc.gov/site/doh/data/health-tools/nycmacroscope.page. Accessed March 1, 2017.

60. 42 U.S.C. §§ 299b-21-26. Patient Safety Improvement.

61. Dunkle M, Pines W, Saltonstall PL. Advocacy groups and their role in rare diseases research. Adv Exp Med Biol. 2010;686:515-25. PMID: 20824463. DOI: 10.1007/978-90-481-9485-8_28.

62. Raval MV, Bentrem DJ, Eskandari MK, et al. The role of Surgical Champions in the American College of Surgeons National Surgical Quality Improvement Program--a national survey. J Surg Res. 2011;166(1):e15-25. PMID: 21176914. DOI: 10.1016/j.jss.2010.10.036.

Case Examples for Chapter 2

Case Example 1. Creating a Registry To Fulfill Multiple Purposes and Using a Publications Committee To Review Data Requests

Description	The National Registry of Myocardial Infarction (NRMI) collected, analyzed, and disseminated data on patients experiencing acute myocardial infarction. Its goal was improvement of patient care at individual hospitals through the hospital team's evaluation of data and assessment of care delivery systems.
Sponsor	Genentech, Inc.
Year Started	1990
Year Ended	2006
No. of Sites	451 hospitals in the final phase of NRMI (NRMI 5). Over 2,150 hospitals participated in NRMI over 16 years.
No. of Patients	2,515,106

Challenge

Over the past 20 years, there have been significant changes in the treatment of acute myocardial infarction (AMI) patients. Evidence from large clinical trials has led to the introduction of new guidelines and therapies for treating AMI patients, including fibrinolytic therapy and percutaneous coronary intervention. While these treatments can improve both morbidity and mortality for AMI patients, they are time sensitive and must be administered very soon after hospital arrival in order to be most effective.

After the release of its first fibrinolytic therapy product in 1987, the sponsor's field representatives learned from their discussions with emergency department physicians, cardiologists, and hospital staff that most clinicians believed they were treating patients quickly, although there was no documentation or benchmarking to confirm this assumption or to identify and correct delays. At that time, many emergency departments did not have readily available diagnostic tools (such as angiography labs), and hospitals with AMI-specific decision pathways and treatment protocols were the exception rather than the rule.

In addition, since fibrinolytic therapy was being widely used for the first time, the sponsor wanted to gather safety information related to its use in real-world situations and in a broader range of patients than those treated in the controlled environment of a clinical trial.

Proposed Solution

The sponsor decided to create the registry to fulfill the multiple purposes of identifying treatment patterns, promoting time-to-treatment and other quality improvements, and gathering real-world safety data. The scope of the data collection necessary to meet these needs could have made such a registry impracticable, so the project team faced the sizable challenge of balancing the data needs with the feasibility of the registry.

The sponsor formed a scientific advisory board with members representing the various clinical stakeholders (emergency department, cardiology, nursing, research, etc.). The scientific advisory board developed the dataset for the registry, keeping a few guiding principles in mind. These principles emphasized maintaining balance between the clinical research and the feasibility of the registry. The first principle was to determine whether the proposed data element was necessary by asking several key questions: How will the data element be used in generating hospital feedback reports or research analyses? Is the data element already collected? If not, should it be collected? If it should be collected, is it feasible to collect those data? The second principle focused on using existing data standards whenever possible. If a data standard did not exist, the team tried to collect the data in the simplest possible way. The third principle emphasized data consistency and making the registry user-friendly by continually refining data element definitions until they were as clear as possible.

In 1990, the sponsor launched the registry. During the 16 years that the registry was conducted, it demonstrated that the advisory board's efforts to create a feasible multipurpose registry were successful. The registry collected data on the clinical presentation, treatment, and outcomes of over 2.5 million patients with AMI from more than 2,150 participating sites.

The success of the registry presented a new challenge for the registry team. The sponsor received a large volume of requests to analyze the registry data, often for research topics that fell outside of the standardized reports developed for the registry. As a guiding principle, the registry team was committed to making the data available for research projects, but it had limited resources. To support these requests, the team developed a process that would allow outside researchers to access the registry data without overburdening the registry team.

The registry team created a publication process to determine when another group could use the data for research. The team set high-level criteria for all data requests: the analysis had to be feasible given the data in the registry, and the request could not represent a duplication of another research effort.

The registry team involved its scientific advisory board, made up of cardiologists, emergency department physicians, nurses, research scientists, pharmacists, and reviewers with specialties in biostatistics and statistical programming, in creating a publication review committee. The review committee evaluated all research proposals to determine originality, interest to peers, feasibility, appropriateness, and priority. The review committee limited its review of research proposals to a set number of reviews per year, and scheduled the reviews and deadlines around the abstract deadlines for the major cardiology conferences. Research analyses had to be intended to result in peer-reviewed presentations and publications. Researchers were asked to submit proposals that included well defined questions and an analysis plan. If the proposal was accepted, the researchers discussed any further details with the biostatisticians and statistical programmers who performed the analyses (and who were employed at an independent clinical research organization). The results were sent directly to the researchers.

The scientific advisory board and review committee remained involved in the process after a data request had been granted. All authors submitted their abstracts to the review committee before sending them to conferences. The review committee offered constructive criticism to help the

authors improve their abstracts. The review committee also reviewed manuscripts before journal submission to help identify any issues or concerns that the authors should address.

Results

This publication process enabled the wealth of data collected in this registry to be used in over 150 scientific abstracts and 100 peer-reviewed articles, addressing each of the purposes of the registry as well as other research topics. By involving the scientific advisory board and providing independent biostatistical support, the registry team developed an infrastructure that enhanced the credibility of the research uses of this observational database.

Key Point

Registries can be developed to fulfill more than one purpose, but this added complexity requires careful planning to ensure that the final registry data collection burden and procedures are feasible. Making sure that the advisory board includes representatives with clinical and operational perspectives can help the board to maintain its focus on feasibility. As a registry database gains large amounts of data, the registry team will likely receive research proposals from groups interested in using the data. The registry team may want to set up a publication process during the registry design phase.

For More Information

- Califf RM. The benefits of moving quality to a national level. Am Heart J. 2008;156(6):1019–22. PMID: 19032994. DOI: 10.1016/j.ahj.2008.09.005.
- Rogers WJ, Frederick PD, Stoehr E, et al. NRMI Investigators. Trends in presenting characteristics and hospital mortality among patients with ST elevation and non-ST elevation myocardial infarction in the NRMI from 1990 to 2006. Am Heart J. 2008;156(6):1026–34. PMID: 19032996. DOI: 10.1016/j.ahj.2008.07.030.
- Gibson CM, Pride YB, Frederick PD, et al. NRMI Investigators. Trends in reperfusion strategies, door-to-needle and door-to-balloon times, and in-hospital mortality among patients with ST-segment elevation myocardial infarction enrolled in the NRMI from 1990 to 2006. Am Heart J. 2008;156(6):1035–44. PMID: 19032997. DOI: 10.1016/j.ahj.2008.07.029.
- Peterson ED, Shah BR, Parsons L, et al. NRMI Investigators. Trends in quality of care for patients with acute myocardial infarction in the NRMI from 1990 to 2006. Am Heart J. 2008;156(6):1045–55. PMID: 19032998. DOI: 10.1016/j.ahj.2008.07.028.

Case Example 2. Determining When To Stop an Open-Ended Registry

Description	The Bupropion Pregnancy Registry was an observational exposure-registration and followup study to monitor prenatal exposure to bupropion and detect any major teratogenic effect.
Sponsor	GlaxoSmithKline
Year Started	1997
Year Ended	2008
No. of Sites	Not applicable
No. of Patients	1,597

Challenge

Bupropion, an antidepressant with the potential for prenatal exposure, was labeled with a pregnancy category C by the U.S. Food and Drug Administration (FDA) due to prior animal data. The manufacturer established a prospective pregnancy registry to monitor pregnancy exposures to bupropion for any potential increased risk of congenital anomalies. Because the purpose of the registry was postmarketing safety surveillance, the duration of the registry was open ended. The registry had collected data on more than 1,500 exposed pregnant women over 10 years when a potential signal suggestive of a bupropion-related increase in cardiovascular birth defects emerged.

Proposed Solution

The advisory committee reviewed the registry data to assess the potential signal. However, due to the potential bias from the large percentage of cases lost to followup (35.8%), retrospective reports, and incomplete descriptions of the reported cardiovascular defects, it was not possible to determine the credibility of the potential signal using registry data alone. Further, the sample size was not adequate to reach definitive conclusions regarding the absolute or relative risk of any specific birth defects in women using bupropion during pregnancy (as the registry was powered only to examine the rate of birth defects overall) and was unlikely to achieve its goal as structured.

The advisory committee recommended a study to expedite the accumulation of pregnancy outcome data among women exposed to bupropion during pregnancy. In response, a large, claims-based, retrospective cohort study was conducted. This study enrolled 1,213 women exposed in the first trimester and did not confirm a consistent pattern of defects (Cole et al., 2007). The prevalence of cardiovascular defects associated with first-trimester exposure to bupropion was 10.7 per 1,000 infants.

Results

The advisory committee reviewed the evidence and concluded that the signal did not represent an increased risk. The committee recommended discontinuation of the registry based on findings from the retrospective cohort and 10 years of surveillance through the registry. The committee took the position that sufficient information had accumulated to meet the scientific objective of

the registry. The high lost-to-followup rate was also taken into consideration. The registry closed to new enrollments on November 1, 2007, and continued to follow existing cases through March 31, 2008.

Key Point

In a registry without a specified end date or target size, it is important to periodically review the registry data to determine if the registry has met its scientific objectives and to ensure that the registry purpose is still relevant.

For More Information

- Cole JA, Modell JG, Haight BR, et al. Bupropion in pregnancy and the prevalence of congenital malformations. Pharmacoepidemiol Drug Safety. 2007;16(5):474–84. PMID: 16897811. DOI: 10.1002/pds.1296.
- The Bupropion Pregnancy Registry. Final Report: September 1, 1997–March 31, 2008. Issued August 2008. http://pregnancyregistry.gsk.com/bupropion.html. Accessed June 10, 2019.
- Alwan S, Reefhuis J, Botto LD, et al. Maternal use of bupropion and risk for congenital heart defects. Am J Obstet Gynecol. 2010;203(1):52.e1–6. Epub 2010 Apr 24. PMID: 20417496. DOI: 10.1016/j.ajog.2010.02.015.
- Thyagarajan V, Robin Clifford C, Wurst KE, et al. Bupropion therapy in pregnancy and the occurrence of cardiovascular malformations in infants. Pharmacoepidemiol Drug Safety. 2012;21(11):1240–42. PMID: 23109236. DOI: 10.1002/pds.3271.

Case Example 3. Modifying a Registry Due to Changes in Standards of Care

Description	The GOLD reGISTry was a prospective, multicenter, 5-year global disease registry designed to collect information on patients with advanced and localized gastrointestinal stromal tumors. The registry collected diagnostic, treatment, and outcomes information in order to identify and compare practice patterns worldwide and assist practitioners in making treatment decisions as standards of care evolved.
Sponsor	Novartis Oncology
Year Started	2007
Year Ended	2011
No. of Sites	More than 200
No. of Patients	1,632

Challenge

When it was launched in 2007, the 5-year GOLD reGISTry enrolled only patients with advanced gastrointestinal stromal tumors (GIST). This population was of interest to researchers because standards of care for advanced GIST were not as clearly defined and widely used as the standard of care for localized GIST, which was complete surgical excision. The sponsor expected that the

outcomes data collected from advanced GIST patients would be valuable in helping to refine standards of care for these patients.

In 2008 and 2009, Gleevec®/Glivec® (imatinib mesylate) received FDA and European Medicines Agency (EMA) approval for adjuvant use in localized GIST after tumor resection. This approval, combined with emerging clinical trial data, prompted new interest in collecting diagnostic, treatment, and outcomes information from patients with localized GIST.

Proposed Solution

The sponsor had selected a steering committee with engaged key opinion leaders who provided guidance for the study and encouraged flexibility in study design to allow for potential changes. In 2009, the steering committee convened and determined that the registry would begin collecting data on patients with localized GIST, in addition to those with advanced disease who were already enrolled in the registry. The study team drafted a protocol amendment to include the localized GIST population and allowing assessment of physician adherence to new clinical guidelines published by the European Society of Medical Oncology the same year. The data management and statistical analysis plans were also revised to allow for the incorporation of the new data.

Significant efforts were then directed at site engagement, including abstract submissions and publicity through the key opinion leaders. The registry also maintained site interest through interim study summaries presented at professional congresses. The sponsor had limited monitoring resources available to accommodate the new patient population, so study designers developed a plan that used remote monitoring and training, reserving onsite visits for research-naïve sites or for-cause audits. This allowed monitors to focus on those sites that required more training and allowed these sites to gain clinical research experience in an observational study.

Results

The registry enrolled 1,632 patients in the two populations within four years: more than 1,000 with advanced GIST, and more than 500 with localized GIST. The registry provided a large dataset on treatment and long-term outcomes for patients with GIST in the real-world setting. The steering committee played an important role in the recruitment and retention of sites, highlighting the importance of the study through publications and interim summaries presented at scientific and professional congresses throughout the enrollment period.

Key Point

Changes in standard of care can significantly impact the design of a study as new treatments are approved or new patient populations become of interest. Registry developers should anticipate that such changes might occur, and should consider what aspects of the registry could be most impacted. A steering committee well regarded in the field and knowledgeable about the disease and treatment can provide significant guidance during registry transitions and keep sites engaged as the changes are implemented.

For More Information

- Barrios CH, Blackstein ME, Blay JY, et al. The GOLD ReGISTry: a Global, Prospective, Observational Registry Collecting Longitudinal Data on Patients with Advanced and Localised Gastrointestinal Stromal Tumours. Eur J Cancer. 2015;51(16):2423-33. PMID: 26248685. DOI: 10.1016/j.ejca.2015.07.010.
- Chacon M, Reichardt P, Gu J, et al. The GOLD reGISTry: A global observational registry collecting longitudinal data on patients with advanced GIST—Second annual summary. J Clin Oncol (Meeting Abstracts). 2010 May;28(15 Suppl.):10092.

Case Example 4. Using Registries To Understand Rare Diseases

Description	The International Collaborative Gaucher Group (ICGG) Gaucher Registry aims to enhance the understanding of the variability, progression, and natural history of Gaucher disease, with the ultimate goals of better guiding and assessing therapeutic intervention, and providing recommendations on patient care to the medical community that will improve the outcomes for patients affected by this disease around the world.
Sponsor	Genzyme, a Sanofi company, Cambridge, MA
Year Started	1991
Year Ended	Ongoing
No. of Sites	700+ sites have enrolled patients
No. of Patients	More than 12,000

Challenge

Rare diseases pose special and unique research challenges. The small number of affected patients often results in limited clinical experience within individual healthcare centers. Therefore, the clinical description of rare diseases may be incomplete or skewed. The medical literature often consists of individual case reports or small case series, limiting understanding of the natural history of rare diseases. Furthermore, randomized controlled trials with adequate sample size and length of followup to assess treatment outcomes may be extremely difficult or not feasible. The challenge is even greater for rare diseases that are chronic in nature, where long-term followup is especially important. As a result, rare diseases are often incompletely characterized and lack published data on symptomatology, disease manifestations, and long-term treatment outcomes.

Gaucher disease, a rare enzyme deficiency that affects fewer than 10,000 known patients worldwide, illustrates many of the challenges facing researchers involved in rare diseases. Gaucher disease has three clinical presentations: Type 1, non-neuronopathic; Type 2, acute neuronopathic; and Type 3, subacute neuronopathic. Physicians who encounter patients with Gaucher disease typically have just one or two affected patients in their practices; only a few physicians around the world have more than 10 to 20 patients with Gaucher disease in their care.

Understanding Gaucher disease is further complicated by the fact that it is a highly heterogeneous and rare disorder with variable progression among patients; a patient cohort from a single center may represent a subset of the entire spectrum of disease phenotypes.

The rarity and chronic nature of Gaucher disease also pose challenges in conducting clinical research. The clinical trial that led to U.S. Food and Drug Administration approval of enzyme replacement therapy (ERT) for Gaucher disease (Ceredase®, alglucerase for injection) in 1991 was a single-arm, open-label study involving only 12 patients followed for from 9 to 12 months. In 1994, a recombinant form of ERT was approved (Cerezyme®, imiglucerase for injection) based on a randomized two-arm clinical trial comparing Ceredase and Cerezyme in 30 patients (15 in each arm) followed for 9 months.

Proposed Solution

Established in 1991, the registry is an ongoing, international, longitudinal disease registry, open to voluntary participation by physicians who care for patients with all subtypes of Gaucher disease, regardless of their treatment status or treatment type. Data on patient demographics; clinical characteristics; treatment regimen; and laboratory, radiologic, and quality-of-life outcome measures are entered and analyzed to address the research challenges of this rare disease. Because of the rarity of Gaucher disease, it is important to create and maintain a reliable, comprehensive registry that serves as an educational resource not only for physicians but also for patients and their families and caregivers. Responsibility for the use, integrity and objectivity of the data and analyses is invested in the ICGG Board of Advisors, which consists of physician-investigators worldwide who are not employees of the sponsor and who advise on the medical and scientific agendas of the registry.

Results

The registry has longitudinal data on more than 12,000 patients from more than 700 healthcare centers in more than 60 countries. The followup period is open-ended and the registry currently has up to 20 years of followup data from individual patients. The registry has collected more than 50,000 patient-years of followup during the past 21 years. Physician participation and patient enrollment have increased consistently from year to year since 1991.

Analyses of the extensive body of longitudinal data have increased knowledge of the disease in a broad range of topics, including the natural history of Gaucher disease; phenotypic and genotypic variation among patients; diagnosis, treatment, and management of the disease; disease manifestations in children; long-term treatment outcomes for ERT; bone disease and complications associated with the disease; and neuronopathic Gaucher disease. Data generated from the registry have been published in nearly 30 key articles and have provided much needed and important insight into this rare genetic disease.

In 2002, the registry published the clinical outcomes of 1,028 patients treated with ERT with up to 5 years of followup. As more data have been gathered through the registry over the past decade, long-term outcomes in patients with Type 1 Gaucher disease after 10 years of ERT have become available, thus providing new reference benchmarks for assessing clinical responses to ERT for various disease parameters. Other more recent publications based on analyses of data from the registry have focused on important specific aspects of Gaucher disease, such as the effects of early intervention with ERT on the incidence of bone pathology, demographic and

clinical characteristics of patients with neuronopathic Gaucher disease, ERT dose-response relationships for disease parameters in patients with Gaucher disease type 1, and phenotypic heterogenicity and genetic variation among patients.

Along with the growth of the registry and the availability of data on Gaucher disease, interest in special patient populations and specific aspects of Gaucher disease continually emerge. As a result, research initiatives into disease subpopulations have been launched recently: the Neurological Outcomes Subregistry, which will begin to evaluate the neurologic manifestations of Gaucher disease and the effects of treatment on these complications; and the Pregnancy Subregistry, which will track the management of Gaucher disease during pregnancy as well as pregnancy outcomes.

The collective clinical experience of the registry led to the development of recommendations for evaluation and monitoring of patients with Gaucher disease. The analysis of registry data on treatment outcomes has facilitated the establishment of therapeutic goals for patients with Type 1 Gaucher disease. Together, these publications have formed the foundation for a consensus- and evidence-based disease management approach, something usually only possible for much more common diseases. In 2008, a benchmark analysis was published that documented the achievement of therapeutic goals after 4 years of ERT among registry patients.

As disease awareness has increased over time, healthcare providers have sought more direct access to general and patient-specific disease information. Therefore, when the registry changed its technology platform in 2011, it established two key objectives: to simplify data entry to help keep data complete and accurate, and to support the community's increased interest in access to data, aggregate reports, and collaborative expertise. To help meet these goals, the registry ensured that the new platform included functionality that allows physicians direct access to aggregate and patient-specific reporting as well as the ability to download their own data to support their own research. This important application of technology enables the registry to "give back" supportive and research tools to those who contribute to the overall registry dataset. This includes the availability of data to address clinical and scientific questions; useful disease management tools, such as interactive patient case reports that a physician can share with other healthcare providers and with patients themselves; and a larger, better-connected worldwide community of physicians and allied health providers who can share information, identify trends, improve best practices, and build awareness of Gaucher disease that will optimize patient outcomes.

Key Point

For rare or ultra-rare conditions, an international, longitudinal disease registry may be the best or only feasible way to comprehensively increase knowledge about the clinical characteristics and natural history of the disease and assess the long-term outcomes of treatment.

For More Information

- Weinreb NJ, Charrow J, Andersson HC, et al. Effectiveness of enzyme replacement therapy in 1028 patients with type 1 Gaucher disease after 2 to 5 years of treatment: a report from the Gaucher Registry. Am J Med. 2002;113(2):112–9. PMID: 12133749. DOI: 10.1016/s0002-9343(02)01150-6.

- Vom Dahl S, Weinreb N, Charrow J, et al. Long-term Clinical Outcomes in Type 1 Gaucher Following 10 Years of Treatment with Imiglucerase; Presented at the 2011 Workshop of the European Study Group on Lysosomal Disease (ESGLD); September 3-6, 2011; Langvik, Finland.
- Mistry PK, Deegan P, Vellodi A, et al. Timing of initiation of enzyme replacement therapy after diagnosis of type 1 Gaucher disease: effect on incidence of avascular necrosis. Br J Haematol. 2009;147(4):561–70. PMID: 19732054. DOI: 10.1111/j.1365-2141.2009.07872.x.
- Tylki-Szymanska A, Vellodi A, El-Beshlawy A, et al. Neuronopathic Gaucher disease: demographic and clinical features of 131 patients enrolled in the International Collaborative Gaucher Group Neurological Outcomes Subregistry. J Inherit Metab Dis. 2010;33(4):339–46. PMID: 20084461. DOI: 10.1007/s10545-009-9009-6.
- Fairley C, Zimran A, Phillips M, et al. Phenotypic heterogeneity of N370S homozygotes with type I Gaucher disease: an analysis of 798 patients from the ICGG Gaucher Registry. J Inherit Metab Dis. 2008;31(6):738–44. PMID: 18979180 DOI: 10.1007/s10545-008-0868-z.

Chapter 3. Registry Design

1. Introduction

This chapter is intended as a high-level practical guide to the application of epidemiologic methods that are particularly useful in the design of registries that evaluate patient outcomes. Since it is not intended to replace a basic textbook on epidemiologic design, readers are encouraged to seek more information from textbooks and scientific articles. Table 3-1 summarizes the key considerations for study design that are discussed in this chapter. Throughout the design process, registry planners may want to discuss options and decisions with the registry stakeholders and relevant experts to ensure that sound decisions are made. The choice of groups to be consulted during the design phase generally depends on the nature of the registry, the registry funding source and funding mechanism, and the intended audience for registry reporting.

Table 3-1. Considerations for study design

Construct	Relevant Questions
Research question	What are the clinical and/or public health questions of interest?
Resources	What resources, in terms of funding, sites, clinicians, experts, and patients, are available for the study?
Exposures and outcomes	How do the clinical questions of interest translate into measurable exposures and outcomes?
Data sources	Where can the necessary data elements be obtained?
Study design	What types of design can be used to answer the questions or fulfill the purpose?
Study population	What types of patients are needed for study? Is a comparison group needed? How should patients be selected for study?
Site and Patient Recruitment	How should the study population be recruited, taking into account the target population(s), types of healthcare providers of interest and study design?
Study size and duration	For how long should data be collected, and for how many patients?
Internal and external validity	What are the potential sources of bias and how much could they distort the study findings (e.g., rate or effect estimates)? What are the concerns about generalizability of the results (external validity)?

2. Research Questions Appropriate for Registries

The questions typically addressed in registries range from purely descriptive questions aimed at understanding the characteristics of people who develop the disease and how the disease

generally progresses, to highly focused questions intended to support decision making. Registries focused on determining clinical effectiveness or cost-effectiveness or assessing safety or harm are generally hypothesis driven and concentrate on evaluating the effects of specific treatments on patient outcomes. Research questions should address the registry's purposes, as broadly described in Table 3-2.

Table 3-2. Overview of registry purposes

Purpose	Description
Natural history	Assessing natural history, including estimating the magnitude of a problem; determining the underlying incidence or prevalence rate of a condition; examining trends of disease over time; conducting surveillance; assessing service delivery and identifying groups at high risk; documenting the types of patients served by a health provider; and describing and estimating survival.
Effectiveness	Determining clinical effectiveness, cost-effectiveness, or comparative effectiveness of a test or treatment, including for the purpose of determining reimbursement, supporting value-based care, or generating real-world evidence for regulatory decision making.
Safety	Measuring or monitoring safety and harm associated with the use of specific products and treatments, including conducting comparative evaluation of safety and effectiveness.
Quality	Measuring or improving quality of care, including conducting programs to measure and/or improve the practice of medicine and/or public health.

Observational studies derived from registries (or "registry-based studies") are an important part of the research armamentarium alongside interventional studies, such as randomized controlled trials (RCTs), registry-based randomized trials, or other pragmatic randomized trials, and retrospective studies, such as studies derived exclusively from administrative claims data. Each of these study designs has strengths and limitations, and the selection of a study design should be guided by the research questions of interest. (See Chapter 2 for a discussion of the factors that influence the study design decision.) In some cases, multiple studies with different designs or a hybrid study that combines study designs will be necessary to address a research question. In fact, this more comprehensive approach to evidence development is likely to become more common as researchers strive to address multiple questions for multiple stakeholders most efficiently. Observational studies and interventional studies are more complementary than competitive, precisely because some research questions are better answered by one method than the other. Interventional studies are considered by many to provide the highest grade evidence for evaluating whether a drug has the ability to bring about an intended effect in optimal or "ideal world" situations, a concept also known as "efficacy."[1] Observational designs, on the other hand, are particularly well suited for studying broader populations, understanding actual results (e.g., some safety outcomes) in real-world practice, and for obtaining more representative quality-of-life information. This is particularly true when the factors surrounding the decision to treat are an important aspect of understanding treatment effectiveness.[2]

In many situations, nonrandomized comparisons either are sufficient to address the research question or, in some cases, may be necessary because of the following issues with randomizing patients to a specific treatment:

- *Equipoise:* Can providers ethically introduce randomization between treatments when the treatments may not be clinically equivalent?
- *Ethics:* If reasonable suspicion about the safety of a product has become known, would it be ethical to conduct a trial that deliberately exposes patients to potential harm? For example, can pregnant women be ethically exposed to drugs that may be teratogenic?
- *Practicality:* Will patients enroll in a study where they might not receive the treatment, or might not receive what is perceived to be the best treatment? How can adherence to a treatment be studied, if not by observing what people do in real-world situations?

Registries are particularly suitable for some types of research questions, such as:

- Natural history studies where the goal is to observe clinical practice and patient experience but not to introduce any intervention.
- Studies of rare diseases or rare exposures that often require working with many sites to study relatively few patients.
- Measures of clinical effectiveness, especially as related to adherence, where the purpose is to learn about what patients and practitioners actually do and how their actions affect real-world outcomes. This is especially important for treatments that have poor adherence.
- Studies of effectiveness and safety for which clinician training and technique are part of the study of the treatment (e.g., a procedure such as placement of carotid stent).
- Studies of heterogeneous patient populations, since unlike randomized trials, registries generally have much broader inclusion criteria and fewer exclusion criteria. These characteristics lead to studies with greater generalizability (external validity) and may allow for assessment of subgroup differences in treatment effects.
- Followup for delayed or long-term benefits or harm, since registries can extend over much longer periods than most clinical trials (because of their generally lower operational costs and lesser burden on participants).
- Surveillance for rare events.
- Studies for treatments in which randomization is unethical, such as intentional exposure to potential harm (as in safety studies of marketed products that are suspected of being harmful).
- Studies for treatments in which randomization is not necessary, such as when certain therapies are only available in certain places owing to high cost or other restrictions (e.g., proton beam therapy).
- Studies for which blinding is challenging or unethical (e.g., studies of surgical interventions, complex or sequential treatments, acupuncture).
- Studies of rapidly changing technology.

- Studies of conditions with complex treatment patterns and treatment combinations.
- Studies of healthcare access and barriers to care.
- Evaluations of actual standard medical practice.
- Studies of diagnostic outcomes, particularly when the outcome of interest is relatively rare and large cohorts are needed to assess test performance metrics.

Registry studies may also include embedded substudies as part of their overall design. These substudies can themselves have various designs (e.g., highly detailed prospective data collection on a subset of registry participants, or a case-control study focused on either incident or prevalent cases identified within the registry). Registries can also be used as a framework for RCTs.[3,4]

3. Translating Clinical Questions Into Measurable Exposures and Outcomes

The specific clinical questions of interest in a registry will guide the definitions of study subjects, exposure, and outcome measures, as well as the study design, data collection, and analysis. In the context of registries, the term "exposure" is used broadly to include treatments and procedures, healthcare services, diseases, and conditions.

The clinical questions of interest can be defined by reviewing published clinical information, soliciting experts' opinions, and evaluating the expressed needs of the patients, healthcare providers, payers, and other stakeholders. Examples of research questions, key outcome and exposure variables, and sources of data are shown in Table 3-3.

Table 3-3. Examples of research questions and key exposures and outcomes

Research Question	Key Exposure (source of data)	Key Outcome (source of data)
What is the expected time to rejection for first kidney transplants among adults, and how does that vary according to immunosuppressive regimen?	All immunosuppressants, including dosage and duration (clinician or medical record)	Organ rejection (clinician or medical record)
Are patients using a particular treatment better able to perform activities of daily living than others?	Treatments for disease of interest (clinician or medical record)	Ability to independently perform key activities related to daily living (patient)
Do patients undergoing gastric bypass surgery for weight loss use fewer healthcare resources in the year following surgery?	Surgery (clinician or medical record)	Number of inpatient and outpatient visits, medications dispensed, associated costs (administrative databases, clinician, or medical record)
Are patients using a particular drug more likely to have serious adverse pregnancy outcomes?	Drug use by mother during pregnancy (clinician, medical record, or patient)	Pregnancy outcome (clinician, medical record, or patient)

As these examples show, the outcomes are the main endpoints of interest posed in the research question. These typically represent measures of health, onset or progression of illness, or adverse events; they also commonly include patient-reported outcome measures, such as quality of life measures, and measures of healthcare utilization and costs. More information on selecting outcome measures is provided in Chapter 4.

In addition to outcomes, relevant exposures also derive from the main research question and relate to why a patient might experience benefit or harm. Evaluation of an exposure includes collection of information that affects or augments the main exposure, such as dose, duration of exposure, route of exposure, or adherence. Other variables of interest include independent risk factors for the outcomes of interest (e.g., comorbidities, age), as well as variables known as potential confounding variables, that are related to both the exposure and the outcome and are necessary for conducting valid statistical analyses. Confounding can result in inaccurate estimates of association between the study exposure and outcome through mixing of effects. To continue with an asthma example, a study of a new asthma medication should collect prior history of treatment resistance or else results may be biased. The bias could occur because treatment resistance may relate both to the likelihood of receiving the new drug (meaning that doctors will be more likely to try a new drug in patients who have failed other therapies) and the likelihood of having a poorer outcome (e.g., hospitalization). Some efforts to standardize outcome measures, such as the OMF project, specify key risk factors and potential confounding variables that should be captured. Refer to Chapters 4 and 5 for more information.

4. Finding the Necessary Data

The identification of key outcome and exposure variables and patients will drive the strategy for data collection, including the choice of data sources. A key challenge to registries, as with all studies that require primary data collection, is that it may not be possible to collect all desired data. As discussed in Chapter 5, data collection should be both purpose-driven and broadly applicable. For example, while experimental imaging studies may provide interesting data, if the imaging technology is not widely available, the data will not be available for enough patients to be useful for analysis. Moreover, the registry findings will not be generalizable if only sophisticated centers that have such technology participate. Instead, registries should focus on collecting the data necessary to achieve their purpose(s) while minimizing the burden on patients and clinicians when feasible.

Registry data can be obtained from patients, clinicians, medical records, and linkage with other sources. While many registries relied on primary data collection in the past, the increasing availability of electronic healthcare data has introduced new opportunities for registries to capture data from secondary sources, such as electronic medical records and administrative databases, thus reducing data collection burden and increasing efficiency. More information on the technical aspects of linking or integrating existing data sources into registries can be found in the supplemental eBook on Registry Informatics.[5] These approaches can yield rich datasets on large patient cohorts that can be used to address the primary objective of the registry as well as numerous secondary objectives. However, significant effort is often needed to clean, standardize, and normalize the data, and these data may not be recorded with the same rigor and quality assurance procedures that are used in some registries. Chapters 6 and 11 explore these issues.

5. Resources and Efficiency

Ideally, a study is designed to optimally answer a research question of interest and funded adequately to achieve the objectives based on the requirements of the design. Frequently, however, finite resources are available at the outset of a project that constrain the approaches that may be pursued. Often, through efficiencies in the selection of a study design and patient population (observational vs. RCT, case-control vs. prospective cohort), selection of data sources (e.g., use of secondary data sources vs. information collected directly from clinicians or patients), restriction of the number of study sites, or other approaches, studies may be planned that provide adequate evidence for addressing a research question within a specified budget. Section 6 below discusses how certain designs may be more efficient for addressing some research questions.

6. Study Designs for Registries

Registries provide a framework for various types of observational study designs. Typically, a registry is designed to support a specific study, but additional studies may be nested as substudies within the registry framework to address secondary objectives or questions that arise during the course of the registry. Additional data may need to be collected to facilitate examination of questions that arise. Before capturing new data elements, the steps outlined in Chapter 2, including assessing feasibility, considering the necessary scope and rigor, and evaluating the regulatory/ethical impact, should be undertaken.

The study models of case series, cohort, case-control, and case-cohort are commonly applied to registry data and are described briefly here. Other models are useful in some situations, but are not covered here. For example, case-crossover studies are efficient designs for studying the effects of intermittent exposures (e.g., use of erectile dysfunction drugs) on conditions with sudden onset. Another example is a pre and post study that enrolls sites prior to introduction of new technology, collects baseline data, and continues data collection after new technology is available. Registries may also provide a platform for pragmatic randomized trials.[6,7] In a pragmatic trial, patients or providers may be randomized as to which intervention or quality improvement tools they use; the comparators are generally one or more other active treatments (generally referred to as standard of care) rather than placebos; and patients are observed without further intervention. Also, there has been recent interest in applying the concept of adaptive clinical trial design to registries. The U.S. Food and Drug Administration defines an adaptive design as “a clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial.[8] While many long-term registries are modified after initiation, the more formal aspects of adaptive trial design have yet to be applied regularly to registries and observational studies.

Determining what framework will be used to analyze the data is important in designing the registry and registry data collection procedures. Readers are encouraged to consult textbooks of epidemiology and pharmacoepidemiology for more information. Many of the references in Chapter 13 relate to study design and analysis.

6.1 Case Series Design

Using a registry population to develop case series is a straightforward application that does not require sophisticated analytics. Depending on the generalizability of the registry itself, case series drawn from the registry can be used to describe the characteristics to be used in comparison to other case series (e.g., from spontaneous adverse event reports). Self-controlled methods, including self-controlled case series, are a relatively new set of methods that lends itself well to registry analyses as it focuses on only those subjects who have experienced the event of interest and uses an internal comparison to derive the relative (not absolute) incidence of the event during the time the subject is "exposed" compared with the incidence during the time when they are "unexposed."[9] This design implicitly controls for all confounders that do not vary over the followup time (e.g., gender, genetics, geographic area), as the subject serves as his or her own control. The self-controlled case series design may also be very useful in those circumstances where a comparison group is not available. Self-controlled case series require that the probability of exposure is not affected by the occurrence of an outcome; in addition, for non-recurrent events, the method works only when the event risk is small and varies over the followup time. Derivative methods, grouped as self-controlled cohort methods, include observational screening,[10] interrupted time series,[11] and temporal pattern discovery.[12] These methods compare the rate of events post-exposure with the rate of events pre-exposure among patients with at least one exposure. Registries that leverage secondary data sources, such as electronic health records, are well-suited for these methods because they typically capture data on most if not all patients at each participating site.

6.2 Cohort Design

Cohort studies follow over time a group of people who possess a characteristic, to see if individuals in the group develop a particular endpoint or outcome. The cohort design is used for descriptive studies as well as for studies seeking to evaluate comparative effectiveness and/or safety or quality of care. Cohort studies may include only people with exposures (such as to a particular drug or class of drugs) or disease of interest. Cohort studies may also include one or more comparison groups for which data are collected using the same methods during the same period. A single cohort study may in fact include multiple cohorts, each defined by a common disease, characteristic, or exposure. Cohorts may be small, such as those focused on rare diseases, but often they target large groups of people (e.g., in safety studies), such as all users of a particular drug or device. Limitations of registry-based cohort studies may include lack of data on treatments provided outside the participating sites (e.g., a surgical registry may have limited information on the patient's use of chiropractic treatments) and underreporting of outcomes if a patient leaves the registry or is not adequately followed up.[13] These pitfalls should be considered and addressed when planning a study.

6.3 Case-Control Design

A case-control study gathers patients who have a particular outcome or exposure or who have suffered an adverse event ("cases") and "controls" who have not but are representative of the source population from which the cases arise.[14] If properly designed and conducted, it should yield results similar to those expected from a cohort study of the population from which the cases

were derived. The case-control design is often employed for understanding the etiology of rare diseases[15] because of its efficiency. In studies where expensive data collection is required, such as some genetic analyses or other sophisticated testing, the case-control design is more efficient and cost effective than a cohort study because a case-control design collects information only from cases and a sample of non-cases. However, if no *de novo* data collection is required, the use of the cohort design may be preferable since it avoids the challenge of selecting a suitable control group and the concomitant danger of introducing more bias.

Depending on the outcome, exposure, or event of interest, cases and controls may be identifiable within a single registry. For example, in the evaluation of restenosis after coronary angioplasty in patients with end-stage renal disease, investigators identified both cases and controls from an institutional percutaneous transluminal coronary angioplasty registry; in this example, controls were randomly selected from the registry and matched by age and gender.[16] Alternatively, cases can be identified in the registry and controls chosen from outside the registry. Care must be taken, however, that the controls from outside the registry meet the requirement of arising from the same source population as the cases to which they will be compared. Matching in case-control designs—for example, ensuring that patient characteristics such as age and gender are similar in the cases and their controls—may yield additional efficiency, in that a smaller number of subjects may be required to answer the study question with a given power. However, matching does not eliminate confounding and must be undertaken with care. Matching variables must be accounted for in the analysis, because a form of selection bias similar to confounding will have been introduced by the matching.[17]

Properly executed, a case-control study can add efficiency to a registry if more extensive data are collected by the registry only for the smaller number of subjects selected for the case-control study. This design is sometimes referred to as a "nested" case-control study, since subjects are taken from a larger cohort. It is generally applied because of budgetary or logistical concerns relating to the additional data desired. Nested case-control studies have been conducted in a wide range of patient registries, from studying the association between oral contraceptives and various types of cancer using the Surveillance Epidemiology and End Results (SEER) program[18-20] to evaluating the possible association of depression with Alzheimer's disease. As an example, in the latter case-control study design, probable cases were enrolled from an Alzheimer's disease registry and compared with randomly selected nondemented controls from the same base population.[21] The increasing availability of electronic healthcare data may make case-control designs unnecessary in some situations, as registries may be able to capture large volumes of data on large numbers of patient efficiently and use advanced statistical techniques, such as propensity score matching, to build cohorts for analysis. This approach is feasible when all necessary data are available in secondary sources, such as electronic health records. In cases where some data are unavailable in the medical record (e.g., patient-reported outcomes), case-control designs may be an appropriate option.

Case-control studies present special challenges with regard to control selection. More information on considerations and strategies can be found in a set of papers by Wacholder.[22-24]

6.4 Case-Cohort Design

The case-cohort design is a variant of the case-control study. As in a case-control study, a case-cohort study enrolls patients who have a particular outcome or who have suffered an adverse event ("cases"), and "controls" who have not, but who are representative of the source population from which the cases arise. In nested case-control studies where controls are selected via risk-set sampling, each person in the source population has a probability of being selected as a control that is, ideally, in proportion to his or her person-time contribution to the cohort. In a case-cohort study, however, each control has an equal probability of being sampled from the source population.[25] This allows for collection of pertinent data for cases and for a sample of the full cohort, instead of the whole cohort. For example, in a case-cohort study of histopathologic and microbiological indicators of chorioamnionitis, which included identification of specific microorganisms in the placenta, cases consisted of extreme preterm infants with cerebral palsy. Controls, which can be thought of as a randomly selected subcohort of subjects at risk of the event of interest, were selected from among all infants enrolled in a long-term study of preterm infants.[26]

With the assumptions that competing risks and loss to followup are not associated with the exposure or the risk of disease, the case-cohort design allows for the selection of one control group that can be compared with various case series since the controls are selected at the beginning of followup. Analogous to a cohort study where every subject in the source population is at risk for the disease at the start of followup, the control series in a case-cohort design represents a sample of the exposed and unexposed in the source population who are disease-free at the start of followup.

7. Choosing Patients for Study

The purpose of a registry is to provide information or describe events and patterns, and often to generate hypotheses about a specific patient population to whom study results are meant to apply. Studies can be conducted of people who share common characteristics, with or without the inclusion of comparison groups. For example, studies can be conducted of:

- People with a particular disease/outcome or condition.
    - Examples include studies of the occurrence of cancer or rare diseases, pregnancy outcomes, and recruitment pools for clinical trials.
- Those with a particular exposure (e.g., to a product, procedure, or other health service, or an environmental or personal exposure).
    - Examples include general surveillance registries, pregnancy registries for particular drug exposures, and studies of exposure to medications and to devices such as stents.[27] They also include studies of people who were treated under a quality improvement program, studies of people with a specific environmental exposure[28] or personal exposure[29] and studies of a exposure that requires controlled distribution, such as drugs with serious safety concerns (e.g., isotretinoin, clozapine, natalizumab [Tysabri®]), where the participants in the registry are identified because of their participation in a controlled distribution/risk management program.

- Those who were part of a program evaluation, disease management effort, or quality improvement project.
    - An example is the evaluation of the effectiveness of evidence-based program guidelines on improving treatment.

7.1 Target Population

Selecting patients for registries can be thought of as a multistage process that begins with understanding the target population (the population to which the findings are meant to apply, such as all patients with a disease or an exposure) and then selecting a sample of this population for study. Some registries will enroll all, or nearly all, of the target population, but most registries will enroll only a subset of the target population. The accessible study population is that portion of the target population to which the participating sites have access. The actual study population is the subset of those who can be identified, invited to participate, and who agree to participate.[30] While it is desirable for the patients who participate in a study to be representative of the target population, representativeness is generally defined in terms of "typical" patients and providers, rather than through an actual sample of all known patients. It is rarely possible to enumerate the complete sample frame to facilitate statistical sampling, either for budgetary reasons or for reasons of practicality.[31] An exception is registries composed of all users of a product (as in postmarketing surveillance studies where registry participation is required as a condition of receiving an intervention), an approach which is becoming more common to manage expensive interventions and/or to track potential safety issues.

Certain populations pose greater difficulties in assembling an actual study population that is truly representative of the target population. Children and other vulnerable populations present special challenges in recruitment, as they typically will have more restrictions imposed by institutional review boards and other oversight groups.

As with any research study, very clear definitions of the inclusion and exclusion criteria are necessary and should be well documented, including the rationale for these criteria. A common feature of registries is that they typically have few inclusion and exclusion criteria, which enhances their applicability to broader populations. Restriction, the strategy of limiting eligibility for entry to individuals within a certain range of values for a confounding factor, such as age, may be considered in order to reduce the effect of a confounding factor when it cannot otherwise be controlled, but this strategy may reduce the generalizability of results to other patients.

These criteria will largely be driven by the study objectives and any sampling strategy. For a more detailed description of target populations and their subpopulations, and how these choices affect generalizability and interpretation, see Chapter 13.

Once the patient population has been identified, attention shifts to selecting the institutions and providers from which patients will be selected. For more information on recruiting patients and providers, see Chapter 10. Depending on the purpose of the registry, direct enrollment of patients may also be appropriate. (See Case Examples 7, 12, and 20.)

7.2 Comparison Groups

Once the target population has been selected and the mechanism for their identification (e.g., by providers) is decided, the next decision involves determining whether to collect data on comparators. Depending on the purpose of the registry, internal or external (contemporaneous or historical) groups can be used to strengthen the understanding of whether the observed effects are different from what would be expected to occur. Comparison groups are most useful in registries where it is important to distinguish between alternative decisions or to assess relative benefits and risks of various treatments. Registries without comparison groups can be used for descriptive purposes, such as characterizing the natural history of a disease or condition, or for exploratory purposes and are often complemented with external benchmarks from contemporaneous or historical data. The addition of a comparison group may add significant complexity, time, and cost to a registry, although the cost can be quite modest if existing data can be used for selected points of comparison.

Although it may be appealing to use more than one comparison group in an effort to overcome the limitations that may result from using any single group, multiple comparison groups pose their own challenges to the interpretation of registry results. For example, the results of comparative safety and effectiveness evaluations may differ depending on the comparison group used, making interpretation of the findings difficult. Generally, it is preferable to make judgments about the best comparison group for study during the design phase and then concentrate resources on these subjects. Also, sensitivity analyses can be used to quantify the likely impact of bias on the study findings (See Chapter 13.)

The choice of comparison groups is more complex in registries than in randomized clinical trials. Whereas clinical trials use randomization to try to achieve an equal distribution of known and unknown risk factors between treatment groups, registry studies need to use various design and analytic strategies to adjust for the confounders that they have measured. The concern for any observational studies is that a person who receives a new intervention has different characteristics than people who receive other treatments or who receive no treatment at all, and that these different characteristics influence the person's likelihood of experiencing benefit or harm. In other words, treatment choices are often related to demographic and lifestyle characteristics, stage of disease, and the presence of coexisting conditions that affect clinician decision making about treatments.[32] To achieve comparability, registries may use inclusion criteria that, for example, restrict the registry focus to patients who have had the disease for a similar duration, are receiving their first treatment for a new condition, or are progressing to second-line treatments.

Other design techniques that can be used in registries, particularly those with large numbers of patients, include matching study subjects using statistical techniques (e.g., propensity scoring) to create strata of patients with similar likelihood of receiving a treatment or of experiencing benefits or risks. As an example, consider a recent study of a rare side effect in coronary artery surgery for patients with acute coronary syndrome. In this instance, the main exposure of interest was the use of antifibrinolytic agents during revascularization surgery, a practice that had become standard for such surgeries. The sickest patients, who were most likely to have adverse events, were much less likely to be treated with antifibrinolytic agents. To address this, the investigators measured more than 200 covariates (by drug and outcome) per patient and used this

information in a propensity score analysis. The results of this large-scale observational study revealed that the traditionally accepted practice (aprotinin) was associated with serious end-organ damage and that the less expensive generic medications were safe alternatives.[33] Registries that capture large volumes of data from secondary sources are particularly well-suited for propensity score analysis because the large amounts of data can produce better models and more comparable cohorts. The drawback to post-hoc matching is the loss of information since some patients in the treatment (or disease) group will be excluded from the analysis if a suitable match is not available; there are also debates concerning the impact of how groups are weighted when propensity scores are used (see Chapter 13.)[34,35]

An internal comparison group refers to simultaneous data collection for patients who are similar to the focus of interest (i.e., those with a particular disease or stage of disease), but who do not have the condition or exposure of interest. For example, a registry might collect information on patients with arthritis who are using acetaminophen for pain control. An internal comparison group could be arthritis patients who are using other medications for pain control. Data regarding similar patients, collected during the same calendar period and using the same data collection methods, are also useful for characterizing treatment heterogeneity and for understanding various risk factors, such as for studying the effects in certain age categories or among people with similar comorbidities. However, the information value and utility of these comparisons depend largely on having adequate numbers of patients in the subgroups of interest, and such analyses may need to be specified *a priori* to ensure that recruitment supports them. Internal comparisons are particularly useful because data are collected using the same tools and endpoints and during the same observation period as for all study subjects, which will account for time-related influences that may be external to the study while also assuring the requisite data are available for all study subjects. For example, if an important scientific article is published that affects general clinical practice, and the publication occurs during the period in which the study is being conducted, clinical practice may change. The effects may be comparable for groups observed during the same period through the same system, whereas information from historical comparisons, for example, would be expected to reflect different practices.

An external comparison group is a group of patients similar to those who are the focus of interest, but who do not have the condition or exposure of interest, and for whom relevant data have been collected outside of the registry. For example, the SEER program maintains national data about cancer and has provided useful comparison information for many registries where cancer is an outcome of interest.[36] External comparison groups can provide informative benchmarks for understanding effects observed and for assessing generalizability, and they are currently being used by regulators in the United States and European Union to study treatment effectiveness for label expansions. In some cases, registry data are being used as the source of external comparators for phase II trials.[37] Additionally, large clinical and administrative claims databases can contribute useful information on comparable subjects for a relatively low cost. Depending on the outcome of interest, a limitation of external comparison groups is that the data are generally not collected the same way and the same information may not be available; however, these differences may not be problematic for some outcomes, such as mortality.

External comparators may be contemporaneous (i.e., referring to data collected during the same timeframe as the registry patients) or historical, referring to patients who are similar to the focus of interest, but who do not have the condition or exposure of interest, and for whom information

was collected in the past (such as before the introduction of an exposure or treatment or development of a condition). Historical controls may actually be the same patients who later become exposed, or they may consist of a completely different group of patients. For example, historical comparators are often used for pregnancy studies since there is a large body of population-based surveillance data available, such as the Metropolitan Atlanta Congenital Defects Program (MACDP).[38] This design provides weak evidence because symmetry is not assured (i.e., the patients in different time periods may not be as similar as desired). Historical controls are susceptible to bias by changes over time in uncontrollable, confounding risk factors, such as differences in climate, management practices, and nutrition. Bias stemming from differences in measuring procedures over time may also account for observed differences.

An approach related to the use of historical comparisons is the use of Objective Performance Criterion (OPC) as a comparator. This research method has been described as an alternative to randomized trials, particularly for the study of devices.[39] An OPC "refers to a numerical target value derived from historical data from clinical studies and/or registries and may be used in a dichotomous (pass/fail) manner by FDA for the review and comparison of safety or effectiveness endpoints."[40] A U.S. Food and Drug Administration guidance document for pivotal clinical investigations for medical devices includes a description of how OPCs may be used in the context of medical device studies. Registries serve as a source of reliable historical data in this context, particularly when combined with trials and other data sources.

There are several situations in which internal comparators may be impractical, unethical, or impossible and a historical comparison may be considered:

- When one cannot ethically continue the use of older treatments or practices, or when clinicians and/or patients refuse to continue their use, so that the researcher cannot identify relevant sites using the older treatments.
- When uptake of a new medical practice has been rapid, concurrent comparisons may differ so markedly from treated patients, with regard to factors related to outcomes of interest, that they cannot serve as valid comparison subjects due to intractable confounding.
- When conventional treatment has been consistently unsuccessful and the effect of new intervention is obvious and dramatic (e.g., first use of a new product for a previously untreatable condition).
- When collecting the comparison data is too expensive.
- When the Hawthorne effect (a phenomenon that refers to changes in the behavior of subjects because they know they are being studied or observed) makes it impossible to replicate actual practice in a comparison group during the same period.
- When the desired comparison is to usual care or "expected" outcomes at a population level, and data collection is too expensive due to the distribution or size of that population.

8. Registry Size and Duration

Precision in measurement and estimation corresponds to the reduction of random error. Depending on available budget, precision can sometimes be improved by increasing the number of study subjects or followup period.[41]

During the registry design stage, it is critical to explicitly state the target number of sites and patients, how long patients should be followed, and the justifications for these decisions. These decisions should be based on the overall purpose of the registry but tempered by budget and whether the proposed registry would fill an important information gap, even if the target study size is not optimal. For example, in addressing specific questions of product safety or effectiveness, the desired level of precision to confirm or rule out the existence of an important effect should be specified, and ideally should be linked to policy or practice decisions that will be made based on the evidence. Nonetheless, registries may make important contributions even if they are only able to evaluate large effects, should they exist.[42] For registries with aims that are descriptive or hypothesis generating, study size may be arrived at through other considerations.

The duration of registry enrollment and followup should be determined both by required number of patients or person-years desired to achieve the target statistical power and by time- and budget-related considerations. The expected (or theoretical) induction period for some outcomes of interest should be considered, and ideally, sufficient followup time allowed for the exposure under study to have induced or promoted the outcome. Calendar time may be a consideration in studies of changes in clinical practice or interventions that have a clear beginning and end. The need for evidence to inform policy may also determine a timeframe within which the evidence must be made available to decision-makers. For practical purposes, it may be useful to evaluate the risk over a time period that is feasible to study, for example, characterizing the benefits and risks for five years after receipt of an artificial hip, recognizing that a much longer time period may also be of interest, e.g., how does the hip perform after 10 years.

A detailed discussion of the topic of study size calculations for registries is provided in Appendix A. For present purposes it is sufficient to briefly describe some of the critical inputs to these calculations that must be provided by the registry developers:

- The expected timeframe of the registry and the time intervals at which analyses of registry data will be performed.
- Either the expected size of clinically meaningful effects (e.g., minimum clinically important differences) or the desired precision of the effect estimates.
- Whether or not the registry is intended to support regulatory decision making. If the results from the registry will affect regulatory action—for example, the likelihood that a product may be pulled from the market—then the precision of the overall risk estimate is important, as is the necessity to predict and account for attrition in designing the target study size.

In a classical calculation of sample size, the crucial inputs that must be provided by the investigators include either the size of clinically important effects or their required precision. For

example, suppose that the primary goal of the registry is to compare surgical complication rates in general practice with those in randomized trials. The inputs to the statistical power calculations would include the complication rates from the randomized trials (e.g., 4 percent) and the complication rate in general practice, which would reflect a meaningful departure from this rate (e.g., 6 percent). If, on the other hand, the goal of the registry is simply to track complication rates (and not to compare the registry with an external standard), then the investigators should specify the required width of the confidence interval associated with those rates. For example, in a large registry, the 95-percent confidence interval for a 5-percent complication rate might extend from 4.5 percent to 5.5 percent. If all of the points in this confidence interval lead to the same decision, then an interval of ±0.5 percent is considered sufficiently precise, and this is the input required for the estimation of sample size.

Specifying the above inputs to sample size calculations is a substantial matter and usually involves a combination of quantitative and qualitative reasoning. The issues involved in making this specification are essentially similar for registries and other study designs, though for registries designed to address multiple questions of interest, one or more primary objectives or endpoints must be selected that will drive the selection of a minimum sample size to meet those objectives.

Other considerations that may be taken into account when estimating study sizes include—

- whether multiple comparisons are being made and subjected to statistical testing, although this notion has been soundly challenged;[43] and
- whether levels of expected attrition or lack of adherence to therapy may require a larger number of patients to achieve the desired number of person-years of followup or exposure.

Although most of the emphasis in estimating study size requirements is focused on patients, it is equally important to consider the number of sites needed to recruit and retain enough patients to achieve a reasonably informative number of person-years for analysis. Many factors are involved in choosing the number and types of sites needed for a given study, including the number of eligible patients seen in a given practice during the relevant time period, desired representativeness of sites with regard to geography, practice size, or other features, and the timeframe within which study results are required, which may also limit the timeframe for patient recruitment.

In summary, the aims of a registry, the desired precision of information sought, and the research question(s) determine the process and inputs for arriving at a target sample size and specifying the duration of followup.

Registries with mainly descriptive aims, or those that provide quality metrics for clinicians or medical centers, may not require the choice of a target study size to be arrived at through statistical power calculations. In these cases, the costs of obtaining study data, in monetary terms and in terms of researcher, clinician, and patient time and effort, may set upper as well as lower limits on study size and scope.

9. Internal and External Validity

The potential for bias refers to opportunities for systematic errors to influence the results. Internal validity is the extent to which study results are free from bias, and the reported association between exposure and outcome is not due to unmeasured or uncontrolled-for variables. Generalizability, also known as external validity, is a concept that refers to the utility of the inferences for the broader population that the study subjects are intended to represent. In considering potential biases and generalizability, we discuss the differences between RCTs and registries, since these are the two principal approaches to conducting clinically relevant prospective research.

The strong internal validity that earns RCTs high grades for evidence comes largely from the randomization of exposures that helps ensure that the groups receiving the different treatments are similar in all measured or unmeasured characteristics, and that, therefore, any differences in outcome (beyond those attributable to chance) can be likely attributed to differences in the efficacy or safety of the treatments. It should be noted that randomization does not guarantee perfect balancing of risk factors and that RCTs are not without their own biases, as illustrated by the "intent-to-treat" analytic approach, in which people are considered to have used the assigned treatment, regardless of actual adherence. The intent-to-treat analyses can minimize a real difference—generating a distortion known as "bias toward the null"—by including the experience of people who did not adhere to the recommended study product along with those who did.

Another principal difference between registries and RCTs is that RCTs generally focus on a relatively homogeneous pool of patients from which significant numbers of patients are purposefully excluded at the cost of external validity—that is, generalizability to the target population of disease sufferers. Registries, in contrast, usually focus on generalizability so that their population will be representative and relevant to decision-makers.

9.1 Generalizability

The strong external validity of registries is achieved by the fact that they include typical patients, which often include more heterogeneous populations than those participating in RCTs (e.g., wide variety of age, ethnicity, and comorbidities). Therefore, registry data can provide a good description of the course of disease and impact of interventions in actual practice. For many purposes, registries may be more relevant for decision making than the data derived from the artificial constructs of the clinical trial because registries generally represent more diverse (and more typical) medical practice as well as more diverse patients. In fact, even though registries have more opportunities to introduce bias (systematic error) because of their nonexperimental methodology, well designed observational studies can approximate the effects of interventions observed in RCTs on the same topic[44,45] and, in particular, in the evaluation of healthcare effectiveness in many instances,[46] and can provide information that may be more relevant to typical clinical practice.

The choice of groups from which patients will be selected directly affects generalizability. No particular method will ensure that an approach to patient recruitment is adequate, but it is worthwhile to note that the way in which patients are recruited, classified, and followed can

either enhance or diminish the external validity of a registry. Some examples of how these methods of patient recruitment and followup can lead to systematic error follow.

9.2 Information Bias

If the registry's principal goal is the estimation of risk, it is possible that adverse events or the number of patients experiencing them will be underreported if the reporter will be viewed negatively for reporting them. It is also possible for those collecting data to introduce bias by misreporting the outcome of an intervention if they have a vested interest in doing so. This type of bias is referred to as information bias (also called detection, observer, ascertainment, or assessment bias), and it addresses the extent to which the data that are collected are valid (represent what they are intended to represent) and accurate. This bias arises if the outcome assessment can be interfered with, intentionally or unintentionally. On the other hand, if the outcome is objective, such as whether or not a patient died or the results of a lab test, then the data are unlikely to be biased.

9.3 Selection Bias

A registry may create the incentive to enroll only patients who either are at low risk of complications or who are known not to have suffered such complications, biasing the results of the registry toward lower event rates. For example, a registry designed to assess complication rates that enrolls hospitals or surgeons who would derive benefit from reporting low complication rates would be at particularly high risk for this type of bias. Another example of how patient selection methods can lead to bias is the use of patient volunteers, a practice that may lead to selective participation from subjects most likely to perceive a benefit, distorting results for studies of patient-reported outcomes.

Enrolling patients who share a common exposure history, such as having used a drug that has been publicly linked to a serious adverse effect, could distort effect estimates for cohort and case-control analyses. Registries can also selectively enroll people who are at higher risk of developing serious side effects, since having a high-risk profile can motivate a patient to participate in a registry.

The term selection bias refers to situations where the procedures used to select study subjects lead to an effect estimate among those participating in the study that is different from the estimate that is obtainable from the target population.[47] Selection bias may be introduced if certain subgroups of patients are routinely included or excluded from the registry. Selection bias also may arise when patients must provide informed consent to participate in the registry. Some research has shown that patients who consent to participate in clinical research are different from patients who elect not to participate.[48] Depending on the registry purpose and design, some registries may be able to obtain a waiver of informed consent from an institutional review board; in these cases, data on all eligible patients are obtained, thus avoiding the potential for bias related to enrollment procedures.

9.4 Channeling Bias (Confounding by Indication)

Channeling bias, also called confounding by indication, is a form of selection bias in which drugs with similar therapeutic indications are prescribed to groups of patients with prognostic differences.[49] For example, physicians may prescribe new treatments more often to those patients who have failed on traditional first-line treatments.

One approach to designing studies to address channeling bias is to conduct a prospective review of cases, in which external reviewers are blinded as to the treatments that were employed and are asked to determine whether a particular type of therapy is indicated and to rate the overall prognosis for the patient.[50] This method of blinded prospective review was developed to support research on ruptured cerebral aneurysms, a rare and serious situation. The results of the blinded review were used to create risk strata for analysis so that comparisons could be conducted only for candidates for whom both therapies under study were indicated, a procedure much like the application of additional inclusion and exclusion criteria in a clinical trial.

For registries with sufficient data, statistical approaches such as matching subjects using propensity scores (i.e., the predicted probability of use of one therapy over another based on medical history, healthcare utilization, and other characteristics measured prior to the initiation of therapy) may be incorporated into study designs to address this type of confounding.[51,52] Propensity scores may be used to create cohorts of initiators of two different treatments matched with respect to probability of use of one of the two therapies, for stratification or for inclusion as a covariate in a multivariate analysis. Studies incorporating propensity scores as part of their design may be planned prior to and implemented shortly following launch of a new drug as part of a risk management program, with matched comparators being selected over time, so that differences in prescribing patterns following drug launch may be taken into account.[53]

Registries with large amounts of data, such as quality improvement registries, may also consider instrumental variables, or factors strongly associated with treatment but related to outcome only through their association with treatment, as an additional means of adjustment for confounding by indication, as well as unmeasured confounding.[54] Types of instrumental variables include providers' preferences for one therapy over another—a variable which exploits variation in practice as a type of natural experiment; variation or changes in insurance coverage or economic factors (e.g., cigarette taxes) associated with an exposure; or geographic distance from a specific type of service.[55-58] Variables that serve as effective instruments of this nature are not always available and may be difficult to identify. While use of clinician or study site may, in some specific cases, offer potential as an instrumental variable for analysis, the requirement that use of one therapy over another be strongly associated with the instrument is often difficult to meet in real-world settings. That said, instrumental variable analysis may either support the conclusions drawn on the basis of the initial analysis, or it may raise additional questions regarding the potential impact of confounding by indication.[51]

In some cases, however, differences in disease severity or prognosis between patients receiving one therapy rather than another may be so extreme and/or unmeasurable that confounding by indication is not remediable in an observational design that compares one group to another.[59] This represents special challenges for observational studies of comparative

effectiveness, as the severity of underlying illness may be a strong determinant of both choice of treatment and treatment outcome.

9.5 Bias From Study of Existing Rather Than New Product Users

If there is any potential for tolerance to affect the use of a product, such that only those who perceive benefit from it or are free from harm continue using it, the recruitment of existing users rather than new product users may lead to the inclusion of only those who have tolerated or benefited from the intervention, and would not necessarily capture the full spectrum of experience and outcomes. This approach is generally used with pharmacotherapy but is not as widely applicable to medical devices studies, since prior use of a medical device may not influence a patient's likelihood of tolerating it again (e.g., a patient's experience with right knee replacement may not predict experience with left knee replacement).[60] Selecting only existing users may introduce any number of biases, including incidence/prevalence bias, survivorship bias, and followup bias. By enrolling new users (an inception or incidence cohort), a study ensures that the longitudinal experience of all users will be captured, and that the ascertainment of their experience and outcomes will be comparable.[61]

9.6 Loss to Followup

Loss to followup or attrition of patients and sites threatens generalizability as well as internal validity if there is differential loss; for example, loss of participants with a particular exposure or disease, or with particular outcomes. Loss to followup and attrition are generally a serious concern only when they are nonrandom (that is, when there are systematic differences between those who leave or are lost and those who remain). The magnitude of loss to followup or attrition determines the potential impact of any bias. Given that the differences between patients who remain enrolled and those who are lost to followup are often unknown (unmeasurable), preventing loss to followup in long-term studies to the fullest extent possible will increase the credibility and validity of the results.[62] Attrition should be considered with regard to both patients and study sites, as results may be biased or less generalizable if only some sites (e.g., teaching hospitals) remain in the study while others discontinue participation.

9.7 Assessing the Magnitude and Impact of Bias

Remaining alert for any source of bias is important, and the value of a registry is enhanced by its ability to provide a formal assessment of the likely magnitude of all potential sources of bias and their impact on the study findings. Any information that can be generated regarding nonrespondents, participants lost to followup, missing data and the like, is helpful, even if it is just an estimation of their raw numbers.

As with many types of survey research, an assessment of differential response rates and patient selection can sometimes be undertaken when key data elements are available for both registry enrollees and nonparticipants or drop-outs. Such analyses can easily be undertaken when the initial data source or population pool is that of a healthcare organization, employer, or practice that has access to data in addition to key selection criteria (e.g., demographic data or data on comorbidities); these types of analyses are more challenging when registries cross health systems, institutions, and borders.

Another tool is the use of sequential screening logs, in which all subjects fitting the inclusion criteria are enumerated and a few key data elements are recorded for all those who are screened to allow some quantitative analysis of nonparticipants and assessments of any differences in key characteristics or events. Whenever possible, quantitative assessment of the likely impact of bias is desirable to determine the sensitivity of the findings to varying assumptions. A text on quantitative analysis of bias through validation studies, and on probabilistic approaches to data analysis, provides a guide for planning and implementing these methods.[63]

Qualitative assessments, although not as rigorous as quantitative approaches, may give users of the research a framework for drawing some conclusions regarding the effects of bias on study results if the basis for the assessment is made explicit in reporting the results.

Accordingly, two items that can be reported to help the user assess the generalizability of research results based on registry data are a description of the criteria used to select the registry sites, and the characteristics of these sites, particularly those characteristics that might have an impact on the purpose of the registry. Consider, for example, a registry designed for the purpose of assessing adherence to lipid screening guidelines that requires sites to have a sophisticated electronic health record for data collection. In this scenario, adherence that is better than usual practice may be reported if the electronic medical record facilitates the generation of real-time reminders to engage in screening. Report of rates of adherence to other screening guidelines (for which there were no reminders), even if these are outside the direct scope of inquiry, would provide some insight into the degree of generalizability to other types of facilities.

Finally, and most importantly, whether or not study subjects need to be evaluated on their representativeness depends on the purpose and kind of inference needed. For example, sampling in proportion to the underlying distribution in the population is not necessary to understand biological effects. However, if the study purpose were to estimate a rate of occurrence of a particular event in a general population then sampling would be necessary to reflect the appropriate underlying distributions.

10. Special Considerations

The study design considerations discussed in this chapter apply to patient registries broadly. Some types of patient registries may need to consider additional factors when determining the most appropriate study design. The following sections summarize design considerations unique to registries designed for product safety assessment, rare diseases, and medical devices as well as pregnancy registries and quality improvement registries.

10.1 Designing Registries for Product Safety Assessment[i]

Patient registries, particularly disease and product registries, that systematically collect data on all eligible patients are a tremendous resource for capturing important information about product safety. When designing a registry for safety, the size of the registry, the enrolled population, and the duration of followup are all critical to understanding the generalizability of results and applicability of the inferences made from the data. In addition, registries designed for safety must clearly define the exposure and risk windows under observation. The registry should record specific information about the products of interest, including route of administration, dose, duration of use, start and stop date, and, ideally, information about whether a generic or branded product was used (and which brand) and/or other pertinent information about the product. Studies of biologic medicines and devices benefit from including device identifiers, as well as information about production lots, and batches. Patterns of real-world product use, such as treatment switches, drug holidays, pill splitting and medication sharing, and patient non-adherence, should also be considered when designing the registry and during data collection. More information on designing registries for product safety assessment can be found in Chapter 19 of the third edition of the User's Guide. Case Example 5 also provides a description of how a registry has provided data for product safety assessments.

10.2 Designing Registries for Medical Devices[ii]

Additional issues must be considered in the design phase of medical device registries to enable the registry to function across the lifecycle of device innovation. While drugs are typically identified through National Drug Codes (NDCs), identification of devices is more complex because of the iterative cycle of device modifications. In 2013, the FDA issued a final rule establishing a unique device identification system and requiring each device to have a Unique Device Identifier (UDI) on device labels and packages.[64] A UDI is a unique alphanumeric or numeric code that contains a device identifier (describing the manufacturer and specific version or model of the device) and a production identifier (describing the lot or batch number, serial number, expiration date, manufactured date, or other distinct identification code). The FDA is requiring devices to have UDIs on a staged 7-year compliance schedule, ending in 2020. While devices are increasingly labeled with UDIs, capturing UDI data within a registry is still complicated. The FDA requirements only extend to the labeling of devices. Adoption and integration of UDIs into the healthcare delivery system is also necessary to facilitate capture of UDIs within registries. Routine and consistent capture of UDIs within electronic health records and administrative claims databases would facilitate re-use of these data for research purposes.

---

[i] *Adapted from Blackburn S, Dreyer NA, Starzyk K. 'Use of Registries in Product Safety Assessment.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014*

[ii] *Adapted from Gross T, Kuntz RE, Mack C. 'Medical Device Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014*

In addition to accurately identifying a device, medical device registries must consider how to capture factors related to device performance. These include potential performance issues, failure modes, and adverse events. The performance issues may be related to software, hardware, biomaterials, sterility, or other issues. In many cases, the device of interest for a registry is either part of a larger system of devices or contains multiple components that are considered devices themselves. Some implantable devices also require assistance from procedural devices, including other commodity devices or operative instruments, or ancillary devices, such as imaging equipment. Device/drug combinations, such as drug-eluting stents, have also become increasingly common in the past decade and necessitate separate collection of concomitant drug dosing information and attention to the medications that the patient is taking during and post implantation to flag possible drug interactions. When studying device safety or effectiveness, researchers should consider the role of these factors in device performance and how these data can be captured in the data collection process.

Lastly, provider experience and training can influence the selection of device, device performance, and patient outcomes, particularly for implantable devices. Device-specific training is an important element of a medical device registry that is not an issue in a drug registry, and experience-related factors such as practitioner annual volume, practitioner lifetime volume, facility volume, and facility characteristics such as academic teaching status should also be considered in analyses and training evaluations. It is ideal to have training and volume information in the registry, but this may not always be realistic. If this is deemed critical, information needs to be collected on provider experience and training at registry initiation and supplemented if any training programs occur during the registry development. More information on designing registries for medical devices can be found in Chapter 23 of the third edition of the User's Guide.

10.3 Designing Registries for Rare Disease[iii]

Rare diseases present special research challenges due to the scarcity of relevant knowledge and experience. Prospective long-term patient registries are critical tools in building a broad and comprehensive knowledge base for these often heterogeneous diseases. Clinicians with relevant expertise who manage patients with rare diseases are limited, and a broad approach may be necessary to identify and recruit sufficient sites and patients to characterize the natural history of the disease. In some cases, multinational efforts may be necessary to enroll sufficient patients.

Registry design is also complicated by the absence of treatments or standards of care for many rare diseases. Use of experimental and adjunctive therapies is common, and it is often unclear how to characterize disease progression, especially start and stop dates of exacerbations. In addition to the typical objectives for disease registries (understand natural history and outcomes, assess effectiveness of treatments, etc.), rare disease registries may be designed to support the drug development process. In these cases, registries may be designed to recruit a readily

[iii] *Adapted from Dale DC, Groft SC, Harrison MJ, et al. 'Rare Disease Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

available pool of patients for potential enrollment into clinical trials, gather baseline data that can inform trial design, and provide external reference groups for understanding potential treatment benefit and risks. Because of the scarcity of eligible patients, patient advocacy and support groups often play more active roles in rare disease registries than in more traditional disease registries;[65] in some cases, these organizations may sponsor registries, or they may work with other sponsors as active partners in the development and operations of the registry. Lastly, the scope of rare disease registries frequently evolves over time, as understanding of the disease increases and/or new treatments become available. More information on designing registries for rare diseases can be found in Chapter 20 of the third edition of the User's Guide.

10.4 Designing Pregnancy Registries[iv]

Pregnancy registry design differs from the design of other types of registries in several respects. First, the population of a pregnancy registry can be defined based on women, pregnancies, or fetuses. A woman might have more than one pregnancy, and she might enroll in the same registry more than once. Clustered analyses are often used in this situation. In addition, multifetal gestations result in more than one fetus "enrolled" within the same pregnancy. Although there may be several ways of dealing with multiple gestations, it is prudent to collect information about all the fetuses. When reporting risks, whether using fetuses or pregnancies as the unit of analysis, both the numerator and denominator should be consistent with the choice. Registries should include women as soon as possible after conception, or even earlier at pregnancy planning stages, to allow the evaluation of early pregnancy events, and women should be enrolled before the pregnancy outcome is known to avoid a selection into the study affected by the outcome. An ideal pregnancy cohort would enroll women at conception and follow them for months beyond delivery, but, in practice, time from enrollment to end of followup can range from 1 month to over 1 year. As with any registry, longer followup periods lead to higher opportunities for diagnosis and therefore both larger cumulative risk estimates and greater statistical power. However, the length of followup may be influenced by the availability of resources and the registry's ability to maintain contact with registry participants and/or healthcare providers over a longer term. It is also very difficult to enroll women early in pregnancy, and differential enrollment may occur according to whether it is a woman's first pregnancy or she has been pregnant before.[66]

Necessary information on exposure, outcome, and key confounders (e.g., history, status, severity, and management of the indication) must be collected. Because treatment strategies often change during pregnancy, detailed information should be collected on treatment start and stop dates, dose, frequency, duration, and indication. Consideration should also be given to the source (mother, obstetrician, pediatrician) of information on outcomes and the potential for selective recall. A method for expert adjudication of birth defect classification, blinded to exposure status, is an important component of a pregnancy registry. In addition, the case of major birth defects occurring in pregnancies that end in embryonic or fetal demise must be considered in the registry

---

[iv] *Adapted from Chambers C, Covington D, Cragan JD, et al. 'Pregnancy Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

design. Failure to include defects detected among terminations can decrease power and introduce bias, particularly for defects for which termination is often chosen after prenatal diagnosis (e.g., neural tube defects).

A critical element for pregnancy exposure registries is the choice of comparator groups. The most valid reference group will have comparable (1) outcome definition (e.g., exclusion of minor anomalies); (2) outcome assessment (e.g., intensity of screening, frequency of terminations, inclusion of prenatal diagnoses, availability of diagnostic tests, start and stop of followup);[67] (3) selection of subjects into the study (e.g., gestational age at enrollment); and (4) baseline risk (e.g., distribution of risk factors, including indication). Ideally, each registry is constructed to include one or more internal reference groups, though this is not common practice. When this is not possible, an external reference group must be selected with care. Each comparison group has its advantages and disadvantages. For example, an external population-based reference group is generally larger and can provide more stable estimates for specific malformations, while an internal comparison group, which may be too small to support assessment of specific malformations, may be able to provide more comparable estimates for malformations overall. More than one comparison group can be used to enhance generalizability. More information on designing pregnancy registries can be found in Chapter 21 of the third edition of the User's Guide. Case Example 6 also provides a description of a long-running pregnancy registry.

10.5 Designing Quality Improvement Registries[v]

Designing a quality improvement (QI) registry presents several challenges, particularly when multiple stakeholders are involved. Like other types of registries, design of QI registries should be purpose-driven. This purpose may require detailed data collection at a single point in time (e.g., to improve care for patients hospitalized with acute coronary syndrome) or long-term followup data from different providers (e.g., to monitor care for patients with coronary artery disease). QI registries may focus on issues within a single institution, or they may address common treatment gaps that are relevant at many institutions.

A unique and critical component of QI registries are quality measures. Quality measures are tools that quantify healthcare processes or outcomes and are designed to help institutions and providers deliver high-quality care that aligns with clinical guidelines or best practices. Quality measures drive the registry data collection and reporting and thus form the backbone of a QI registry. Since QI registries are part of healthcare operations, it is critical that they do not overly interfere with the efficiency of those operations, and therefore the data collection must be limited to those data elements that are essential for calculating the relevant quality measures. In addition, the appropriate level of analysis and reporting of quality measures is an important consideration in QI registry design. Reports on compliance with quality measures may provide data at the individual patient, provider, or institution level, or they may provide aggregate data on groups of patients, providers, and institutions. The registry may also provide reports to the registry

[v] *Adapted from Asher AL, Gliklich RE, Hernandez AF, et al. 'Quality Improvement Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

participants, to patients, or to the public. Reports may be unblinded (e.g., the provider is identifiable) or blinded, and they may be provided through the registry or through other means. In designing the registry, consideration should be given to what types of reports will be most relevant for achieving the registry's goals, what types of reports will be acceptable to participants, and how those reports should be presented and delivered. More information on designing quality improvement registries can be found in Chapter 22 of the third edition of the User's Guide. Case Examples 8 and 11 also describe quality improvement registries.

10.6 Designing Multinational Registries

In cases where the registry intends to collect data in more than one country, it is desirable to gather input from clinicians and patients in the other country (or countries) to understand potential variations in treatment patterns and data elements. Treatment patterns often vary across geographic regions due to multiple factors, including differences in approved indications, coverage decisions, and clinical guidelines. Products may be approved for different indications in different countries or regions, which can lead to the use of the product by patients with different characteristics, including varying levels of severity of conditions in each country or region. For example, natalizumab is approved in the European Union (EU) for patients who have failed two or more therapies for relapsing-remitting multiple sclerosis, while, in the United States, the therapy is used more widely. Differences in health insurance coverage decisions may affect treatment patterns in a similar manner; access and reimbursement levels may differ among countries, which can impact providers' and patients' ability and willingness to use a specific product. The use of different clinical guidelines also can have a substantial impact on treatment patterns. The American Gastroenterological Association, for example, recommends annual or biannual colonoscopic surveillance for neoplasia in patients with inflammatory bowel disease-related colitis, depending on whether patients are considered high risk or average risk. In contrast, the British Society of Gastroenterology recommends colonoscopic surveillance on a 1-year, 3-year, or 5-year basis, depending on risk assessment.[68]

Registries implemented in multiple countries must plan for these types of differences in standard of care. Because registries are observational, additional diagnostic or monitoring procedures such as laboratory tests are not undertaken unless they are within the scope of normal practice. Combined with differences in national guidelines, policies, and regulations, this makes variation in data availability commonplace for multinational registries.

11. Summary

In summary, the key points to consider in designing a registry include study design, selection of patients and healthcare practitioners, data collection, comparison groups, recruitment strategies, and considerations of possible sources of bias, their likely impact, and ways to address them to the extent that is practical and achievable. Additional design considerations apply for some specialized types of patient registries.

Lastly, it is important to keep in mind that it may be necessary to revisit the registry design if it becomes apparent that the initial plan will not meet expectations. For example, the original criteria for defining the target population (patients and/or healthcare providers) may not yield

enough patients, such as when a treatment of interest is only slowly coming into use for the intended population or in the sites that have been recruited for study; moreover, recruitment can be more difficult if the treatment or product of interest is not covered by health insurers. More information on modifying patient registries can be found in Chapters 2 and 11.

References for Chapter 3

1. Strom BL. Pharmacoepidemiology. 3rd ed. Chichester, England: John Wiley; 2000.

2. Gliklich R. A New Framework For Comprehensive Evidence Development. IN VIVO. October 22, 2012. https://invivo.pharmaintelligence.informa.com/IV003921/A-New-Framework-For-Comprehensive-Evidence-Development. Accessed June 10, 2019.

3. James S, Rao SV, Granger CB. Registry-based randomized clinical trials--a new clinical trial paradigm. Nat Rev Cardiol. 2015;12(5):312-6. PMID: 25781411. DOI: 10.1038/nrcardio.2015.33.

4. Li G, Sajobi TT, Menon BK, et al. Registry-based randomized controlled trials- what are the advantages, challenges, and areas for future research? J Clin Epidemiol. 2016;80:16-24. PMID: 27555082. DOI: 10.1016/j.jclinepi.2016.08.003.

5. Gliklich RE, Leavy MB, Dreyer NA, (eds). Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide. 3rd Edition, Addendum II. (Prepared by L&M Policy Research, LLC under Contract No. 290-2014-00004-C.) AHRQ Publication No. 17(18)-EHC017-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2019. https://effectivehealthcare.ahrq.gov/.

6. Lauer MS, D'Agostino RB, Sr. The randomized registry trial--the next disruptive technology in clinical research? N Engl J Med. 2013;369(17):1579-81. PMID: 23991657. DOI: 10.1056/NEJMp1310102.

7. Rao SV, Hess CN, Barham B, et al. A registry-based randomized trial comparing radial and femoral approaches in women undergoing percutaneous coronary intervention: the SAFE-PCI for Women (Study of Access Site for Enhancement of PCI for Women) trial. JACC Cardiovasc Interv. 2014;7(8):857-67. PMID: 25147030. DOI: 10.1016/j.jcin.2014.04.007.

8. U.S. Food and Drug Administration. Adaptive Designs for Clinical Trials of Drugs and Biologics. Draft Guidance for Industry. October 2018. https://www.fda.gov/downloads/drugs/guidances/ucm201790.pdf. Accessed June 10, 2019.

9. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. Biometrics. 1995;51(1):228-35. PMID: 7766778.

10. Ryan PB, Powell GE, Pattishall EN, et al. Performance of screening multiple observational databases for active drug safety surveillance. International Society of Pharmacoepidemiology; Providence, RI, USA. 2009.

11. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46(1):348-55. PMID: 27283160. DOI: 10.1093/ije/dyw098.

12. Norén GN, Hopstadius J, Bate A, et al. Temporal pattern discovery in longitudinal electronic patient records. Data Mining and Knowledge Discovery. 2009;20(3):361-87. DOI: 10.1007/s10618-009-0152-3.

13. Travis LB, Rabkin CS, Brown LM, et al. Cancer survivorship--genetic susceptibility and second primary cancers: research strategies and recommendations. J Natl Cancer Inst. 2006;98(1):15-25. PMID: 16391368. DOI: 10.1093/jnci/djj001.

14. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology. Boston: Little, Brown and Company; 1985. p. 228.

15. Hennekens CH, Buring JE, Mayrent SL. Epidemiology in medicine. 1st ed. Boston: Little, Brown and Company; 1987.

16. Schoebel FC, Gradaus F, Ivens K, et al. Restenosis after elective coronary balloon angioplasty in patients with end stage renal disease: a case-control study using quantitative coronary angiography. Heart. 1997;78(4):337-42. PMID: 9404246. DOI: 10.1136/hrt.78.4.337.

17. Rothman K, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p. 175–9.

18. Oral contraceptive use and the risk of endometrial cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA. 1983;249(12):1600-4. PMID: 6338265.

19. Oral contraceptive use and the risk of ovarian cancer. The Centers for Disease Control Cancer and Steroid Hormone Study. JAMA. 1983;249(12):1596-9. PMID: 6338264.

20. Long-term oral contraceptive use and the risk of breast cancer. The centers for Disease Control Cancer and Steroid Hormone Study. JAMA. 1983;249(12):1591-5. PMID: 6338262.

21. Speck CE, Kukull WA, Brenner DE, et al. History of depression as a risk factor for Alzheimer's disease. Epidemiology. 1995;6(4):366-9. PMID: 7548342.

22. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. Am J Epidemiol. 1992;135(9):1019-28. PMID: 1595688. DOI: 10.1093/oxfordjournals.aje.a116396.

23. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. III. Design options. Am J Epidemiol. 1992;135(9):1042-50. PMID: 1595690. DOI: 10.1093/oxfordjournals.aje.a116398.

24. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. II. Types of controls. Am J Epidemiol. 1992;135(9):1029-41. PMID: 1595689. DOI: 10.1093/oxfordjournals.aje.a116397.

25. Rothman K, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p. 108.

26. Vigneswaran R, Aitchison SJ, McDonald HM, et al. Cerebral palsy and placental infection: a case-cohort study. BMC Pregnancy Childbirth. 2004;4(1):1. PMID: 15005809. DOI: 10.1186/1471-2393-4-1.

27. Ong AT, Daemen J, van Hout BA, et al. Cost-effectiveness of the unrestricted use of sirolimus-eluting stents vs. bare metal stents at 1 and 2-year follow-up: results from the RESEARCH Registry. European heart journal. 2006;27(24):2996-3003. PMID: 17114234. DOI: 10.1093/eurheartj/ehl357.

28. Farfel M, DiGrande L, Brackbill R, et al. An overview of 9/11 experiences and respiratory and mental health conditions among World Trade Center Health Registry enrollees. J Urban Health. 2008;85(6):880-909. PMID: 18785012. DOI: 10.1007/s11524-008-9317-4.

29. E-Cigarette Effects on Markers of Cardiovascular and Pulmonary Disease ClinicalTrials.gov. NCT03863509. https://clinicaltrials.gov/ct2/show/NCT03863509. Accessed June 10, 2019.

30. Hulley SB, Cumming SR. Designing clinical research. Baltimore: Williams & Wilkins; 1988.

31. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013;42(4):1012-4. PMID: 24062287. DOI: 10.1093/ije/dys223.

32. Hunter D. First, gather the data. N Engl J Med. 2006;354(4):329-31. PMID: 16436764. DOI: 10.1056/NEJMp058235.

33. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006;354(4):353-65. PMID: 16436767. DOI: 10.1056/NEJMoa051379.

34. Ripollone JE, Huybrechts KF, Rothman KJ, et al. Implications of the Propensity Score Matching Paradox in Pharmacoepidemiology. Am J Epidemiol. 2018;187(9):1951-61. PMID: 29750409. DOI: 10.1093/aje/kwy078.

35. Jackson JW, Schmid I, Stuart EA. Propensity Scores in Pharmacoepidemiology: Beyond the Horizon. Curr Epidemiol Rep. 2017;4(4):271-80. PMID: 29456922. DOI: 10.1007/s40471-017-0131-y.

36. National Cancer Institute. Surveillance Epidemiology and End Results. https://seer.cancer.gov/. Accessed June 10, 2019.

37. Becker JC, Lorenz E, Ugurel S, et al. Evaluation of real-world treatment outcomes in patients with distant metastatic Merkel cell carcinoma following second-line chemotherapy in Europe. Oncotarget. 2017;8(45):79731-41. PMID: 29108353. DOI: 10.18632/oncotarget.19218.

38. Metropolitan Atlanta Congenital Defects Program (MACDP). National Center on Birth Defects and Developmental Disabilities. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/birthdefects/macdp.html. Accessed June 10, 2019.

39. Hatfield L, Zusterzeel R, Daluwatte C, Normand SL. Improving Access To Medical Devices: The Use And Evolution Of Objective Performance Criteria. Health Affairs Blog. July 26, 2018. https://www.healthaffairs.org/do/10.1377/hblog20180726.907775/full/. Accessed June 10, 2019.

40. U.S. Food and Drug Administration; Center for Devices and Radiological Health. Design Considerations for Pivotal Clinical Investigations for Medical Devices. Guidance for Industry, Clinical Investigators, Institutional Review Boards and Food and Drug Administration Staff. November 2013. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/design-considerations-pivotal-clinical-investigations-medical-devices. Accessed June 10, 2019.

41. Cochran WG. Sampling Techniques. 3rd ed. New York: John Wiley & Sons; 1977.

42. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med. 1971;284(15):878-81. PMID: 5549830. DOI: 10.1056/NEJM197104222841604.

43. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1(1):43-6. PMID: 2081237.

44. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342(25):1887-92. PMID: 10861325. DOI: 10.1056/NEJM200006223422507.

45. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med. 2000;342(25):1878-86. PMID: 10861324. DOI: 10.1056/NEJM200006223422506.

46. Black N. Why we need observational studies to evaluate the effectiveness of health care. BMJ. 1996;312(7040):1215-8. PMID: 8634569. DOI: 10.1136/bmj.312.7040.1215.

47. Rothman K. Modern Epidemiology. Boston: Little Brown and Company; 1986. p. 83.

48. Kho ME, Duffett M, Willison DJ, et al. Written informed consent and selection bias in observational studies using medical records: systematic review. BMJ. 2009;338:b866. PMID: 19282440. DOI: 10.1136/bmj.b866.

49. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med. 1991;10(4):577-81. PMID: 2057656.

50. Johnston SC. Identifying confounding by indication through blinded prospective review. Am J Epidemiol. 2001;154(3):276-84. PMID: 11479193. DOI: 10.1093/aje/154.3.276.

51. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98(3):253-9. PMID: 16611199. DOI: 10.1111/j.1742-7843.2006.pto_293.x.

52. Sturmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006;59(5):437-47. PMID: 16632131. DOI: 10.1016/j.jclinepi.2005.07.004.

53. Loughlin J, Seeger JD, Eng PM, et al. Risk of hyperkalemia in women taking ethinylestradiol/drospirenone and other oral contraceptives. Contraception. 2008;78(5):377-83. PMID: 18929734. DOI: 10.1016/j.contraception.2008.06.012.

54. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf. 2010;19(6):537-54. PMID: 20354968. DOI: 10.1002/pds.1908.

55. Evans W, Ringel J. Can higher cigarette taxes improve birth outcomes? J Public Economics. 1997;72(1):135-54. DOI: 10.3386/w5998.

56. Schneeweiss S, Seeger JD, Landon J, et al. Aprotinin during coronary-artery bypass grafting and risk of death. N Engl J Med. 2008;358(8):771-83. PMID: 18287600. DOI: 10.1056/NEJMoa0707571.

57. Secemsky EA, Kirtane A, Bangalore S, et al. Use and Effectiveness of Bivalirudin Versus Unfractionated Heparin for Percutaneous Coronary Intervention Among Patients With ST-Segment Elevation Myocardial Infarction in the United States. JACC Cardiovasc Interv. 2016;9(23):2376-86. PMID: 27838271. DOI: 10.1016/j.jcin.2016.09.020.

58. Vertullo CJ, Graves SE, Peng Y, et al. The effect of surgeon's preference for hybrid or cemented fixation on the long-term survivorship of total knee replacement. Acta Orthop. 2018;89(3):329-35. PMID: 29528754. DOI: 10.1080/17453674.2018.1449466.

59. Bosco JL, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. J Clin Epidemiol. 2010;63(1):64-74. PMID: 19457638. DOI: 10.1016/j.jclinepi.2009.03.001.

60. Dreyer NA, Velentgas P, Westrich K, et al. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. J Manag Care Spec Pharm. 2014;20(3):301-8. PMID: 24564810. DOI: 10.18553/jmcp.2014.20.3.301.

61. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003;158(9):915-20. PMID: 14585769. DOI: 10.1093/aje/kwg231.

62. Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? Eur J Epidemiol. 2004;19(8):751-60. PMID: 15469032.

63. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data: Springer; 2009.

64. U.S. Food and Drug Administration. Unique Device Identification System Final Rule. September 24, 2013. https://www.federalregister.gov/documents/2013/09/24/2013-23059/unique-device-identification-system. Accessed June 10, 2019.

65. Gliklich RE, Dreyer NA, Leavy MB, Christian JB (eds). 21st Century Patient Registries. EBook addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. (Prepared by L&M Policy Research, LLC under Contract No. 290-2014- 00004-C.) AHRQ Publication No. 17(18)-EHC013-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018. www.effectivehealthcare.ahrq.gov.

66. Dreyer NA, Blackburn SC, Mt-Isa S, et al. Direct-to-Patient Research: Piloting a New Approach to Understanding Drug Safety During Pregnancy. JMIR Public Health Surveill. 2015;1(2):e22. PMID: 27227140. DOI: 10.2196/publichealth.4939.

67. Cragan JD, Gilboa SM. Including prenatal diagnoses in birth defects monitoring: Experience of the Metropolitan Atlanta Congenital Defects Program. Birth Defects Res A Clin Mol Teratol. 2009;85(1):20-9. PMID: 19089857. DOI: 10.1002/bdra.20508.

68. Lutgens M, van Oijen M, Mooiweer E, et al. A risk-profiling approach for surveillance of inflammatory bowel disease-colorectal carcinoma is more cost-effective: a comparative cost-effectiveness analysis between international guidelines. Gastrointest Endosc. 2014;80(5):842-8. PMID: 25088918. DOI: 10.1016/j.gie.2014.02.1031.

Case Examples for Chapter 3

Case Example 5. Using a Registry To Assess Long-Term Product Safety

Description	The British Society for Rheumatology Biologics Registers in Rheumatoid Arthritis (BSRBR-RA) is a prospective observational study conducted to monitor the routine clinical use and long-term safety of biologics (including biosimilars) and other targeted therapies in patients with rheumatoid arthritis and other rheumatic conditions.
Sponsor	Research Governance Sponsor: The University of Manchester Funder: The British Society for Rheumatology (BSR)
Year Started	2001
Year Ended	Ongoing
No. of Sites	All consultant rheumatologists in the UK who have prescribed anti-TNF and other targeted therapies have an opportunity to participate in the register.
No. of Patients	More than 30,000

Challenge

Rheumatoid arthritis (RA) is a progressive inflammatory disease characterized by joint damage, pain, and disability. Among the pharmacologic treatments, nonbiologic disease-modifying antirheumatic drugs (DMARDs) are considered the first-line treatment. Biologic therapies were introduced approximately 20 years ago and offered patients and providers a new class of agents with demonstrated efficacy in RA patients. The most commonly used biologics are tumor necrosis factors (TNF) inhibitors (etanercept, infliximab, adalimumab, certolizumab, and golimumab), although the use of other classes of advanced therapies, including andi-CD20 (rituximab), anti-IL6 (tocilizumab, sarilumab), and JAK inhibitors (baricitinib, tofacitinib) is also increasing. However, results from clinical trials and pharmacovigilance studies raised potential safety concerns, and limited long-term data on these therapies are available at the time of regulatory approval. Of particular concern is the risk of serious infections including tuberculosis and malignancy.

Proposed Solution

A prospective observational registry was launched in 2001 to monitor the safety and effectiveness of biologic treatments. This United Kingdom-wide national project was launched after the introduction of the first tumor necrosis factors (TNF) alpha inhibitors and has now expanded to include therapies across a wide range of biologic and targeted therapies, including the recent inclusion of biosimilar drugs. The registry collects data on response to treatment and adverse events (AEs) every six months, and patients are followed for the life of the registry. In addition to patients receiving biologic and targeted therapies, the registry has enrolled a control cohort of patients receiving nonbiologic DMARDs, although recruitment to this cohort (n=3800) ended in 2008.

Results

The registry has now published over 60 papers looking at a wide range of outcomes, including treatment effectiveness and safety, such as the risk of infections, malignancy, cardiovascular disease, and thromboembolic disease. Details of publications, study protocols and further information can be found on the registry website. Datasets can be made available to researchers upon approval of an application, and information on how to apply for a dataset may found on the registry website.

Key Point

As novel drugs and treatments are developed and licensed, registries may be useful tools for collecting long-term data to assess known and emerging safety concerns.

For More Information

- www.bsrbr.org
- Nikiphorou E, Buch MH, Hyrich KL. Biologics registers in RA: methodological aspects, current role and future applications. Nat Rev Rheumatol. 2017 Aug;13(8):503-510. doi: 10.1038/nrrheum.2017.81. PMID: 28569267. DOI: 10.1038/nrrheum.2017.81.

Case Example 6. Expanding an Ongoing Pregnancy Registry

Description	The Antiretroviral Pregnancy Registry is the oldest ongoing pregnancy exposure registry. This multisponsor, international, voluntary, collaborative registry monitors prenatal exposures to all marketed antiretroviral drugs for potential risk of birth defects.
Sponsors	AbbVie, Accord Healthcare Inc., Alvogen Inc., Amneal Pharmaceuticals LLC, Apotex Inc., Aurobindo Pharma Ltd., Boehringer Ingelheim Pharmaceuticals Inc., Bristol-Myers Squibb Company, Celltrion, Inc., Cipla Ltd., F. Hoffman La-Roche, Gilead Sciences Inc., Hetero Labs Ltd., Hikma Pharmaceuticals USA Inc., Janssen Scientific Affairs, Lannett Company Inc., Lupin Pharmaceuticals, Macleods Pharmaceuticals Ltd., Merck & Co. Inc., Mylan Laboratories, Novartis Pharmaceuticals, Prinston, Qilu Pharmaceutical Co, Ltd., Sandoz Inc, SigmaPharm Laboratories LLC, Strides Shasun Ltd., Teva Pharmaceuticals USA Inc., and ViiV Healthcare
Year Started	1989
Year Ended	Ongoing
No. of Sites	Not site-based; open to all healthcare providers.
No. of Patients	20,375

Challenge

Antiretroviral treatments represent an area of particular concern for monitoring safety in pregnancy. Women may need to take the drugs during pregnancy to manage their own HIV infection and to reduce the risk of transmitting HIV to the infant, but these benefits must be

weighed against the risk of teratogenic effects. Because of these factors, it is extremely important for clinicians and patients to understand the risks of using antiretroviral drugs during pregnancy in order to make an informed decision. However, ethical and practical concerns make a randomized trial to gather these data difficult, if not impossible.

In 1989, the first manufacturer of an antiretroviral drug voluntarily initiated a pregnancy exposure registry to track the outcomes of women who had used its product during pregnancy. The purpose of the registry is to collect information on any teratogenic effects of the product by prospectively enrolling women during the course of their pregnancy and following up with them to determine the outcome of the pregnancy. Physicians enroll a patient by providing information on the pregnancy dates, characteristics of the HIV infection, drug dosage, length of therapy, and trimester of exposure to the antiretroviral drug. Information on the pregnancy outcome is gathered through a followup form sent to the physician after the expected delivery date.

In 1993, the registry was expanded to include all antiretroviral drugs, as other manufacturers voluntarily joined the registry once their drugs were on the market. The registry is international in scope and allows any healthcare provider to enroll a patient who intentionally or unintentionally has used an antiretroviral drug during pregnancy. The U.S. Food and Drug Administration, which has used this registry as a model for new pregnancy registries, now requires all new and generic antiretroviral drugs marketed in the U.S to be monitored in a registry.

The year 2019 marks 30 years of active enrollment with the registry expanding to now monitor 153 antiretroviral drugs including 55 brand-name single-entity or fixed-dose combinations and 98 generic versions from 28 companies. The registry has also increased enrollment as well as its geographic representation by incorporating the datasets of comparable, completed epidemiological studies. For example, the registry added data on nearly 1,000 women from a study conducted in Brazil and Argentina of antiretroviral-exposed pregnant women who delivered between the years 2002 and 2007. In addition, electronic data capture (EDC) was introduced in 2010 as a data collection method for the registry.

In summary, early challenges for the registry included establishing standard processes for monitoring and assessing the safety of drugs during pregnancy. Key challenges in recent years have included managing the methodological and analytic implications of a rapid growth in size, complex drug regimens and the operational implications of long-term EDC system management.

Proposed Solution

To ensure both rigor and consistency early on, the registry put in place predefined analytic methods and criteria for recognizing a potential teratogenic signal. Tools for coding and classifying birth defects were developed for the registry to maximize the likelihood of identifying a teratogenic signal. This unique system groups birth defects by etiology or embryology rather than by general location or category, as does the Medical Dictionary for Regulatory Activities (MedDRA). Grouping like defects together increases the likelihood of detecting a potential signal. The registry also codes the temporal association between timing of exposure and formation of the birth defect, aiding in signal detection.

Specific monitoring criteria were developed for evaluating signals at various levels, including the Rule of Three (the rule that three exposure-specific cases with the same birth defect require immediate evaluation). This rule is based on the statistical principle that the likelihood of finding at least three of any specific defect in a cohort of 600 or fewer by chance alone is less than 5 percent.

More recently, large increases in enrollment required re-evaluation of the adequacy of existing signal detection rules. The Rule of Three continues to serve an important role; however, understanding weak signals is methodologically challenging. Incorporating enrollments from comparable epidemiological studies into the registry population has boosted enrollment, increased cultural diversity, and enhanced signal detection capabilities. Each merger of external data prompts the need to re-examine the potential for selection and ascertainment bias.

Operationally, each new participating manufacturer undergoes a series of trainings and is required to obtain institutional review board approval before participation in the registry. Registry trainings and standard operating procedures are reviewed at biannual steering committee meetings and revised as appropriate.

In expanding the options for data entry into the registry, a hybrid EDC-paper approach was deemed operationally feasible in lieu of an EDC-only approach. This allowed a subset of established reporters to use EDC, while limiting disruption for reporters who preferred to report data on paper CRFs.

Results

The registry now contains data on 20,375 prospective pregnancies with exposure to 55 medications. Registry data have been used in 15 publications, 15 presentations, and more than 40 conference abstracts and posters, and the registry design and operation have been the subject of many publications and presentations. The registry findings can help provide clinicians and patients with information to make informed decisions regarding use of antiretroviral drugs during pregnancy.

Key Point

A pregnancy exposure registry can employ continuous quality improvement practices to identify and define key quality processes and keep the registry current and innovative throughout its life cycle. The fact that the registry had established, standard policies and procedures for coding, monitoring, and analysis was critical in incorporating new partners and data sources quickly and easily. Regular review of these policies and procedures is essential to respond to the changing registry environment.

For More Information

- Antiretroviral Pregnancy Registry Steering Committee. Antiretroviral Pregnancy Registry International Interim Report for 1 January 1989 through 31 January 2019. Wilmington, NC: Registry Coordinating Center; 2019. http://www.apregistry.com/. Accessed June 10, 2019.
- Tilson H, Roberts S, Watts H, et al. The Antiretroviral Pregnancy Registry: A 20th anniversary celebration. Pharmacoepidemiology and Drug Safety. 2011;20(S1):S190. PMID: 17229330. DOI: 10.1097/01.ogx.0000253377.14647.80.

- Tilson H, Doi PA, Covington DL, et al. The antiretrovirals in pregnancy registry: A fifteenth anniversary celebration. Obstet Gynecol Surv. 2007;62:137–48. PMID: 17229330. DOI: 10.1097/01.ogx.0000253377.14647.80.
- Covington D, Tilson H, Elder J, et al. Assessing teratogenicity of antiretroviral drugs: monitoring and analysis plan of the Antiretroviral Pregnancy Registry. Pharmacoepidemiol Drug Saf. 2004;13:537–45. PMID: 15317035. DOI: 10.1002/pds.982.
- Scheuerle A, Covington D. Clinical review procedures for the Antiretroviral Pregnancy Registry. Pharmacoepidemiol Drug Saf. 2004;13:529–36. PMID: 15317034 DOI: 10.1002/pds.971.

Case Example 7. Designing a Registry To Address Unique Patient Enrollment Challenges

Description	The Anesthesia Awareness Registry is a survey-based registry that collects detailed data about patient experiences of anesthesia awareness. Patient medical records are used to assess anesthetic factors associated with the patient's experience. An optional set of psychological assessment instruments measure potential trauma-related sequelae including depression and post-traumatic stress disorder (PTSD).
Sponsor	American Society of Anesthesiologists
Year Started	2007
Year Ended	Ongoing
No. of Sites	Not applicable
No. of Patients	366

Challenge

Anesthesia awareness is a recognized complication of general anesthesia, defined as the unintended experience and explicit recall of events during surgery. The incidence of anesthesia awareness has been estimated at 1–2 patients per 1,000 anesthetics and may result in development of serious and long-term psychological sequelae including PTSD. The causes of the phenomenon and preventive strategies have been studied, but there is disagreement in the scientific community about the effectiveness of monitoring devices for prevention of anesthesia awareness.

The population of patients experiencing anesthesia awareness is difficult to identify. Although standard short questionnaires designed to identify anesthesia awareness are sometimes administered to patients postoperatively, many patients experience delayed recollection and do not realize that they were awake during their procedure until several weeks later. These patients may or may not report their experience to their provider. In addition, because of the often unsettling and traumatic nature of their experience, even patients who recognize their anesthesia awareness before being discharged from the hospital may not feel comfortable reporting it to their surgeon or other healthcare providers.

With ongoing coverage in the media, anesthesiologists were facing increasing concern and fear about anesthesia awareness among their patients. The American Society of Anesthesiologists sought a patient-oriented approach to this problem.

Proposed Solution

Because this population of patients is not always immediately recognized in the healthcare setting, the registry was created to collect case reports of anesthesia awareness directly from patients. A patient advocate was invited to consult in the registry's development and provides ongoing advice from the patient perspective. The registry hosts a website that provides information about anesthesia awareness and directions for enrolling in the registry. Any patient who believes they have experienced anesthesia awareness may voluntarily submit a survey and medical records to the registry. Psychological assessments are optional. An optional open-ended discussion about the patient's anesthesia awareness experience provides patients with an opportunity to share information that may not be elicited through the survey.

Results

The registry has enrolled 366 patients since 2007. Patients who enroll are self-selected, and the sample is likely biased towards patients with emotional sequelae. While the information provided to potential enrollees clearly states that eligibility is restricted to awareness during general anesthesia, a surprising number of enrollments are patients who were supposed to be awake during regional anesthesia or sedation. This revealed a different side to the problem of anesthesia awareness: clearly, some patients did not understand the nature of the anesthetic that would be provided for their procedure, or patients had expectations that were not met by their anesthesia providers. Most enrollees experienced long-term psychological sequelae regardless of anesthetic technique.

Key Point

Allowing the registry's purpose to drive its design produces a registry that is responsive to the expected patient population. Employing direct-to-patient recruitment can be an effective way of reaching a patient population that otherwise would not be enrolled in the registry, and can yield surprising and important insights into patient experience.

For More Information

- http://www.awaredb.org
- Domino KB. Committee on Professional Liability opens anesthesia awareness registry. ASA Newsletter. 2007. p. 29.p. 34.
- Kent CD, Metzger NA, Posner KL, et al. Anesthesia Awareness Registry: psychological impacts for patients. Anesthesiology. 2011:A003.
- Domino KB, Metzger NA, Mashour GA. Anesthesia Awareness Registry: patient responses to awareness. Br J Anaesth. 2012;108(2):338P.
- Kent CD, Mashour GA, Metzger NA, et al. Psychological impact of unexpected explicit recall of events occurring during surgery performed under sedation, regional anaesthesia, and general anaesthesia: data from the Anesthesia Awareness Registry. Br J Anaesth. 2013;110(3):381-7. PMID: 23161356. DOI: 10.1093/bja/aes386.

- Kent CD, Posner KL, Mashour GA, et al. Patient perspectives on intraoperative awareness with explicit recall: report from a North American anaesthesia awareness registry. Br J Anaesth. 2015;115 Suppl 1:i114-i21. PMID: 26174296. DOI: 10.1093/bja/aev211.

Case Example 8. Using Registries To Drive Quality Improvement in Chronic Conditions

Description	The National Parkinson Foundation Quality Improvement Initiative is a registry-based quality care program that captures longitudinal data on clinical interventions and patient-reported outcomes to identify, implement, and disseminate best practices for the treatment and management of Parkinson's disease.
Sponsor	National Parkinson Foundation
Year Started	2009
Year Ended	Ongoing
No. of Sites	21 sites in North America, the Netherlands, and Israel
No. of Patients	>8,000 patients

Challenge

Parkinson's disease (PD), an incurable, progressive neurogenerative disorder associated with a high burden of disease, presents unique challenges for quality improvement initiatives. Treatments for PD generally focus on reducing patients' symptoms and improving quality of life. Unlike other chronic conditions where improvement can be measured in terms of well-defined outcomes such as survival or cardiovascular events, quality improvement in PD can best be measured using patient-based outcomes. However, identifying appropriate patient-based outcomes for this disease can be a challenge. In addition, variability exists in the clinical diagnosis, management, and treatment of PD. Studies have shown that PD patients treated by a neurologist experience better outcomes, such as a decrease in hip fractures or nursing home placement. However, the specific management and treatment strategies used by these specialists have not been studied or well-described. The lack of evidence-based treatment standards warranted a data-driven approach to identify and understand best practices that improve the quality of care and quality of life for PD patients.

Proposed Solution

In 2009, the National Parkinson Foundation launched an initiative to improve the quality of care in PD. To support an evidence-based approach, the foundation initiated a PD registry to capture clinical interventions and patient-reported outcomes over time from multiple centers across the United States, Canada, and internationally. The initiative, led by a steering committee of movement disorders neurologists, is a unique effort in PD research because of its ability to collect long-term, longitudinal data from multiple centers and its focus on patient-based outcomes data, rather than process of care measures. The aims of the registry are to accelerate clinical discovery, promote collaborative science, and drive advancements in clinical practice toward patient-centered care.

Results

The registry includes data on more than 8,000 patients from 21 centers. Patients' encounter-based data, including demographics, comorbidities, hospitalizations, falls, medications, treatments, and outcomes, are collected annually on brief data collection forms. The registry database includes a diverse population of PD patients, and analyses have confirmed variation in practice patterns across centers. The registry data have yielded important findings, including enhanced understanding of factors and predictors of patients' quality of life and caregiver burden. Additional cross-sectional and longitudinal analyses are planned using physician care and patient outcome data to describe practice patterns across the registry, identify and improve understanding of best practices, and support the development of guidelines.

Many neurologists were initially doubtful about the value of a registry in this disease area. For the most part, their past experience was with mortality-based registries based around interventions or fatal illnesses; these failed to model a disease with complex, heterogeneous symptomology, where the pathology could not be directly measured. Increasingly providers have recognized the value of the statistical power and nuanced insight that can be leveraged in this large and detailed registry of expert care.

Key Point

Registry-based quality improvement programs can be useful in many clinical settings, from in-hospital care (e.g., heart failure) to chronic progressive diseases (e.g., PD). The design of the registry and the quality improvement initiative must reflect the nature of the disease and the state of existing evidence. For chronic, progressive diseases, registries can be useful tools for identifying, developing, and disseminating guidelines for best practices to improve quality of care.

For More Information

- Margolius A, Cubillos F, He Y, et al. Predictors of clinically meaningful change in PDQ-39 in Parkinson's disease. Parkinsonism & related disorders. 2018;56:93-7. PMID: 30056039 DOI: 10.1016/j.parkreldis.2018.06.034.
- Rafferty MR, Schmidt PN, Luo ST, et al. Regular Exercise, Quality of Life, and Mobility in Parkinson's Disease: A Longitudinal Analysis of National Parkinson Foundation Quality Improvement Initiative Data. J Parkinsons Dis. 2017;7(1):193–202. PMID: 27858719. DOI: 10.3233/JPD-160912.
- Hassan A, Wu SS, Schmidt P, et al. The Profile of Long-term Parkinson's Disease Survivors with 20 Years of Disease Duration and Beyond. Journal of Parkinson's disease. 2015;5(2):313-9. PMID: 25720446. DOI: 10.3233/JPD-140515.
- Parashos SA, Wielinski CL, Giladi N, et al. Falls in Parkinson disease: analysis of a large cross-sectional cohort. Journal of Parkinson's disease. 2013;3(4):515-22. PMID: 24113557 DOI: 10.3233/JPD-130249.

Chapter 4. Selecting and Defining Outcome Measures for Registries

1. Introduction

As discussed in Chapter 3, the outcomes captured in a registry-based study should be selected primarily based on the research questions of interest, with consideration given to the feasibility of capturing the desired outcomes within the study scope and budget. It is also important to consider the perspectives of multiple stakeholders when determining which outcomes are most relevant.

The selection and definition of patient outcomes of interest is a critical step in designing a patient registry. The outcomes of interest, together with the exposures(s) of interest, drive many of the decisions regarding the study duration, the necessary data elements, and the source(s) of the data. For example, in determining the study duration and frequency of followup, registry developers should consider when the outcomes of interest may be observed (e.g., three months after treatment, one year after treatment). When selecting data elements, it is important to define the critical data elements to capture the outcomes of interest, along with any information that is necessary for risk adjustment. In evaluating potential data sources, registry developers should consider whether the outcomes of interest are available in the data source and the reliability of such data. Decisions about data management, such as the need for adjudication or validation of outcomes, are also informed by the specific outcomes of interest.

This chapter describes a framework that can be used to guide the selection and definition of outcome measures for use within patient registries. Types of outcome measures and considerations in defining outcome measures are discussed, as well as the rationale for standardization of outcome measures and resources for finding standardized outcome measures. Considerations related to study design, data collection and management, and analysis are addressed in Chapters 3, 11, and 13, respectively.

2. Outcome Measures Framework

2.1 Development of the Outcome Measures Framework

Over the past eight years, the Agency for Healthcare Research and Quality (AHRQ) has supported a series of projects to understand how registries select and define outcome measures and to develop tools to support harmonization of outcome measures. This work launched in 2011 with a series of stakeholder meetings designed to gather information on how outcome measures were collected in existing patient registries and how stakeholders would like to see information on outcome measures presented. In parallel, background research was conducted to identify existing models or systems designed to categorize and/or present information on data elements, outcome measures, or quality measures. Based on the background research and stakeholder feedback, the initial Outcome Measures Framework (OMF) was created in early 2012 and revised following a series of web-based meetings and document review cycles with stakeholders. The OMF was finalized in December 2012.[1]

The second phase of the OMF project began in 2013 with a systematic literature review of systems used to standardize language and definitions for outcome measures and other data elements, including systems for registries, clinical trials, electronic health records (EHRs), and quality reporting systems. The literature review identified 61 publications on three major topics: harmonizing data elements, key components of outcome measures, and governance plans for existing models. Many of the publications described efforts to harmonize data elements or create core sets of outcome measures; these efforts were identified as useful models for developing standardized outcome measures through a consensus-driven process. At the time this review was completed (2014), no existing efforts with the same or substantially similar goals as the OMF project were identified.[2]

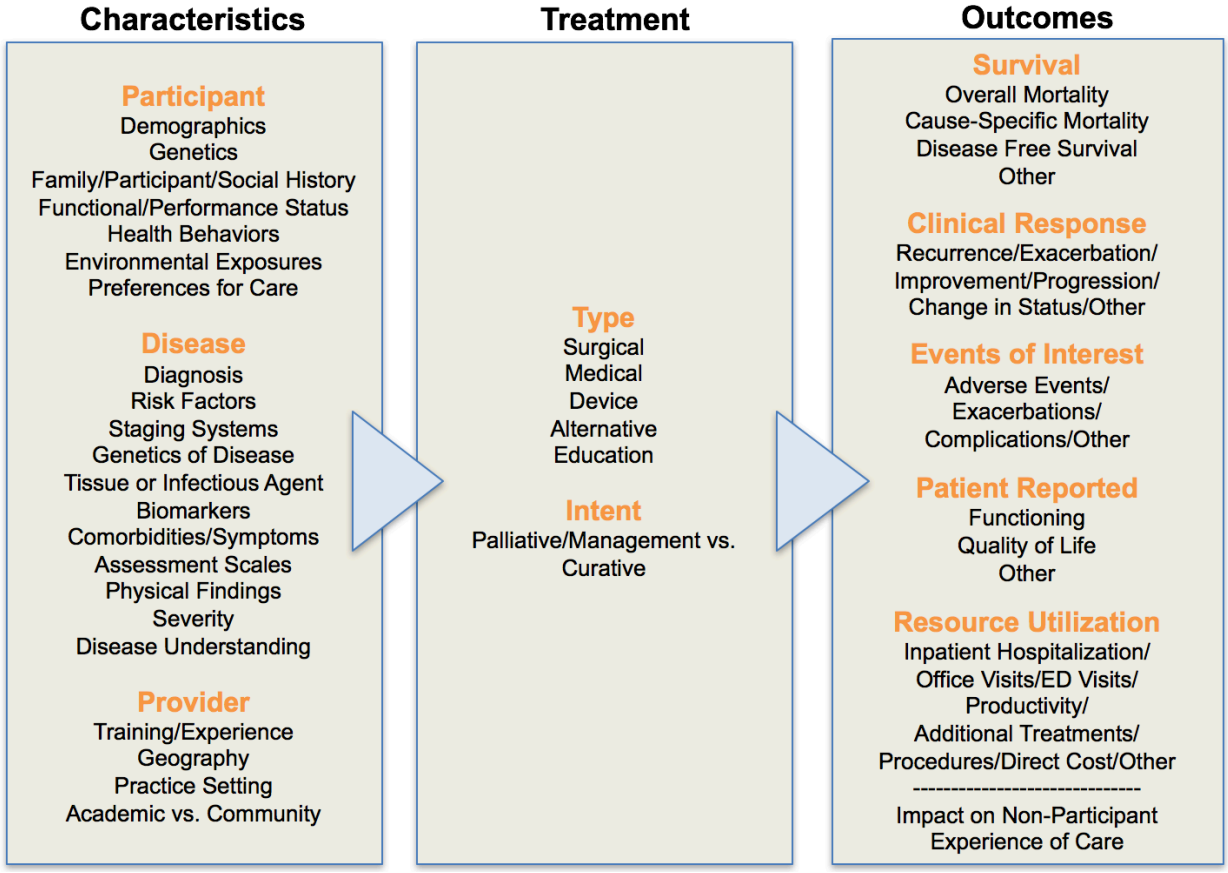
In 2015, a qualitative analysis was conducted to test the robustness of the OMF and identify any areas for improvement. Outcome measures from four diverse condition areas – depression, asthma, rheumatoid arthritis, and cardiac surgery – were abstracted from patient registries listed on ClinicalTrials.gov in June 2015 and mapped to the OMF. The condition areas were selected to represent different types of conditions, treatment options, providers, care settings, and patient populations. Two of the condition areas (rheumatoid arthritis and cardiac surgery) were selected for further analysis, and additional outcome measures were abstracted from patient registry-run websites and the published literature and mapped to the OMF. Across the four condition areas, 416 outcome measures were identified and reviewed. Most measures mapped directly to the OMF; analysis of the measures that did not map directly to the OMF resulted in minor modifications to the framework. The analysis demonstrated the robustness of the OMF for classifying a diverse group of outcome measures and highlighted its potential for supporting the development of standardized outcome measures in a range of condition areas.[3]

Throughout each phase of the development of the OMF, stakeholder feedback has been actively sought and incorporated into the framework. Over 400 stakeholders representing registry stewards, healthcare provider organizations, professional societies, academia, research and consulting organizations, government agencies, patient/consumer organizations, journal editors, payers, and pharmaceutical and medical device companies have participated in the various meetings and review activities.

2.2 Structure of the Outcome Measures Framework

The OMF (Figure 4-1) is a hierarchy with three levels: domains, subcategories of data elements, and data elements. The domains – characteristics, treatments, and outcomes – represent the process by which characteristics of the participant, disease, and provider influence treatment, and by which characteristics and treatment together influence outcomes. The process may be iterative, in that outcomes of one treatment may determine additional courses of treatment. At the second level, subcategories of data elements are presented to help guide the definition of an outcome measure. For example, information on the intent of a treatment (palliative vs. curative vs. management) is important when determining the appropriate outcomes to measure. Lastly, at the third level are the categories of data elements that would be used to define an outcome measure, such as those that capture the patient demographics and diagnosis. These categories are intentionally broad so that the framework can be used across condition areas; not all categories will be relevant in a specific condition area.

Figure 4-1. Outcome Measures Framework*



*Modified from Gliklich RE, Leavy MB, Karl J, et al. J Comp Eff Res. 2014;3(5):473-80.

In the Outcomes domain, outcome measures are grouped into five main categories: survival, clinical response or status, events of interest, patient-reported, and resource utilization. These categories represent both final outcomes, such as mortality, as well as intermediate outcomes, such as clinical response. While final outcomes may be most important in some condition areas, inclusion of intermediate outcomes such as clinical response makes the framework applicable to chronic conditions such as asthma or diabetes, where tracking patient-reported outcomes and disease progression over time is critical. It is also important to note that outcome measures may fit in more than one category. As an example, patient-reported outcomes may be used to assess clinical response (or status) for some conditions (e.g., depression).

Finally, two categories—Experience of Care and Impact on Non-Participant—are included below the Outcomes domains section. These measures fall outside of the structure of the OMF, in that they do not reflect an outcome of treatment for an individual patient; however, these are important concepts to capture in some condition areas. For example, a registry may wish to capture a birth outcome for a woman receiving treatment during pregnancy. Registries also may wish to understand patients' experiences of care, particularly as they relate to specific issues encountered during treatment, such as care coordination and provider communication in oncology. These categories are discussed in more detail in the "Types of Outcome Measures" section below.

The Characteristics domain describes attributes of the patient, disease, and provider that may be important for risk adjustment. The framework provides examples that can be modified for specific clinical areas. For example, in asthma, key characteristics to collect include the patient's age, race, ethnicity, age of onset of symptoms, history of near fatal asthma exacerbation, comorbidities, and type of provider.

The framework is a common model intended to be applied to specific conditions in potentially differing ways. For that reason, recommendations for measurement frequency are not specified in the model, but should be specified when applying the OMF to specific condition areas. Different timeframes and measurement frequencies may be appropriate depending on the condition area and outcome measure of interest. Further, some decisions regarding frequency of measurement are made by registries with a goal to minimize administrative and respondent burden. As these data elements are incorporated into interoperable health IT systems, those limitations may become fewer, allowing for new time points for some measures to be added (e.g., longer followup). Chapters 2 and 3 discuss considerations related to determining the duration of observation and the frequency of followup.

3. Types of Outcome Measures

As shown in the OMF, outcome measures can be grouped into five major categories that are relevant across a broad range of condition areas.

3.1 Survival Measures

Survival measures are important endpoints for many registries. Some survival measures, such as all-cause mortality, can be defined and captured consistently across many types of registries. All-cause mortality is broadly relevant for most condition areas and is useful for registry operations (e.g., determining that a patient has died instead of classifying the patient as lost to followup). Cause-specific mortality can be more challenging to capture because of the difficulty of ascertaining cause of death in a consistent and accurate fashion. For example, in lung cancer, pneumonia may be the immediate cause of death, while lung cancer is the underlying cause of death. Other causes of death, such as suicide, may be underreported. Because of these issues, registries interested in cause-specific mortality may consider capturing all-cause mortality as well. In registries that focus on a specific procedure or treatment, treatment-related mortality may be of interest. For example, the definition of procedure-related deaths following catheter ablation is "all-cause mortality within 30 days of the procedure or during the index procedure hospitalization (if the postoperative length of stay is > than 30 days). Procedure-related deaths include those related to a complication of the procedure or treatment for a complication of the procedure."[4]

In some condition areas, such as oncology, survival measures include the concepts of progression-free survival and disease-free survival. In the context of cancer drugs and biologics, the U.S. Food and Drug Administration (FDA) defines progression-free survival (PFS) as the time from randomization until objective tumor progression or death. Disease-free survival (DFS) is defined as the time from randomization until recurrence of tumor or death from any cause.[5] Overall survival is a critical outcome in oncology research, but it presents challenges in some

contexts, such as when the natural course of the disease is lengthy or when a new treatment results in only incremental improvements in survival. Other survival measures, such as DFS and PFS, can be important endpoints in these circumstances. These endpoints are also useful in studies that examine multiple rounds of treatment (e.g., first line, second line, etc.), as each treatment can be examined individually. However, unlike overall survival, PFS and DFS are not precisely measured, can be subject to assessment bias (e.g., measurement of tumor size), and may be defined differently in different studies.[6] The FDA has developed guidance on the use of PFS and DFS in the context of oncology clinical trials; much of this information is relevant for registry developers as well.[5]

3.2 Clinical Response or Status

Clinical Response measures capture the clinician's assessment of whether the patient is responding to treatment – meaning improving, worsening, or remaining stable – or, for patients not receiving treatment, whether the patient's clinical status is changing. These measures can be challenging to capture for several reasons. First, for many condition areas, a uniform approach to assessing clinical response has not been clearly articulated by the providers who treat those conditions. Moreover, clinicians freely admit that it can be difficult to date the onset and resolution of exacerbations of chronic diseases. It should also be noted that in some condition areas, different outcomes may be used depending on the intent of treatment. For example, in atrial fibrillation, recurrence of atrial fibrillation is an important outcome for patients undergoing ablation procedures, but this outcome is not relevant for patients receiving anticoagulation therapy. In some cases, clinical response is best measured using patient-reported outcomes (e.g., improvement or worsening in pain, asthma control). In general, clinical response measures should be valid and reproducible across different care settings and different providers and should be relevant to patients and providers.

3.3 Events of Interest

Events of interest typically include complications, adverse events related to treatment, or events associated with disease progression. For example, stroke is an important event for studies of atrial fibrillation, while exacerbation is an important event in studies of asthma. Clear, unambiguous definitions are critical for capturing events of interest consistently across sites.

3.4 Patient-Reported Outcomes

Patient-reported outcomes (PROs) reflect the patients' perceptions of their status and their perspective on health and disease. PROs have become an increasingly important avenue of investigation in many condition areas, and their importance is widely recognized. However, identification and selection of specific PROs for use within registries can be challenging. These challenges are discussed further in the "Selecting PROs" section below.

3.5 Resource Utilization

Resource utilization measures capture the patient's interactions with the healthcare system. In some cases, the outcomes of interest are specific events (e.g., hospitalizations), while in other

cases, the overall economic burden of the condition is important to capture (e.g., office visits, medications, hospitalizations, etc.). For some conditions, impact on work productivity and missed days of school are also outcomes of interest.

3.6 Composite Endpoints

Composite endpoints are composed of a specified set of outcomes of interest and are often used when the individual outcomes of interest are rare and/or when the outcomes are related clinically.[7] In a composite endpoint, the patient is considered to have reached the endpoint if any of the individual outcomes occurs. An example of a composite endpoint is major adverse cardiovascular or neurological events (MACNE), defined as a composite of cardiovascular death, myocardial infarction, stroke/non- central nervous system (CNS) systemic embolism, or transient ischemic attack.[8]

A related approach to tracking outcomes over time is the characterization of the patient's condition by a set of scores that leverage a range of patient data and can be assessed as repeat measures over time. Disease activity indices in rheumatoid arthritis are one example.[9,10]

4. Selecting Patient-Reported Outcomes[i]

The process of choosing which PRO measure(s) to include in a registry can be challenging, largely because the number of available measures is overwhelming. As discussed in Chapter 3, clear and careful definition of the target population, concept to be measured, and purpose of the registry is an important first step. In addition, when selecting measures, burden on the participant is a major consideration. The inclusion of multiple PROs can be tempting, but they may deter patient participation if the burden is excessive.

As a first step, researchers should search for existing PRO instruments that will assess the outcomes of interest. Traditional literature searches can yield results, but may be quite time-consuming. The Mapi Institute maintains the Patient-Reported Outcome and Quality of Life Instruments Database (https://eprovide.mapi-trust.org/about/about-proqolid), allowing users to search a large and relatively comprehensive database for PRO instruments that best address the specific needs identified. The Online Guide to Quality-of-life Assessment (http://www.olga-qol.com/) is another database of existing QOL instruments. Additionally, the U.S. National Institutes of Health PROMIS Initiative (http://www.healthmeasures.net/) is developing rigorously tested item banks across a broad range of domains and subdomains (functioning, disability, symptoms, distress, and role participation).[11] The PROMIS Initiative is also actively evaluating methods to achieve brevity in instruments through techniques such as computer adaptive testing. Importantly, these measures are publicly available.

[i] *Adapted from Abernethy A, Basch E, Kulig K, et al. Use of Patient-Reported Outcomes in Registries. In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

Item banks represent another option for developing PRO surveys. In general, item banks contain comprehensive collections of items that pertain to a particular construct (e.g., dyspnea).[12] Item banks generally rely on item response theory (IRT), in which the unit of focus is the item rather than the entire instrument. As such, instruments can be constructed using IRT that employ only those items which provide the most useful and relevant information, eliminating questions with little added value, without compromising psychometric qualities.[13] The PROMIS Initiative is an example of an item bank. A Computer Adaptive Test (CAT) is the dynamic application of an item bank using an algorithm that can narrow the number of items that need to be presented to a patient in order to arrive at a scale score. This can be a useful tool for limiting respondent burden for some PRO uses, although CAT scales require a continuous connection to the internet. An example of this are the CAT versions of the PROMIS scales.

Many properties of PRO instruments should be considered when choosing the appropriate instrument for a specific registry. These include the developmental history and conceptual framework; psychometric properties; content, construct, and criterion validity; reliability; and ability to detect change. The interpretability of the scores and the availability of alternate forms (e.g., different languages, different modalities for administration) are also important. Extensive literature exists on these topics; in particular, the COSMIN study (COnsensus-based Standards for the selection of health status Measurement INstruments) checklist is a useful tool for helping to guide the selection of a measurement instrument.[14] Registries should also consider the intended use of the data; for example, registries that are intended to inform regulatory decision making should follow the U.S. Food and Drug Administration (FDA) guidance on PROs.[15] Case Examples 8 and 9 provide an examples of the use of PRO instruments in registries.

5. Standardized Outcome Measures

5.1 Rationale for Standardization

Currently, registries, clinical trials, quality improvement initiatives, and other data collection efforts frequently measure different outcomes or use different definitions of the same outcome measure. For example, a technology assessment to determine the safety and efficacy of retinal prosthesis systems for halting disease progression in patients with retinitis pigmentosa reported 74 different outcome measures used in 11 studies.[16] Only three of the 74 outcome measures were reported by three or more studies, and only four of the outcome measures had evidence of validity and reliability. This type of variation in the selection of outcome measures is common across condition areas and has been well-documented in the literature.[17-20]

Variation in the definition of a specific outcome measure is equally problematic. Consider, for example, the definitions of bleeding that are used in cardiovascular research. A systematic review and meta-analysis published in 2014 found that 10 different definitions of major bleeding are currently used in clinical trials and patient registries for patients undergoing percutaneous coronary intervention (PCI).[21] The definitions include different clinical events (e.g., blood transfusion, hemorrhage), different laboratory parameters, and different outcomes (e.g., mortality), and the incidence of major bleeding, naturally, varies depending on the definition used by the study. In one example cited by the authors, non-coronary artery bypass graft related major bleeding occurred in 0.87% of patients according to one definition but in 3.1% of the same population according to another definition. While PCI studies are measuring the same outcome,

"major bleeding," comparison across studies is challenging because of the variations in definition. An earlier review, published in 2007, identified the same issue with bleeding definitions in PCI studies, leading the authors to conclude that "different bleeding definitions can lead to markedly different conclusions about the safety of an antithrombotic regimen."[22]

To address these issues, many consensus-based efforts with different intended uses and scopes have been launched. For example, the National Institutes of Health (NIH) has focused on harmonization of data elements by supporting multiple efforts to develop common data elements (CDEs), both for specific disease areas as well as for general use. The Office of Rare Diseases Research (ORDR), within NIH, has developed CDEs for use in any rare disease registry in conjunction with the Global Rare Diseases Patient Registry (GRDR) being developed through the ORDR.[23] NIH also has launched a repository to facilitate access to CDE resources.[24] The Pew Charitable Trusts is also working on a collaborative project with the Duke Clinical Research Institute to develop registry data standards for concepts collected frequently in registries.[25]

At the outcome measure level, some efforts have focused on standardizing the definition of a single outcome, such as myocardial infarction,[26] while others have focused on harmonizing the outcome measure concepts captured across studies in a specific disease area. OMERACT (Outcome Measures in Rheumatology), a long-standing, independent, and international initiative, is an example of the latter type of effort. Over the past 20 years, OMERACT has developed core sets of outcome measures for use in rheumatoid arthritis, osteoarthritis, psoriatic arthritis, fibromyalgia, and other rheumatic disease research through a well-documented, repeatable process that has served as a model for other efforts.[27] The International Consortium for Health Outcomes Measurement (ICHOM) also develops standard sets of outcome measures in different clinical areas, with the goal of improving healthcare quality and patient outcomes.[28] Finally, some efforts have focused on improving the methodology used to develop and report on consensus-based standards[29,30] or increasing access to standards that have already been developed.[31] A full review of existing efforts is beyond the scope of this chapter; more information can be found in a 2014 literature review on this topic published by AHRQ,[2] the COMET Initiative database,[31] and the NIH CDE Repository.[32]

The use of established standardized outcome measures or other data standards, when available, is essential so that registries can maximally contribute to evolving medical knowledge. Standard terminologies—and to a greater degree, higher level groupings into core datasets for specific conditions—not only improve efficiency in establishing registries but also promote more effective sharing, combining, or linking of datasets from different sources. Furthermore, the use of well-defined standards for data elements and data structure ensures that the meaning of information captured in different systems is the same. This is critical for "semantic" interoperability between information systems and to maximize the value of registries as tools in learning health systems and a national research infrastructure.

Yet, despite many efforts, many new registries do not use existing standardized measures or data elements. Researchers may not be aware of existing standards, may disagree with the standards or wish to measure different outcomes, or may be uncertain about the quality or value of using the existing standards.[17] A 2016 report from The Pew Charitable Trusts examined barriers to use of existing data standards in patient registries and found that registry stewards frequently have

not participated in the development of data standards, resulting in standards that may not meet the needs of registries and their stakeholders.[33]

5.2 OMF Standardized Measures

Recently, AHRQ supported an effort to develop minimum sets of harmonized outcome measures in five condition areas using the OMF as a conceptual model.[34] These minimum measure sets contain outcome measures that are feasible to capture in registries and routine clinical practice and that are important to providers, patients, payers, and other stakeholders. In addition to narrative definitions, the outcome measures were mapped to standardized terminologies to facilitate consistent collection and implementation within electronic health records and other systems.

For this project, standardized outcome measures were developed for the five condition areas using a reproducible process involving registry sponsors and other stakeholders, such as clinicians and representatives from patient advocacy organizations, payers, funding agencies, regulatory bodies, and research organizations. The five condition areas – atrial fibrillation, asthma, depression, non-small cell lung cancer, and lumbar spondylolisthesis – were selected to represent different types of conditions (chronic, acute, mental health), treatment modalities, care providers and care settings, and patient populations. Within each condition area, workgroups made up of registry sponsors and other stakeholders produced a minimum set of standardized measures that could be captured in future registries as well as in clinical practice in the condition area of interest; workgroups also identified characteristics of the patient, disease, and provider that are necessary to support appropriate risk adjustment for the measures included in the minimum set. Measure sets for atrial fibrillation[4] and asthma[35] have been published, and publications describing the other measure sets are forthcoming.

6. Conclusions

The selection of outcome measures is a critical step in designing a patient registry. When selecting and defining outcome measures, consideration should be given to the outcome's relevance to patients, providers, and other key stakeholders; whether it can be collected accurately and consistently across participating registry sites; and whether it is feasible to capture within the registry scope and budget. The OMF offers a useful model for selecting and defining outcome measures within registries. In addition, the use of standardized outcome measures is encouraged whenever feasible to facilitate consistency in data collection and comparability of results across registries and other efforts in learning health systems.

References for Chapter 4

1. Gliklich RE, Leavy MB, Karl J, et al. A framework for creating standardized outcome measures for patient registries. J Comp Eff Res. 2014;3(5):473-80. PMID: 25350799. DOI: 10.2217/cer.14.38.

2. L&M Policy Research, LLC, Quintiles Outcome. Registry of Patient Registries Outcome Measures Framework: Literature Review Findings and Implications. OMF Literature Review Report. (Prepared under Contract No. 290-2014-00004-C.) AHRQ Publication No. 16-EHC036-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2016. www.effectivehealthcare.ahrq.gov/reports/final/cfm.

3. Gliklich RE BK, Eisenberg F, Hanna J, Leavy MB, Campion D, Christian JB. Registry of Patient Registries Outcome Measures Framework: Information Model Report. Methods Research Report. (Prepared by L&M Policy Research, LLC, under Contract No. 290-2014-00004-C.) AHRQ Publication No. 17(18)-EHC012-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018. www.effectivehealthcare.ahrq.gov/reports/final/cfm. DOI: https://doi.org/10.23970/AHRQROPRMETHODS.

4. Calkins H, Gliklich RE, Leavy MB, et al. Harmonized outcome measures for use in atrial fibrillation patient registries and clinical practice: Endorsed by the Heart Rhythm Society Board of Trustees. Heart Rhythm. 2019;16(1):e3-e16. PMID: 30449519. DOI: 10.1016/j.hrthm.2018.09.021.

5. U.S. Food and Drug Administration. Guidance for Industry. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. December 2018. https://www.fda.gov/media/71195/download. Accessed June 10, 2019.

6. Gutman SI, Piper M, Grant MD, et al. Progression-Free Survival: What Does It Mean for Psychological Well-Being or Quality of Life? Methods Research Report. (Prepared by the Blue Cross and Blue Shield Association Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-2007-10058-I.) AHRQ Publication No. 13-EHC074-EF. Rockville, MD: Agency for Healthcare Research and Quality. April 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

7. U.S. Food and Drug Administration. Multiple Endpoints in Clinical Trials Guidance for Industry. DRAFT Guidance for Industry. January 2017. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry. Accessed June 10, 2019.

8. Inohara T, Shrader P, Pieper K, et al. Association of of Atrial Fibrillation Clinical Phenotypes With Treatment Patterns and Outcomes: A Multicenter Registry Study. JAMA Cardiol. 2018;3(1):54-63. PMID: 29128866. DOI: 10.1001/jamacardio.2017.4665.

9. Salaffi F, Cimmino MA, Leardini G, et al. Disease activity assessment of rheumatoid arthritis in daily practice: validity, internal consistency, reliability and congruency of the Disease Activity Score including 28 joints (DAS28) compared with the Clinical Disease Activity Index (CDAI). Clin Exp Rheumatol. 2009;27(4):552-9. PMID: 19772784.

10. Anderson J, Caplan L, Yazdany J, et al. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. Arthritis Care Res (Hoboken). 2012;64(5):640-7. PMID: 22473918. DOI: 10.1002/acr.21649.

11. Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care. 2007;45(5 Suppl 1):S3-S11. PMID: 17443116. DOI: 10.1097/01.mlr.0000258615.42478.55.

12. Flynn KE, Dombeck CB, DeWitt EM, et al. Using item banks to construct measures of patient reported outcomes in clinical trials: investigator perceptions. Clin Trials. 2008;5(6):575-86. PMID: 19029206. DOI: 10.1177/1740774508098414.

13. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010;63(11):1179-94. PMID: 20685078. DOI: 10.1016/j.jclinepi.2010.04.011.

14. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49. PMID: 20169472. DOI: 10.1007/s11136-010-9606-8.

15. U.S. Food and Drug Administration. Guidance for Industry: Patient Reported Outcome Measures: Use in Medical Product Development and Labeling Claims. December 2009. https://www.fda.gov/media/77832/download. Accessed June 5, 2019.

16. Fontanarosa J TJ, Samson DJ, VanderBeek BL, Schoelles, K. Retinal Prostheses in the Medicare Population. AHRQ Project ID: RPST0515. Rockville, MD: Agency for Healthcare Research and Quality; 2016.

17. Tunis SR, Clarke M, Gorst SL, et al. Improving the relevance and consistency of outcomes in comparative effectiveness research. J Comp Eff Res. 2016;5(2):193-205. PMID: 26930385. DOI: 10.2217/cer-2015-0007.

18. Curtis JR, Jain A, Askling J, et al. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. Semin Arthritis Rheum. 2010;40(1):2-14 e1. PMID: 20674669. DOI: 10.1016/j.semarthrit.2010.03.003.

19. Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. BMJ. 2017;357:j2262. PMID: 28515234. DOI: 10.1136/bmj.j2262.

20. Maddox TM, Albert NM, Borden WB, et al. The Learning Healthcare System and Cardiovascular Care: A Scientific Statement From the American Heart Association. Circulation. 2017;135(14):e826-e57. PMID: 28254835. DOI: 10.1161/CIR.0000000000000480.

21. Kwok CS, Rao SV, Myint PK, et al. Major bleeding after percutaneous coronary intervention and risk of subsequent mortality: a systematic review and meta-analysis. Open Heart. 2014;1(1):e000021. PMID: 25332786. DOI: 10.1136/openhrt-2013-000021.

22. Steinhubl SR, Kastrati A, Berger PB. Variation in the definitions of bleeding in clinical trials of patients with acute coronary syndromes and undergoing percutaneous coronary interventions and its impact on the apparent safety of antithrombotic drugs. Am Heart J. 2007;154(1):3-11. PMID: 17584547. DOI: 10.1016/j.ahj.2007.04.009.

23. Rubinstein YR, McInnes P. NIH/NCATS/GRDR(R) Common Data Elements: A leading force for standardized data collection. Contemp Clin Trials. 2015;42:78-80. PMID: 25797358. DOI: 10.1016/j.cct.2015.03.003.

24. Sheehan J, Hirschfeld S, Foster E, et al. Improving the value of clinical research through the use of Common Data Elements. Clin Trials. 2016;13(6):671-6. PMID: 27311638. DOI: 10.1177/1740774516653238.

25. Registry Data Standards. Duke Clinical Research Institute. https://dcri.org/registry-data-standards/. Accessed June 10, 2019.

26. Thygesen K, Alpert JS, Jaffe AS, et al. Third universal definition of myocardial infarction. Journal of the American College of Cardiology. 2012;60(16):1581-98. PMID: 22958960. DOI: 10.1016/j.jacc.2012.08.001.

27. OMERACT. Outcome Measures in Rheumatology. https://omeract.org/. Accessed June 10, 2019.

28. International Consortium for Health Outcomes Measurement (ICHOM). https://www.ichom.org/. Accessed June 10, 2019.

29. Kirkham JJ, Davis K, Altman DG, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. PLoS Med. 2017;14(11):e1002447. PMID: 29145404. DOI: 10.1371/journal.pmed.1002447.

30. Kirkham JJ, Gorst S, Altman DG, et al. Core Outcome Set-STAndards for Reporting: The COS-STAR Statement. PLoS Med. 2016;13(10):e1002148. PMID: 27755541. DOI: 10.1371/journal.pmed.1002148.

31. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative. http://www.comet-initiative.org/. Accessed June 4, 2019.

32. NIH Common Data Elements (CDE) Repository. National Library of Medicine. National Institutes of Health. https://cde.nlm.nih.gov/. Accessed June 4, 2019.

33. Next Steps to Encourage Adoption of Data Standards for Clinical Registries. The Pew Charitable Trusts. November 2016. https://www.pewtrusts.org/en/research-and-analysis/fact-sheets/2016/11/next-steps-to-encourage-adoption-of-data-standards-for-clinical-registries. Accessed June 10, 2019.

34. Agency for Healthcare Research and Quality. Outcome Measures Framework. https://effectivehealthcare.ahrq.gov/topics/registry-of-patient-registries/outcome-measures-framework. Accessed June 10, 2019.

35. Gliklich RE, Castro M, Leavy MB, et al. Harmonized outcome measures for use in asthma patient registries and clinical practice. J Allergy Clin Immunol. 2019 Sep;144(3):671-681.e1. PMID: 30857981. DOI: 10.1016/j.jaci.2019.02.025.

Case Examples for Chapter 4

Case Example 9. Developing and validating a patient-administered questionnaire

Description	The Benign Prostatic Hypertrophy (BPH) Registry and Patient Survey was a multicenter, prospective, observational registry examining the patient management practices of primary care providers and urologists, and assessing patient outcomes, including symptom amelioration and disease progress. The registry collected patient-reported and clinician-reported data at multiple clinical visits.
Sponsor	sanofi-aventis
Year Started	2004
Year Ended	2007
No. of Sites	403
No. of Patients	6,928

Challenge

Lower urinary tract symptoms associated with benign prostatic hyperplasia (LUTS/BPH) have a strong relationship to sexual dysfunction in aging males. Sexual dysfunction includes both erectile dysfunction (ED) and ejaculatory dysfunction (EjD), and healthcare providers treating patients with symptoms of BPH should evaluate men for both types of dysfunction. Providers can use the Male Sexual Health Questionnaire (MSHQ), a validated, self-administered, sexual function scale, to assess dysfunction, but the 25-item scale can be perceived as too long. To assess EjD more efficiently, it was necessary to develop a brief, patient-administered, validated questionnaire.

Proposed Solution

The team used representative, population-based samples to develop a short-form scale for assessing EjD. The team administered the 25-item MSHQ to three populations: a sample of men from the Men's Sexual Health Population Survey, a subsample of men from the Urban Men's Health Study, and a sample of men enrolled in the observational registry.

Using the data from the sample populations, the team conducted a series of analyses to develop the scale. The team used factor analysis to help select the items from the scale that had the highest correlations with the principal factors. Using conventional validation, the team examined reliability (both internal consistency and test-retest repeatability). To assess validity, tests of repeatability and discriminant/convergent validity were used to determine that the short form successfully discriminated between men with no to mild LUTS/BPH and those with moderate to severe LUTS/BPH. Lastly, the team examined the correlation between the 7-item ejaculation domain of the 25-item MSHQ and the new short-form scale, using data from the observational registry.

Results

Based on the results of these analyses, the team selected three ejaculatory function items and one ejaculation bother item for inclusion in the new MSHQ-EjD Short Form. The new scale demonstrates a high degree of internal consistency and reliability, and it provides information to identify men with no to mild LUTS/BPH and those with moderate to severe LUTS/BPH.

Key Point

Developing new instruments for collecting patient-reported outcomes requires careful testing of the new tool in representative populations to ensure validity and reliability. Registries can provide a large sample population for validating new instruments.

For More Information

- Rosen RC, Catania JA, Althof SE, et al. Development and validation of four-item version of Male Sexual Health Questionnaire to assess ejaculatory dysfunction. Urology. 2007;69(5):805–9. PMID: 17482908 DOI: 10.1016/j.urology.2007.02.036.
- Rosen R, Altwein J, Boyle P, et al. Lower urinary tract symptoms and male sexual dysfunction: the Multinational Survey of the Aging Male. Eur Urol. 2003;44:637–49. PMID: 14644114.

Case Example 10. Using validated measures to collect patient-reported outcomes

Description	The Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD) is a household panel registry designed to assess the prevalence and incidence of diabetes mellitus and cardiovascular disease; disease burden and progression; risk predictors; and knowledge, attitudes, and behaviors regarding health in the U.S. population. The study involves three distinct phases: an initial screening survey, a baseline survey, and yearly followup surveys for 5 years.
Sponsor	AstraZeneca Pharmaceuticals LP
Year Started	2004
Year Ended	2009
No. of Sites	Not applicable
No. of Patients	More than 211,000 individuals were included in the screening survey; approximately 15,000 individuals were followed for 5 years.

Challenge

The SHIELD registry used survey methodologies to collect health information from a large sample of adults. The goal of the study was to capture participants' perspectives and views on diabetes and cardiovascular disease, risk factors for the diseases, and burden of the diseases. The study investigators, noting that treatment for diabetes and cardiovascular disease relies heavily on patient self-management, felt that it was particularly important to gather information on activities, weight control, health attitudes, quality of life, and other topics directly from the

participant, without a physician as an intermediary. The investigators also wanted to follow participants over time to better understand disease progression and changes in health behaviors or activities.

To achieve the study goals, the registry needed to collect health-related data directly from participants in such a way that the data would be reliable, valid, and comparable across participant groups and over time.

Proposed Solution

The investigators decided to use validated patient-reported outcomes measures (PROs) to collect information on health status and behaviors. The PROs allowed the data from the registry to be compared with data collected in other registries to assess the generalizability of data on the study population. In addition, the PROs already took into account issues such as recall bias and interpretability of the questions, and self-administered instruments eliminated the possibility of introducing interviewer bias.

The registry included seven PROs: (1) the 12-item Short Form Health Survey (SF-12) and European Quality of Life (EuroQoL) EQ-5D instrument, to assess health-related quality of life; (2) the Sheehan Disability Scale, to assess the level of disruption in work, social life, and family/home life; (3) the 9-item Patient Health Questionnaire, to assess depression; (4) the Work Productivity and Activity Impairment Questionnaire: General Health, to assess work productivity and absenteeism; (5) the Diet and Health Knowledge Survey; (6) the Press-Ganey Satisfaction questionnaire; and (7) the International Physical Activity Questionnaire, to assess health-related physical activity and sedentary behaviors.

The investigators considered many factors, such as length, ease of use, format, and scoring system, when selecting the PROs to include in the survey. For example, a major reason for selecting the SF-12 rather than the SF-36 as a measure of quality of life was the length of the forms (12 vs. 36 items). The survey was entirely paper-based, with participants mailing back completed forms. The validated scoring algorithms were used to account for missing or illegible values on the completed forms. All participants were able to read and write English.

Results

The registry had a generally high response rate for the surveys. The response rates were 63.7 percent for the screening survey, 71.8 percent for the baseline survey, and between 71 and 75 percent for the annual surveys. In terms of missing data, participants who returned the survey forms tended to complete all of the questions in the appropriate manner. However, the registry was missing longitudinal data from some participants. For example, a participant may have returned the completed form in 2005, failed to return the form in 2006, and returned the form again in 2007. The investigators must account for the missing 2006 values when conducting longitudinal analyses. The data from the survey were sufficient to support comparisons over time and across participant groups, leading to several publications.

Key Point

Utilization of standardized, validated instruments in a registry can offer many benefits, including enhanced scientific rigor, the ability to compare patient views over time, and the ability to compare registry data with data from other sources to assess the representativeness of the registry

population. It should be noted that significant initial planning is necessary to identify appropriate PROs, obtain the necessary permissions, and include them in a registry. Issues with missing data must be considered in the planning phases for a registry. This registry considered missing data within returned survey questionnaires. In addition, an acceptable followup rate should be stated a priori so that response rates can be better interpreted with respect to their potential for introducing bias.

For More Information

- Gavin JR III, Rodbard HW, Fox KM. Association of overweight and obesity with health status, weight management, and exercise behaviors among individuals with type 2 diabetes mellitus or with cardiometabolic risk factors. Risk Management and Healthcare Policy. 2009;2:1–7. PMID: 22312203. DOI: 10.2147/RMHP.S4562.
- Grandy S, Chapman RH, Fox KM, et al. Quality of life and depression of people living with type 2 diabetes mellitus and those at low and high risk for type 2 diabetes: findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Int J Clin Pract. 2008;62:562–8. PMID: 18266708. DOI: 10.1111/j.1742-1241.2008.01703.x
- Grandy S, Fox KM. EQ-5D visual analog scale and utility index values in individuals with diabetes and at risk for diabetes: findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). Health Qual Life Outcomes. 2008;6:18. PMID: 18304340. DOI: 10.1186/1477-7525-6-18.
- Grandy S, Fox KM, Bazata DD, et al. Association of self-reported weight change and quality of life, and exercise and weight management behaviors among adults with type 2 diabetes mellitus. Cardiol Res Pract. 2012;2012:892564. PMID: 22645696. DOI: 10.1155/2012/892564.
- Rodbard HW, Bays HE, Gavin JR III, et al. Rate and risk predictors for development of self-reported type 2 diabetes mellitus over a 5-year period: the SHIELD study. Int J Clin Pract. 2012;66:684–691. PMID: 22698420. DOI: 10.1111/j.1742-1241.2012.02952.x.

Chapter 5. Data Elements for Registries

1. Introduction

Selection of data elements for a registry requires a balancing of potentially competing considerations. These considerations include the importance of the data elements to the integrity of the registry, their reliability, their necessity for the analysis of the primary outcomes, their contribution to the overall response burden, and the incremental costs associated with their collection. Registries are generally designed for a specific purpose, and data elements not critical to the successful execution of the registry or to the core planned analyses should not be collected unless there are explicit plans for their analysis.

The selection of data elements for a registry begins with the identification of the domains that must be quantified to accomplish the registry purpose. The specific data elements can then be selected, with consideration given to standardized outcome measures, other data standards, common data definitions, and the use of patient identifiers. Next, the data element list can be refined to include only those elements that are necessary for the registry purpose. Once the selected elements have been incorporated into a data collection tool, the tool can be pilot tested to identify potential issues, such as the time required to complete the form, data that may be more difficult to access than realized during the design phase, and practical issues in data quality (such as appropriate range checks). This information can then be used to modify the data elements and reach a final set of elements.

2. Identifying Domains

Registry design requires explicit articulation of the goals of the registry and close collaboration among disciplines, such as epidemiology, health outcomes, statistics, informatics, and clinical specialties. Once the goals of the study are determined, the domains most likely to influence the desired outcomes must be identified and defined. Registries generally capture data on the characteristics of the patient, the disease or condition of interest; exposure(s), including treatments, and outcomes. The *characteristics* domain consists of data that describe the patient, such as information on patient demographics, medical history, health status, and any necessary patient identifiers. The *exposure* domain describes the patient's experience with the product, disease, device, procedure, or service of interest to the registry. Exposure can also include other treatments that are known to influence outcome but are not necessarily the focus of the study, so that their confounding influence can be adjusted for in the planned analyses. The *outcomes* domain consists of information on the patient outcomes that are of interest to the registry; this domain should include both the primary endpoints and any secondary endpoints that are part of the overall registry goals. These domains are illustrated in the Outcome Measures Framework[1] (see Figure 4-1).

In addition to the goals and desired outcomes, it is necessary to consider any important subsets when defining the domains. Measuring potential confounding factors (variables that are linked with both the exposure and outcome) should be taken into account in this stage of registry development. Collecting data on potential confounders will allow for analytic or design control. (See Chapters 3 and 13.)Variables that can change over time must include a time reference in

order to distinguish cause-and-effect relationships. For example, a drug taken after an outcome is observed cannot possibly have contributed to the development of that outcome. Time reference periods can be addressed by including start and stop dates for variables that can change; they can also be addressed categorically, as is done in some quality improvement registries. For example, the Paul Coverdell National Acute Stroke Registry organized its patient-level information into categories to reflect the timeframe of the stroke event from onset through treatment to followup. In this case, the domains were categorized as prehospital, emergency evaluation and treatment, in-hospital evaluation and treatment, discharge information, and post-discharge followup.[2]

3. Selecting Data Elements

The process of selecting data elements begins with identification of the data elements that best quantify each domain and the source(s) from which those data elements can be collected. When selecting data elements, gaining consensus among the registry stakeholders is important, but this must be achieved without undermining the purpose of the registry by including elements solely to please a stakeholder. Each data element should support the purpose of the registry and answer an explicit scientific question or address a specific issue or need. The most effective way to select data elements is to start with the study purpose and objective, and then decide what types of groupings, measurements, or calculations will be needed to analyze that objective. Data elements may also be selected based on performance or quality measures in a clinical area; this is a particularly relevant approach for registries that focus on quality improvement. (See Case Example 11.)

Once the plan of analysis is clear, it is possible to work backward to define the data elements necessary to implement that analysis plan. This process keeps the group focused on the registry purpose and limits the number of extraneous ("nice to know") data elements that may be included.[3] When selecting data elements, it is often helpful to gather input from statisticians, epidemiologists, psychometricians, and experts in health outcomes assessment who will be analyzing the data, as they may notice potential analysis issues that need to be considered at the time of data element selection

3.1 Data Standards

The data element selection process can be simplified if standardized outcome measures or clinical data standards for a disease area exist, as discussed in Chapter 4. In cases where clinical data standards for the disease area do not exist, established datasets or common data elements may be widely used in the field. (See Case Example 12.) For example, United Network of Organ Sharing (UNOS) collects a large amount of data on organ transplant patients. Creators of a registry in the transplant field should consider aligning their data definitions and data element formats with those of UNOS to simplify the training and data abstraction process for sites. The National Institutes of Health (NIH) maintains a repository of common data elements (CDEs) that can be used to find CDEs relevant for use in a wide range of condition areas. Another example of an established dataset is the U.S. Core Data for Interoperability (USCDI). The USCDI, which was developed by the Office of the National Coordinator for Health Information Technology (ONC), is a standardized set of health data classes and constituent data elements that are intended to support national, interoperable health information exchange.[4]

If clinical data standards for the disease area and established datasets do not exist, it is still possible to incorporate standard terminology into a registry. This will make it easier to compare the registry data with the data of other registries and reduce the training needs and data abstraction burden on sites. Examples of several standard terminologies used to classify important data elements are listed in Table 5-1.

Table 5-1. Standard terminologies

Category	Standard	Acronym	Description and Website	Developer
Billing Related	Current Procedural Terminology	CPT®	Medical service and procedure codes commonly used in public and private health insurance plans and claims processing. website: https://www.ama-assn.org/practice-management/cpt-current-procedural-terminology	American Medical Association
	International Classification of Diseases	ICD, ICD-O, ICECI, ICF, ICPC	International standard for classifying diseases and other health problems recorded on health and vital records. ICD-10-CM is used for billing and claims data in the United States. The ICD is also used to code and classify mortality data from death certificates in the United States. ICD adaptations include ICD-O (oncology), ICECI (External Causes of Injury), ICF (Functioning, Disability and Health), and ICPC-2 (Primary Care, Second Edition). website: http://www.who.int/classifications/icd/en	World Health Organization
Clinical	Systemized Nomenclature of Medicine	SNOMED CT	Clinical healthcare terminology that maps clinical concepts with standard descriptive terms. website: http://www.ihtsdo.org/snomed-ct	International Health Terminology Standards Development Organization
	Unified Medical Language System	UMLS	Database of 100 medical terminologies with concept mapping tools. website: http://www.nlm.nih.gov/research/umls/	National Library of Medicine
	Classification of Interventions and Procedures	OPCS-4	Code for operations, surgical procedures, and interventions. Mandatory for use in National Health Service (England). website: http://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp	Office of Population, Censuses, and Surveys

Category	Standard	Acronym	Description and Website	Developer
	Diagnostic and Statistical Manual	DSM	The standard classification of mental disorders used in the United States by a wide range of health and mental health professionals. The version currently in use is the DSM-V. website: https://www.psychiatry.org/psychiatrists/practice/dsm	American Psychiatric Association
Drugs	Medical Dictionary for Regulatory Activities	MedDRA	Terminology covering all phases of drug development, excluding animal toxicology. Also covers health effects and malfunctions of devices. Replaced COSTART (Coding Symbols for a Thesaurus of Adverse Reaction Terms). website: https://www.meddra.org/	International Conference on Harmonisation (ICH)
	VA National Drug File Reference Terminology	NDF-RT	Extension of the VA National Drug File; used for modeling drug characteristics, including ingredients, chemical structure, dose form, physiologic effect, mechanism of action, pharmacokinetics, and related diseases. website: https://healthdata.gov/dataset/national-drug-file-reference-terminology-api.	U.S. Department of Veterans Affairs
	National Drug Code	NDC	Unique 3-segment number used as the universal identifier for human drugs. website: http://www.fda.gov/cder/ndc/	U.S. Food and Drug Administration
	RxNorm	RxNorm	Standardized nomenclature for clinical drugs. The name of a drug combines its ingredients, strengths, and/or form. Links to many of the drug vocabularies commonly used in pharmacy management and drug interaction software. website: http://www.nlm.nih.gov/research/umls/rxnorm/	National Library of Medicine
	World Health Organization Drug Dictionary	WHODRUG	International drug dictionary. website: https://www.who-umc.org/whodrug/whodrug-portfolio/	World Health Organization
Lab Specific	Logical Observation Identifiers Names and Codes	LOINC®	Concept-based terminology for lab orders and results. website: http://www.regenstrief.org/loinc/	Regenstrief Institute for Health Care
Other	HUGO Gene Nomenclature Committee	HGNC	Recognized standard for human gene nomenclature. website: http://www.genenames.org/	Human Genome Organization

Category	Standard	Acronym	Description and Website	Developer
	Dietary Reference Intakes	DRIs	Nutrient reference values developed by the National Academy of Medicine (formerly the Institute of Medicine) to provide the scientific basis for the development of food guidelines in Canada and the United States. website: https://ods.od.nih.gov/Health_Information/Dietary_Reference_Intakes.aspx	National Academy of Medicine
	Substance Registry Services	SRS	The central system for standards identification of, and information about, all substances tracked or regulated by the Environmental Protection Agency. website: https://ofmpub.epa.gov/sor_internet/registry/substreg/LandingPage.do	Environmental Protection Agency

In addition to these standard terminologies, numerous useful commercial code listings target specific needs, such as proficiency in checking for drug interactions or compatibility with widely used electronic medical record systems. Mappings between many of these element lists are also increasingly available. For example, SNOMED CT® (Systemized Nomenclature of Medicine Clinical Terminology) can currently be mapped to ICD-10-CM (International Classification of Diseases, 10th Revision, Clinical Modification).[5]

Despite progress in the use of vocabulary and terminology standards, challenges still exist. Multiple standards are still used for some areas (e.g., medications), and some systems that capture electronic health data use local terminologies instead of existing standards. In addition, some types of electronic health data, such as radiographic images, pathology slides, and clinical notes, may not be recorded using vocabulary and terminology standards. After investigating data standards, registry planners may find that there are no useful standards or established datasets for the registry, or that these standards comprise only a small portion of the dataset. In these cases, the registry will need to select and define data elements with the guidance of its project team, which may include an advisory board.

3.2 Enrollment and Followup Data Elements

When beginning the process of selecting and defining data elements, it can be useful to start by considering the registry design. Since many registries are longitudinal, sites often collect data at multiple visits. In these cases, it is necessary to determine which data elements can be collected once and which data elements should be collected at every visit. Data elements that can be collected once are often collected at the enrollment or baseline visit. Other data elements may be collected at every followup visit or on a specified schedule (e.g., once per year) that reflects routine care. In other cases, the registry may collect data at an event level, meaning all data elements will be collected during the course of the event rather than in separate visits. In considering when to collect a data element, it is also important to determine the most appropriate order of data collection. Data elements that are related to each other temporally (e.g., dietary

information and a fasting blood sample for glucose or lipids) should be collected in the same visit rather than in different visits.

Table 5-2 provides examples of possible data elements to be collected at registry enrollment and followup visits, organized into the *characteristics*, *exposures*, and *outcomes* domains described above. The actual data elements selected for a specific registry will vary depending on the design, nature, and goals of the registry.

Table 5-2. Examples of potential registry data elements

Domain	Category	Example Data Element	Enrollment*	Followup
Participant Characteristics	Contact information	• Patient contact information (for registries with direct-to-enrollee contact) • Another individual who can be reached for followup (address, telephone, email)	X	
	Patient identifiers	• Name (last, first, middle initial) • Date of birth • Place of birth • Social Security number	X	
	Enrollment criteria	• Permission/consent to participate • Source of enrollment (e.g., provider, institution, phone number, address, contact information) • Enrollment criteria • Enrollment in clinical trials (if patients enrolled in clinical trials are eligible for the registry)	X	
	Demographics	• Race/ethnicity • Gender • Age • Preferred language • Country, State, city, county, ZIP Code of residence	X	

Domain	Category	Example Data Element	Enrollment*	Followup
	Social determinants of health	• Education • Economic status, income, living situation • Employment, industry, job category • Health insurance status/access to care barriers • Social history • Marital status • Family history • Social support networks • Sexual history • Foreign travel, citizenship • Legal characteristics (e.g., incarceration, legal status) • Reproductive history • Health literacy • Individual understanding of medical conditions and the risks and benefits of interventions • Social environment (e.g., community services)	X	
	Genetics	• Results of genetic tests (e.g., BRCA1 or BRCA2)	X	
	Functional/ performance status	• Ability to perform tasks related to daily living • Quality of life • Symptoms	X	X
	Preferences for care	• Preference for medical vs. surgical management	X	
Disease Characteristics	Diagnosis	• Diagnosis • Diagnostic tests and results • Date of diagnosis/time since diagnosis	X	
	Severity/staging systems	• Risk scores (e.g., bleeding risk scores) • Disease severity classification • Cancer stage at diagnosis	X	X
	Biomarker testing results	• EGFR mutation in lung cancer	X	
	Comorbidities / medical history	• Comorbidities • Developmental history (pediatric/adolescent)	X	
	Assessment scales	• Disease activity scores	X	X
	Prior treatments	• Medical • Procedures • Alternative	X	

Domain	Category	Example Data Element	Enrollment*	Followup
	Physical findings	• Vital signs • Laboratory findings • Imaging results • Physical exam findings	X	X
Provider Characteristics	Training/ Experience	• Specialty • Years in practice • Procedural volume	X	
	Practice setting	• Geographical location • Academic vs. community • Quality improvement programs • Disease management, case management • Adherence programs • Information technology use (e.g., computerized physician order entry, e-prescribing, electronic medical records)	X	
Exposure(s)	Pharmaceutical or biologic medicine	• Medication type • Route of administration • Dose • Duration of use • Start and stop date • Generic or branded (and which brand), if relevant • Adherence	X	X
	Device	• Device type • Unique device identifier (UDI), if available • Any device-related procedures (e.g., for implantable devices)	X	X
	Procedure	• Procedure type • Periprocedural complications • Date	X	X
	Other exposures	• Alternative treatments • Education	X	X
	Intent of treatment	• Palliative • Management • Curative	X	X
Outcomes	Survival outcomes	• Death • Date of death • Cause of death		X
	Clinical response or status outcomes	• Disease progression • Improvement/worsening in symptoms • Recurrence		X

Domain	Category	Example Data Element	Enrollment*	Followup
	Events of interest outcomes	• Adverse events • Hospitalization • Emergency room visits • Surgical complications		X
	Patient-reported outcomes	• Social functioning • Physical functioning • Pain intensity and interference • Psychological well-being • General quality of life • Disease-specific quality of life	X	X
	Resource utilization outcomes	• Disability, work attendance (days lost from work), or absenteeism/presenteeism Out-of-pocket costs • Healthcare utilization behavior, including outpatient visits, hospitalizations (and length of stay), and visits to the emergency room or urgent care • Patients' assessments of the degree to which they avoid healthcare because of its costs • Destination when discharged from a hospitalization (home, skilled nursing facility, long-term care, etc.) • Emergency room visits, hospitalizations (including length of stay), long-term care, or stays in skilled nursing facilities • Medical costs, often derived from data clinician office visits, hospitalizations (especially length of stay), and/or procedures		X

*Data elements collected at enrollment may be updated during the course of the registry, if necessary; for example, patient contact information or sociodemographic information may change.

In addition, data elements that may be needed for specific types of registries are outlined here:

- For registries examining questions of safety for drugs, vaccines, procedures, or devices, key information includes history of the exposure and data elements that will permit analysis of potential confounding factors that may affect observed outcomes, such as enrollee characteristics (e.g., comorbidities, concomitant therapies, socioeconomic status, ethnicity, environmental and social factors) and provider characteristics. For drug exposures, data on duration of use (start and stop dates), as well as data providing continuing evidence that the drug was actually used (data on medication persistence and/or adherence), may be important. The registry should also record specific information about the products of interest, including route of administration, dose, and, ideally, information about whether a generic or branded product was used (and which brand). Studies of biologic medicines and devices benefit from including device identifiers, as well as information about production lots and batches. When a device of

interest is dependent on accompanying devices, the registry should consider capturing relevant data about the other devices in the same system (e.g., hip and knee implants). Patterns of real-world product use, such as treatment switches, drug holidays, pill splitting and medication sharing, and patient adherence, should also be considered when selecting data elements.

- For registries examining questions of effectiveness and cost-effectiveness, key information includes the history of exposure and data elements that will permit analysis of potential confounding factors that may affect observed outcomes. It may be particularly useful to collect information to assess confounding by indication, such as the reason for prescribing a medication. In addition to the data elements mentioned above for safety, data elements may include individual behaviors and provider and/or system characteristics. For assessment of cost-effectiveness, information may be recorded on the financial and economic burden of illness, such as office visits, visits to urgent care or the emergency room, and hospitalizations, including length of stay. Information on indirect or productivity costs (such as absenteeism and disability) may also be collected. For some studies, a quality-of-life instrument that can be analyzed to provide quality-adjusted life years or similar comparative data across conditions may be useful.
- For registries assessing quality of care and quality improvement, data that categorize and possibly differentiate among the services provided (e.g., equipment, training, or experience level of providers, type of healthcare system) may be sought, as well as information that identifies individual patients as potential candidates for the treatment. Data elements that are necessary for risk-adjustment should be included. In addition, patient-reported outcomes are valuable to assess the patients' perception of quality of care.
- For registries examining the natural history of a condition, the selection of data elements would be similar to those of effectiveness registries. Rare disease registries may also consider collecting biomarkers, or, in disease areas for which biomarkers have not been identified, biological specimens, physiological tests, or radiographic studies, in the hope of furthering efforts to develop and validate biomarkers.[6]

If one goal of a registry is to identify patient subsets that are at higher risk for particular outcomes, more detailed information on patient and provider characteristics should be collected, and a higher sample size also may be required. This information may be important in registries that look at the usage of a procedure or treatment. Quality improvement registries also use this information to understand how improvement differs across many types of institutions.

Lastly, it is important to plan for patients who will leave the registry. While the intention of a registry is generally for all patients to remain in the study until planned followup is completed, planning for patients to leave the study before completion of full followup may reduce analysis problems. By designing a final study visit form, registry planners can more clearly document when losses to followup occurred and possibly collect important information about why patients left the study. Not all registries will need a study discontinuation form, as some studies collect data on the patient only once and do not include followup information (e.g., in-hospital procedure registries).

3.3 Data Definitions

Documentation of explicit data definitions for each variable to be collected is essential to the process of selecting data elements. This is important to ensure internal validity of the proposed study so that all participants in data collection are acquiring the requisite information in the same reproducible way. (See Chapter 11.) This process may be simplified if standardized data elements and data definitions are used (e.g., CDEs or data standards). Use of existing, standardized definitions also improves the ability of the registry to compare and exchange data with other systems in the future. However, registries may need to develop data definitions when existing standards do not meet the needs of the registry. The data definitions should include the ranges and allowable values for each individual data element, as well as the potential interplay of different data elements. For example, logic checks may be created for data elements that should be mutually exclusive.

When deciding on data definitions, it is important to determine which data elements are required and which elements may be optional. This is particularly true in cases where the registry may collect some "nice to know" data elements. The determination will differ depending on whether the registry is using existing medical record documentation to obtain a particular data element or whether the clinician is being asked directly. For example, the New York Heart Association Functional Class for heart failure is an important staging element but is often not documented.[7] However, if clinicians are asked to provide the data point prospectively, they can readily do so. Consideration should also be given to accounting for missing or unknown data. In some cases, a data element may be unknown or not documented for a particular patient, and followup with the patient to answer the question may not be possible. Including an option on the form for "not documented" or "unknown" will allow the person completing the CRF to provide a response to each question rather than leaving it blank. Depending on the analysis plans for the registry, the distinction between undocumented data and missing data may be important.

3.4 Patient Identifiers

When selecting patient identifiers, there are a variety of options, such as the patient's name, date of birth, or some combination thereof, that are subject to legal and security considerations. More specific patient information may be needed when linkage to or integration with other data sources is planned, depending on the planned method of patient matching. In selecting patient identifiers, thought should be given primarily to patient privacy and security, as well as to the possibility that patient identifiers may change during the course of the registry. For example, patients may change their names during the course of the registry following marriage/divorce, or patients may move or change their telephone numbers. Patient identifiers can also be inaccurate because of intentional falsification by the patient (e.g., for privacy reasons in a sexually transmitted disease registry), unintentional misreporting by the patient or a parent (e.g., wrong date of birth), or typographical errors. In these cases, having more than one patient identifier for matching patient records can be invaluable. In addition, identifier needs will differ based on the registry goals. For example, a registry that tracks children will need identifiers related to the parents, and registries that are likely to include twins (e.g., immunization registries) should plan for the duplication of birth dates and other identifiers. In selecting patient identifiers for use in a

registry, registry planners will need to determine what data are necessary for their purpose and plan for potential inaccurate and changing data.

Generally, patient identifiers can simplify the process of identifying and tracking patients for followup. Patient identifiers also allow for the possibility of identifying patients who are lost to followup due to death (i.e., through the National Death Index) and linking to birth certificates for studies in children. In addition, unique patient identifiers allow for analysis to remove duplicate patients.

When considering the advantages of patient identifiers, it is important to take into account the potential challenges that collecting patient identifiers can present and the privacy and security concerns associated with the collection and use of patient identifiers. Obtaining consent for the use of patient-identifiable information can be an obstacle to enrollment, as it can lead to the refusal of patients to participate. Chapter 7 contains more information on the ethical and legal considerations of using patient identifiers.

3.5 Multinational Registries

Registries are commonly multinational, and data elements must be tailored appropriately for each country. Even when the same concepts are captured, examination and laboratory test results or units may differ among countries, making standardization of data elements necessary at the data-entry level. Data elements relating to cost-effectiveness studies may be particularly challenging, since there is substantial variation among countries in healthcare delivery systems and practice patterns, as well as in the cost of medical resources that are used as "inputs." Alternatively, if capture of internationally standardized data elements is not desirable or cannot be achieved, registry stakeholders should consider provisions to capture data elements according to local standards. Later, separate data conversions and merging outside the database for uniform reporting or comparison of data elements captured in multiple countries can be evaluated and performed as needed if the study design ensures that all data necessary for such conversions have been collected.

Multinational registries also must carefully consider translation of data elements and case report forms (CRFs) into different languages. Appropriate translation and linguistic validation of CRFs is critically important to maintain a high quality of systematic data collection in the registry and to ensure that data captured from different countries have the same definition and meaning. Linguistic validation is important even when the same language is spoken in different countries. For example, though persons in the United States and the United Kingdom (UK) both commonly speak English, content validity of the same case CRF may differ between the two nations due to different cognitive interpretations. Consider patient-reported weight; a patient in the United States would typically write the full amount in pounds, while a patient in the UK would typically write the amount in stone or pounds or possibly kilograms.

4. Registry Data Map

Once data elements have been selected, a data map should be created. The data map identifies all sources of data (Chapter 6) and explains how the sources of data will be integrated. Data maps

are useful to defend the validity and/or reliability of the data, and they are typically an integral part of the data management plan (Chapter 11). Clear operational definitions for each data element are also important to facilitate eventual interpretation of the data.

5. Pilot Testing

After the data elements have been selected and the data map created, it is important to pilot test the data collection tools to determine the time and costs of obtaining the data and the resulting clinician and patient burden. For example, through pilot testing, registry planners might determine that it is wise to collect certain data elements that are either highly burdensome or only "nice to know" in only a subset of participating sites (nested registry) that agree to the more intensive data collection, so as not to endanger participation in the registry as a whole. Pilot testing should also help to identify the rate of missing data and any validity issues with the data collection system.

The burden of data collection is a major factor determining a registry's success or failure, with major implications for the cost of participation and for the overall acceptance of the registry by healthcare personnel and patients. Moreover, knowing the anticipated time needed for patient recruitment/enrollment will allow better communication to potential sites regarding the scope and magnitude of commitment required to participate in the study. Registries that obtain information directly from patients include the additional issue of participant burden, with the potential for participant fatigue, leading to failure to answer all items in the registry. Highly burdensome questions can be collected in a prespecified subset of subjects. The purpose of these added questions should be carefully considered when determining the subset so that useful and accurate conclusions can be achieved.

Pilot testing the registry also allows the opportunity to identify issues and make refinements in the registry-specific data collection tools, including alterations in the format or order of data elements and clarification of item definitions. Piloting may also uncover problems in registry logistics, such as the ability to accurately or comprehensively identify subjects for inclusion. A fundamental aspect of pilot testing is evaluation of the accuracy and completeness of registry questions and the comprehensiveness of both instructional materials and training in addressing these potential issues. Gaps in clarity concerning questions can result in missing or misclassified data, which in turn may cause bias and result in inaccurate or misleading conclusions. For example, time points, such as time to radiologic interpretation of imaging test, may be difficult to obtain retrospectively and, if they do exist in the chart, may not be consistently documented. Without additional instruction, some hospitals may indicate the time the image was read by the radiologist and others may use the time when the interpretation was recorded in the chart. The two time points can have significant variation, depending on the documentation practices of the institution.

Pilot testing ranges in practice from ad hoc assessments of the face validity of instruments and materials in clinical sites, to trial runs of the registry in small numbers of sites, to highly structured evaluations of inter-rater agreement. The level of pilot testing is determined by multiple factors. Accuracy of data entry is a key criterion to evaluate during the pilot phase of the registry. When a "gold standard" exists, the level of agreement with a reference standard

(construct validity) may be measured.[8] Data collected by seasoned abstractors or auditors following strict operational criteria can serve as the gold standard by which to judge accuracy of abstraction for chart-based registries.[9]

In instances where no reference standard is available, reproducibility of responses to registry elements by abstractors (inter-rater reliability) or test-retest agreement of subject responses may be assessed.[10] Reliability and/or validity of a data element should be tested in the pilot phase whenever the element is collected in new populations or for new applications. Similar mechanisms to those used during the pilot phase can be used during data quality assurance (Chapter 11). A kappa statistic measure of how much the level of agreement between two or more observers exceeds the amount of agreement expected by chance alone is the most common method for measuring reliability of categorical and ordinal data. The intraclass correlation coefficient, or inter-rater reliability coefficient, provides information on the degree of agreement for continuous data. It is a proportion that ranges from zero to one. Item-specific agreement represents the highest standard for registries; it has been employed in cancer registries and to assess the quality of data in statewide stroke registries. Other methods, such as the Bland and Altman method,[10] may also be chosen, depending upon the type of data and registry purpose.

6. Summary

The selection of data elements requires balancing such factors as their importance for the integrity of the registry and for the analysis of primary outcomes, their reliability, their contribution to the overall burden for respondents, and the incremental costs associated with their collection. Data elements should be selected with consideration for established clinical data standards, common data definitions, and whether patient identifiers will be used. The role of PROs and any other information provided directly by the patient is also important to consider. Lastly, it is important to determine which elements are absolutely necessary and which are desirable but not essential. Once data elements have been selected, a data map should be created with clear operational definitions for each variable, and the data collection tools should be pilot tested. Overall, the choice of data elements should be guided by parsimony, validity, and a focus on achieving the registry's purpose.

References for Chapter 5

1. Gliklich RE, Leavy MB, Karl J, et al. A framework for creating standardized outcome measures for patient registries. J Comp Eff Res. 2014;3(5):473-80. PMID: 25350799. DOI: 10.2217/cer.14.38.

2. Wattigney WA, Croft JB, Mensah GA, et al. Establishing data elements for the Paul Coverdell National Acute Stroke Registry: Part 1: proceedings of an expert panel. Stroke. 2003;34(1):151-6. PMID: 12511767.

3. Good PI. A manager's guide to the design and conduct of clinical trials. New York: John Wiley & Sons, Inc.; 2002.

4. U.S. Core Data for Interoperability. Office of the National Coordinator for Health Information Technology. Version 1. https://www.healthit.gov/isa/us-core-data-interoperability-uscdi. Accessed June 18, 2019.

5. National Library of Medicine. SNOMED CT to ICD-10-CM Map. https://www.nlm.nih.gov/research/umls/map

ping_projects/snomedct_to_icd10cm.html. Accessed June 18, 2019.

6. Institute of Medicine. Rare Diseases and Orphan Products: Accelerating Research and Development. Washington, DC: National Academies Press; 2010.

7. Yancy CW, Fonarow GC, Albert NM, et al. Influence of patient age and sex on delivery of guideline-recommended heart failure care in the outpatient cardiology practice setting: findings from IMPROVE HF. Am Heart J. 2009;157(4):754-62 e2. PMID: 19332206. DOI: 10.1016/j.ahj.2008.12.018.

8. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev. 1980;2:210-20. PMID: 7000537. DOI: 10.1093/oxfordjournals.epirev.a036224.

9. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996;25(2):435-42. PMID: 9119571. DOI: 10.1093/ije/25.2.435.

10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10. PMID: 2868172.

Case Examples for Chapter 5

Case Example 11. Using Recognition Measures To Develop a Dataset

Description	Get With The Guidelines® is the flagship program for in-hospital quality improvement of the American Heart Association and American Stroke Association . The Get With The Guidelines—Stroke program supports point of care data collection and real-time reports aligned with the latest evidence-based guidelines. The reports include achievement, quality, reporting, and descriptive measures that allow hospitals to trend their performance related to clinical and process outcomes.
Sponsor	American Heart Association/American Stroke Association
Year Started	2003
Year Ended	Ongoing
No. of Sites	Over 2,000 hospitals have participated
No. of Patients	> 5,000,000

Challenge

The primary purpose of the Get With The Guidelines—Stroke program is to improve the quality of in-hospital care for stroke patients. The program uses the PDSA (plan, do, study, act) quality improvement cycle, in which hospitals plan quality improvement initiatives, implement them, study the results, and then make adjustments to the initiatives. To help hospitals implement this cycle, the program uses a registry to collect data on stroke patients and generate real-time reports showing compliance with a set of standardized stroke recognition and quality measures. The reports also include benchmarking capabilities, enabling hospitals to compare themselves with other hospitals at a national and regional level, as well as with similar hospitals based on size or type of institution.

In developing the registry, the team faced the challenge of creating a dataset that would be comprehensive enough to satisfy evidence-based medicine but manageable by hospitals participating in the program. The program does not provide reimbursements to hospitals entering data, so it needed to keep the dataset as small as possible while still maintaining the ability to measure quality improvement.

Proposed Solution

The team began developing the dataset by working backward from the recognition measures. Recognition measures, based on the sponsor's guidelines for stroke care, contain detailed inclusion and exclusion criteria to determine the measure population, and they group patients into denominator and numerator groups. Using these criteria, the team developed a dataset that framed the questions necessary to determine compliance with each of the guidelines. The team then added questions to gather information on the patient population characteristics. Since the inception of the program, data elements and measure reports have been added or updated to maintain alignment with the current stroke guidelines. Over time, certain measures have also been promoted to or demoted from the higher tiers of recognition measures, depending on current science and changes in quality improvement focus.

Results

By using this approach, the registry team was able to create the necessary dataset for measuring compliance with stroke guidelines. The program was launched in 2003. As of 2019, over 2,000 hospitals have participated, entering data on more than five million stroke patients. The data from the program have been used in many abstracts and have resulted in dozens of manuscripts since 2007.

Key Point

Registry teams should focus on the outcomes or endpoints of interest when selecting data elements. In cases where compliance with guidelines or quality measures is the outcome of interest, teams can work backward from the guidelines or measures to develop the minimum necessary dataset for their registry.

For More Information

- http://www.heart.org
- Ormseth CH, Sheth KN, Saver JL, et al. The American Heart Association's Get With the Guidelines (GWTG)-Stroke development and impact on stroke care. Stroke and vascular neurology. 2017;2(2):94-105. PMID: 28959497. DOI: 10.1136/svn-2017-000092.
- Schwamm L, Fonarow G, Reeves M, et al. Get With the Guidelines—Stroke is associated with sustained improvement in care for patients hospitalized with acute stroke or transient ischemic attack. Circulation. 2009;119:107–11. PMID: 19075103. DOI: 10.1161/CIRCULATIONAHA.108.783688.

Case Example 12. Patient-Powered Registries: Developing Scalable and Reusable Infrastructure To Support Harmonized Data Collection Across Rare Diseases

Description	NORD created the IAMRARE™ Registry Program to address the lack of rare disease natural history data, developing a disease-agnostic registry platform to support harmonized longitudinal data collection for all rare diseases, with the goal of informing patient decision making, standards of care, and drug development.
Sponsor	National Organization for Rare Disorders (NORD)
Year Started	2014
Year Ended	Ongoing
No. of Sites	1 Platform; >20 disease-specific registries
No. of Patients	Ongoing enrollment; > 6,000 participants

Challenge

Rare diseases pose unique research challenges, such as geographically dispersed patient populations, lack of information on the natural history of the disease, absence of standards of care or treatment guidelines, and limited numbers of clinicians with relevant expertise and experience managing patients with the condition. Longitudinal, observational data captured through patient registries can provide important information about the prevalence, characteristics, and natural history of the disease. These data may be used to supplement clinical trial data and identify meaningful endpoints during drug development, and the registries may serve as vehicles for identifying potential clinical trial participants. Some rare disease registries are developed and managed by patient organizations, however, patient organizations often lack the resources needed to develop and manage a registry, underscoring the importance of a common rare disease registry infrastructure - not only to minimize the burden of conducting longitudinal research studies, but also to facilitate cross-disease analyses and community ownership of the data.

Proposed Solution

NORD has partnered with rare disease stakeholders, including patients, caregivers, researchers, clinicians, and regulatory agencies, to develop a cloud-based registry platform and supplemental support program that facilitates longitudinal and episodic data entry by both patients and caregivers. Patient organizations can leverage the platform infrastructure and the support program to develop and manage patient registries. Common data elements (CDEs) serve as the foundation for each registry and are supplemented by validated disease-specific measures or, in cases where these do not exist, custom surveys. In addition to the platform, NORD provides support for facilitating the development of research consortia to encourage collaboration among partners working in the same disease space and organizing treatment and guideline review meetings that bring together experts across stakeholder groups to utilize registry data to inform the revision of standards of care. NORD works closely with partners to refine the study design, supporting documents, and overall study management protocols, and facilitates the registry launch process through its partnership with a centralized Institutional Review Board (IRB).

NORD also provides training and resources through educational webinars, in-person workshops, and individualized guidance. Once a registry has launched, the platform supports concurrent sub-studies that branch off from existing registries to capture specific and complementary data, thereby reducing redundant registry efforts and community fragmentation.

Results

Since 2014, the program has grown to over 35 registries representing 9000+ users who have submitted more than 70,000 surveys (statistics as of May 2019). Throughout this period, continued program development has been driven by ongoing stakeholder engagement and input, collected via targeted questionnaires and meetings, as well as through consistent, open-ended dialogue. The program's community portal and in-person leadership meetings offer forums for the registry partners to consider new concepts, share resources and lessons learned, and celebrate key milestones.

With NORD's technical and programmatic infrastructure, harmonized data are collected across registries, from basic demographics to patient-reported outcomes. Preliminary registry data have been presented at national and international conferences, submitted for peer-reviewed publication, and analyzed to inform the development of patient-focused clinical trials.

Key Point

The use of CDEs, repeatable processes, and scalable infrastructure can produce efficiencies in registry development and operations and create opportunities for cross-disease analysis.

For More Information

- IAMRARE™ Registry Program. https://rarediseases.org.

Chapter 6. Data Sources for Registries

1. Introduction

Identification and evaluation of suitable data sources should be completed within the context of the registry purpose and availability of the data of interest. A single registry may have multiple purposes and integrate data from various sources. (See Case Example 14.) While some data in a registry are collected directly for registry purposes (primary data collection), the push to use real-world data and real-world evidence in decision making has resulted in an increased focus on the incorporation of data collected primarily for other purposes. Examples include demographic information from a hospital admission, discharge, and transfer system; medication use from a pharmacy database; disease and treatment information, such as details of diagnostic and therapeutic procedures or treatment plans from ancillary clinical systems (e.g., cardiology information systems, radiology information systems, surgical systems, oncology systems, etc.); electronic health records (EHRs); and medical claims databases. In addition, observational studies can generate as many hypotheses as they test, and other sources of data can be merged with the primary data collection to allow for analyses of questions that emerge during the course of the registry.

The burden of registry participation can be significantly reduced with broader use of these sources. However, high standards for quality, including documentation of transformations and traceability of data in the registry to the source, are important considerations. This chapter will review the various sources of data, comment on their strengths and weaknesses, and provide some examples of how data collected from different sources can be integrated to help answer important questions. Information on the technical aspects of linking or integrating existing data sources into registries can be found in the supplemental eBook on Tools and Technologies for Registry Interoperability.[1]

2. Intended Uses for Data Elements

The types of data to be collected are guided by the registry purpose, design, and data collection methods. The form, organization, and timing of required data are important components in determining appropriate data sources. Data elements can be grouped into categories that identify the specific variable or construct they are intended to describe. One framework for grouping data elements into categories follows:

- *Identify patients*—Rather than incorporate all possible data of interest, many registries use patient identifiers to link data from secondary sources in order to support a specific analysis. In these registries, data elements are linked to the specific patient through a unique patient identifier or registry identification number. However, the potential for mismatch errors and duplications must be managed. The use of patient identifiers may not be possible in all registries due to the additional legal requirements that usually apply to the use and disclosure of such data. (See Chapter 7.)
- *Determine eligibility*—The eligibility criteria in a registry protocol or study plan determine the group that will be included in the registry. These criteria may be very broad or restrictive, depending on the purpose. Criteria often include demographics (e.g., target

age group), a disease diagnosis, a treatment, or diagnostic procedures or laboratory tests. Healthcare provider, healthcare facility or system, and insurance criteria may also be included in certain types of registries (e.g., following care patterns of specific conditions at large medical centers compared with small private clinics).

- *Describe treatments and tests*—Treatments and tests are necessary to describe the natural history of patients. Treatments can include pharmaceutical, biological, or device therapies, or procedures such as surgery or radiation. Evaluation of the treatment itself is often a primary focus of registries (e.g., treatment safety and effectiveness over 5 years). Results of laboratory testing or diagnostic procedures may be included as registry outcomes and may also be used in defining a diagnosis or condition of interest.
- *Understand confounders*—Confounders are elements or factors that have an independent association with the outcomes of interest. These are particularly important because patients are typically not randomized to therapies in registries. Confounders such as comorbidities (disease diagnoses and conditions) can confuse analysis results and interpretation of causality. Information on the healthcare provider, treatment facility, concomitant therapies, or insurance may also be considered. Unknown confounders, or those not recorded in the registry, pose particular challenges for the analysis of patient outcomes. If external, or linked, data sources may provide values for these confounder variables not included in the registry, they could ultimately help reduce bias in the analysis and interpretation of patient outcomes.
- *Measure outcomes*—The focus of this document is on patient outcomes. Outcomes are end results and must be defined for each condition. Outcomes may include patient-reported outcomes (PROs). In some registries, surrogate markers, such as biomarkers, or other intermediate outcomes (e.g., hemoglobin A1c levels in diabetes) that are highly reflective of the longer-term end results are used.

Within this framing, a given type of data may be present in multiple categories. Consider, for example, diagnosis. One diagnosis (e.g., diabetes) may be used to determine eligibility for enrollment into a registry, while other diagnoses (e.g., heart failure, atrial fibrillation) may be captured as potential confounding variables. While both pieces of information may be present in the same secondary source, different quality requirements (e.g., more stringent requirements to verify eligibility for enrollment) may mean that the source does not satisfy both purposes.

3. Types of Data

Before considering the potential sources for registry data, it is important to understand the types of data that may be collected in a registry. Several types of data that may be gathered from other sources in some registries are described below. A given data source may contain data from more than one of these categories.

Patient identifiers—Depending on the data sources required, some registries may use certain personal identifiers for patients in order to locate them in other databases and link the data. For example, Social Security numbers (SSNs) in combination with other personal identifiers can be used to identify individuals in the National Death Index (NDI). Combinations of variables are also used by many hashing algorithms (e.g., gender, date of birth, last 4 digits of SSN, etc.).

Patient contact information, such as address and phone numbers, may be collected to support tracking of participants over time. Information for additional contacts (e.g., family members) may be collected to support followup in cases where the patient cannot be reached. In many registries, patient informed consent and appropriate privacy authorizations are required so that personal identifiers can be used for registry purposes. In some registries, the use of personal identifiers may not be possible. Chapter 7 discusses the legal requirements for including patient identifiers. Systems and processes must be in place to manage security and confidentiality of these data. Confidentiality can be enhanced by assigning a registry-specific identifier via a crosswalk algorithm, as discussed below. Demographics, such as date of birth (to calculate age at any time point), gender, and ethnicity, are typically collected and may be used to stratify the registry population.

Disease/condition—Disease or condition data include those related to the disease or condition of focus for the registry and may incorporate comorbidities. Elements of interest related to the confirmation of a diagnosis or condition could include the date of diagnosis and the specific diagnostic results that were used to make the diagnosis, depending on the purpose of the registry. Disease or condition is often a primary eligibility or outcome variable in registries, whether the intent is to answer specified treatment questions (e.g., measure effectiveness or safety) or to describe the natural history. This information may also be collected in constructing a medical history for a patient. In addition to "yes" or "no" to indicate presence or absence of the diagnosis, it may be important to capture responses such as "missing" or "unknown."

Treatment/therapy—Treatment or therapy data include specific identifying information for the primary treatment (e.g., drug name or code, biologic, device product or component parts, or surgical intervention, such as organ transplant or coronary artery bypass graft) and may include information on concomitant treatments. Dosage (or parameters for devices), route of administration, and prescribed exposure time (such as daily or 3 times weekly for 4 weeks), should be collected. Pharmacy data may include dispensing information, such as the primary date of dispensation and subsequent refill dates. Data in device registries can include the initial date of dispensation or implantation and subsequent dates and specifics of required evaluations or modifications as well as the Unique Device Identifier (UDI). Compliance data may also be collected if pharmacy representatives or clinic personnel are engaged to conduct and report pill counts or volume measurements on refill visits, or return visits for device evaluations and modifications.

Anthropometrics/vital signs—Measurements about the registry participant, such as height, weight, body mass index (BMI), pulse, temperature, and oxygen saturation can be important data in a registry. While these data may be obtained through primary data collection, they are increasingly available through secondary sources such as the electronic health record (EHR) or from patient monitors. If incorporating via a secondary source, it is important to understand whether all measurements or a subset of measurements (median, mean, maximum, minimum) are required, as the vital signs documented within some secondary sources can be both sparse (e.g., height not recorded in an adult if they had a visit within the last two weeks) and overly abundant (e.g., oxygen saturation levels recorded every five seconds on a patient monitor).

Laboratory/procedures—Laboratory and ancillary data include a broad range of testing, such as blood, tissue, catheterization, and radiology. Specific test results, units of measure, and

laboratory reference ranges or parameters are typically collected. National efforts towards interoperability mean that laboratory data are becoming increasingly standardized, making electronic transfer more feasible. A few specialized types of laboratory testing (e.g., imaging, genomics) are called out below. Diagnostic testing or evaluation may include procedures such as psychological or behavioral assessments. Results of these procedures and clinician examinations may be difficult to obtain through data sources other than the patient medical record.

Imaging—The result of a given imaging test (e.g., echocardiogram, x-ray, CT scan) may be an important element for a registry, but a distinction should be made as to whether the actual images are needed or simply the interpretations or variables derived from the images themselves (e.g., left ventricular volume from an echocardiogram). A registry that seeks to develop more sophisticated image processing or automated interpretation algorithms may need the former, while the latter is likely sufficient for a registry that uses imaging results as part of their inclusion criteria.

Biosamples/Genomic results—The increased collection, testing, and storage of biological specimens as part of a registry (or independently as a potential secondary data source such as those described further below) provides another source of information. Biorepositories are employed to store information about the specimens themselves, while genomic repositories handle the sequencing results, which may range from a handful of genetic variants or specific genotypes all the way to entire individual genomes. Increasingly with genomic data, a distinction is made between the underlying "raw" results and those specific, limited findings that might be reported out back as part of a given genomic test. Operationally, a registry may need the results of the test itself, for instance, to stratify patients based on a specific genotype, but a given research analysis may need access to all of the unreported raw results. Due to the size and specialized nature of these data, it is often more feasible to link to these datasets as part of a specific analysis, as opposed to incorporating all of the "raw" data directly into the registry. See Chapters 7 and 8 for more information on the regulatory and ethical issues related to the use of biosamples and genomic results in registries.

Survey/questionnaire data—Surveys or questionnaires can be administered directly to patients, families/caregivers, or providers to collect data for the registry. In some cases, surveys or questionnaires are used to capture patient-reported outcomes (PROs). Surveys or questionnaires can be administered in many ways (e.g., paper forms, online surveys, mobile applications). Registries often use standardized, validated instruments that may also include computer-adaptive testing (CAT) to minimize the number of questions that are presented to the respondent.

Patient-generated data—Data from devices recording a range of data from patients can be incorporated into the registry. This can range from medical devices such as internet-connected scales or continuous glucose monitors, to consumer devices like activity trackers or readings from sensors contained within smartphones or smartwatches. As with patient monitors used in the medical setting, it is important to decide whether to include the raw values (which may be numerous) or some type of summary or derivation.

Healthcare provider characteristics—Information on the healthcare provider (e.g., physician, nurse, or pharmacist) may be collected, depending on the purpose of the registry. Training,

education, or specialization may account for differences in care patterns. Geographic location has also been used as an indicator of differences in care or medical practice.

Hospital/clinic/health plan encounter details—System interactions include office visits, outpatient clinic visits, emergency room visits, inpatient hospitalizations, procedures, and pharmacy visits, as well as associated dates. The events that occurred within a given encounter are covered in the sections above (e.g., therapy/treatment, laboratory/procedure, disease/condition, etc.), but descriptive information related to the encounter itself may be useful in capturing differences in care patterns and can also be used to track patterns of referral (e.g., outpatient clinic, inpatient hospital, academic center, emergency room, pharmacy).

Cost/resource utilization—Cost and/or resource utilization data may be necessary to examine the cost-effectiveness of a treatment. Resource utilization data reflect the resources consumed (both services and products), while cost data reflect a monetary value assigned to those resources. Examples include the actual cost of the treatment (e.g., medication, screening, procedure) and the associated costs of the intervention (e.g., treatment of side effects, expenses incurred traveling to and from clinicians' appointments). Costs that are avoided due to the treatment (e.g., the cost to treat the avoided disease) and costs related to lost workdays may also be important to collect, depending on the objectives of the study. Registries that collect cost data over long periods of time (i.e., many years) may need to adjust costs for inflation during the analysis phase of the study. The types of data elements included in this framework are further described in Chapter 5 and above with respect to their source or the utility of the data for linking to other sources. Many of these may be available through data sources outside of the registry system.

Insurance—The insurance system or payer claims data can provide useful information on interactions with the healthcare systems, including visits, procedures, inpatient stays, and costs associated with these events. When using these data, it is important to understand what services were covered under the various insurance plans at the time the data were collected, as this may affect utilization patterns, but it can be reasonable to assume that these data may represent the complete capture for all reimbursed health outcomes or exposures of interest.

Environmental factors/social determinants of health—The social or environmental factors related to a patient's community are increasingly being recognized as important drivers of health disparities and a cause of variations in patient outcomes. Social determinants of health can be collected directly from patients but may also be available through secondary sources via proxy measures such as socioeconomic status, pollution levels, or community characteristics. These measures are typically assigned to the patient by geocoding their address information and linking at the appropriate geo-spatial level (neighborhood, census tract, zip code, etc.). Since address information is usually considered a patient identifier, additional regulatory approvals may be required to obtain these data.

Social media–An emerging type of data is information related to a patient's social media activity. This could include the content of the posts themselves, or simply metadata about the time and date of the posting, who viewed or commented on it, etc. These data can be synthesized into measures of community engagement. Some social media companies have restrictions on how a member's data may be obtained, so it is important to understand the potential terms of use.

4. Data Sources

Data sources are classified as primary or secondary based on the relationship of the data to the registry purpose. Primary data sources incorporate data collected for direct purposes of the registry (i.e., primarily for the registry). Primary data sources are typically used when the data of interest are not available elsewhere or, if available, are unlikely to be of sufficient accuracy and reliability for the planned analyses and uses. Primary data collection increases the probability of completeness, validity, and reliability because the registry drives the methods of measurement and data collection (See Chapter 5). Primary data collection can occur via patients/caregivers or clinicians. These data are prospectively planned and collected under the direction of a protocol or study plan, using common procedures and the same format across all registry sites and patients. The data are readily integrated for tracking and analyses. Since the data entered can be traced to the individual who collected them, primary data sources are more readily reviewed through automated checks or followup queries from a data manager than is possible with many secondary data sources (See Chapter 11).

Secondary data sources are comprised of data collected for purposes other than or in addition to the registry under consideration (e.g., routine medical care, insurance claims processing). Data that are collected as primary data for one registry are considered secondary data from the perspective of a second registry if linking was done. These data are often stored in electronic format and may be available for use with appropriate permissions. Data from secondary sources may be used in two ways: (1) the data may be transferred and imported into the registry, becoming part of the registry database, or (2) the secondary data and the registry data may be linked to create a new, larger dataset for analysis. (See Case Example 15.) This chapter primarily focuses on describing commonly used secondary sources. Chapter 11 discusses strategies for transferring secondary data into a registry (abstraction with double-data entry, direct import, transformation, algorithmic derivation).

When considering secondary data sources, it is important to note that health professionals are accustomed to entering data for defined purposes. Data in secondary sources are not constrained by a data collection protocol and therefore represent the diversity observed in real-world practice. Thus, there may be increased probability of errors and underreporting because of inconsistencies in measurement, reporting, and collection. Staff changes can further complicate data collection and may affect data quality. There may also be increased costs for linking the data from the secondary source to the primary source and dealing with any potential duplicate or unmatched patients.

The potential for data completeness, variation, and specificity must be evaluated in the context of the registry purpose and intended use of secondary data. It is crucial to have a solid understanding of the original purpose of the secondary data collection, including the processes for collection and submission, and any verification and validation practices. Questions to ask include: Is data collection passive or active? Are standard definitions or codes used in reporting data? Are standard measurement criteria or instruments used (e.g., diagnoses, symptoms, quality of life)? The existence and completeness of claims data, for example, will depend on insurance company coverage policies. One company may cover many preventive services, whereas another may have more restricted coverage. One company may cover a treatment without restriction,

while another may require prior authorization or require that the patient must have first failed on a previous, less expensive treatment. Also, coverage policies can change over time. These variations must be known and carefully documented to prevent misinterpretation of use rates. Additionally, secondary data may not all be collected in the format (e.g., units of measure) required for registry purposes and may require transformation for integration and analyses.

An overview of some secondary data sources that may be used for registries is given below. Table 6-1 identifies some key strengths and limitations of these data sources.

Table 6-1. Key data sources—strengths and limitations

Data Source Category	Data Source	Strength and Uses	Limitations
Primary Sources	**Patient-reported data/patient-generated data**	• Patient and/or caregiver outcomes. • Unique perspective. • Obtaining information on treatments not necessarily prescribed by clinicians (e.g., over-the-counter drugs, herbal medications). • Obtaining adherence information. • Useful when timing of followup may not be concordant with timing of clinical encounter. • Obtaining information about the patient not available otherwise (e.g., device data)	• Literacy, language, or other barriers that may lead to under-enrollment of some subgroups. • Validated data collection instruments may need to be developed. • Loss to followup or refusal to continue participation. • Non-response • Limited confidence in reporting clinical information and utilization information. • May not be usable in its raw form; may be necessary to compute a summary metric.
	Clinician-reported data	• More specific information than available from coded data or medical record.	• Clinicians are highly sensitive to burden. • Missing data • Consistency in capture of patient signs, symptoms, use of nonprescribed therapy varies.

Data Source Category	Data Source	Strength and Uses	Limitations
Secondary Sources	**Electronic health records (EHRs)**	• Information on routine medical care and practice, with more clinical context than coded claims. • Potential for comprehensive view of patient medical and clinical history within a given health system, or from multiple health systems, if obtaining EHR data directly from patient. • Efficient access to medical and clinical data. • Use of data transfer and coding standards (including handling of missing data) will increase the quality of data incorporated into the registry.	• The underlying information is not collected in a systematic way. For example, a diagnosis of bacterial pneumonia by one physician may be based on a physical exam and patient report of symptoms, while another physician may record the diagnosis only in the presence of a confirmed laboratory test. The movement to standardized value sets for electronic medical records addresses this issue, but such sets are not yet generally adopted. • It is difficult to interpret missing data. For example, absence of a specific symptom in the visit record may indicate that the symptom was not present or that the physician did not actively inquire about this specific symptom or set of symptoms. • Consistency of data quality and breadth of data collected varies across sites. • Difficult to handle information that has been uploaded into the EHRs (e.g., scanned clinician reports) vs. direct entry into data fields. • Historical data capture may require manual chart abstraction prior to implementation date of medical records system. • Complete medical and clinical history may not be available (e.g., new patient to clinic). • EHR systems vary widely. If data come from multiple systems, the registry should plan to work with each one individually to understand the quality of the underlying information and its suitability for use.

Data Source Category	Data Source	Strength and Uses	Limitations
	Ancillary clinical information systems	• May include more comprehensive information on laboratory results, diagnostic evaluations or treatments than what is available in the EHR.	• Important to be knowledgeable about coding systems used in entering data into the original systems. • The use of ancillary clinical information systems varies by health system. The registry should plan to work with each system individually to understand the quality of the underlying information and its suitability for use.
	Clinical data warehouses (CDWs) or integrated data repositories (IDRs)	• Harmonized information from the EHR and potentially other ancillary clinical systems. • May include legacy clinical information not present in the EHR. • Potential resource utilization (e.g., days in hospital). • May incorporate cost data (e.g., billed and/or paid amounts from insurance claims submissions).	• Important to be knowledgeable about the underlying data model, the coding systems used in the original source system(s) and the transformation processes used to populate the repository. • The use of CDWs and IDRs varies widely by institution. The registry should plan to work with each system individually to understand the quality of the underlying information and its suitability for use.

Data Source Category	Data Source	Strength and Uses	Limitations
	Administrative (claims) databases	• Useful for tracking healthcare resource utilization and cost-related information. • Range of data includes anything that is reimbursed by health insurance, generally including visits to physicians and allied health providers, most prescription drugs, many devices, hospitalization(s), if a lab test was performed, and in some cases, actual lab test results for selected tests (e.g., blood test results for cholesterol, diabetes). • In some cases, demographic information (e.g., gender, date of birth from billing files) can be obtained. • Potential for efficient capture of large populations.	• Represents clinical cost drivers rather than complete clinical diagnostic and treatment information. • Important to be knowledgeable about the process and standards used in claims submission. For example, only primary diagnosis may be coded and secondary diagnoses not captured. In other situations, value-laden claims may not be used (e.g., an event may be coded as a "nonspecific gynecologic infection" rather than a "sexually transmitted disease"). • Important to be knowledgeable about data handling and coding systems used when incorporating the claims data into the administrative systems. • Can be difficult to gain the cooperation of partner groups, particularly in regard to receiving the submissions in a timely manner. • May require that data be purchased.

Data Source Category	Data Source	Strength and Uses	Limitations
	Death indexes	• Completeness—death reporting is mandated by law in the United States. • Strong backup source for mortality tracking (e.g., patient lost to followup). • National Death Index (NDI)— centralized database of death records from State vital statistics offices; database updated annually. • NDI causes of death relatively reliable (93–96%) compared with State death certificates. • Social Security Administration's (SSA) Death Master File—database of deaths reported to SSA; database updated weekly.	• Time delay—indexes depend on information from other data sources (e.g., State vital statistics offices), with delays of 12 to 18 months or longer (NDI). It is important to understand the frequency of updates of specific indexes that may be used. • Absence of information in death indexes does not necessarily indicate "alive" status at a given point in time. • Most data sources are country specific and thus do not include deaths that occurred outside of the country. • As of November 2011, Death Master File no longer includes protected State records. • Lack of complete patient identifier may pose challenge linking with data from other data sources.
	Aggregate/non-patient level databases (e.g., U.S. Census Bureau, Health Care Utilization Project, Area Health Resources File)	• Can be used to provide population estimates or socio-economic characteristics of a given area or region (U.S. Census Bureau databases). • Provide additional details on providers or medical facilities. • Allow additional understanding of target registry population.	• Each database targets participants via different survey sampling methodology and estimates. • Does not provide subject-level data. • May not be linkable with the registry database.

Data Source Category	Data Source	Strength and Uses	Limitations
	Existing registries	• Can be merged with another data source to answer additional questions not considered in the original registry protocol or plan. • May include specific data not generally collected in routine medical practice. • Can provide historical comparison data. • Reduces data collection burden for sites, thereby encouraging participation.	• Important to understand the existing registry protocol or plan to evaluate data collected for element definitions, timing, and format, as it may not be possible to merge data unless many of these aspects are similar. • Creates a reliance on the other registry. Other registry may end. • Other registry may change data elements (which highlights the need for regular communication). • Some sites may not participate in both. Must rely on the data quality of the other registry.
	Distributed research networks	• May have EHR and/or claims data available for large populations of patients in a standardized CDM. • Can be used to augment registry data without having to work with each individual health system on data transformations	• Important to understand the processes used by the network to transform data into the CDM and to assess the quality of the underlying source data. • Creates a reliance on another entity. • Networks may change their underlying data model, which can affect the availability/quality of certain data. • Some sites may not participate in both. • Regulatory/legal requirements for data linkage.

Medical chart abstraction—When secondary sources are unstructured (e.g. notes) or registry variables require human interpretation for completion (several areas of the record need to be consulted to make a determination), abstraction may be utilized. Inter-rater reliability measurements of abstractors can assist in understanding the quality of the abstraction. Abstraction may be done manually or by using computational methods to extract information from free text that is stored electronically. Computational methods are referred to as *natural language processing* (NLP). Chapter 11 discusses abstraction methods in more detail.

Electronic health records—Electronic health records (EHRs) are computer systems that are used to document and manage patient care within and across health systems. The last decade has seen a tremendous uptake in the adoption and use of EHRs, due to the EHR Incentive Program ("Meaningful Use" incentives) that was included in the Health Information Technology for Economic and Clinical Health (HITECH) Act (part of the American Recovery and Reinvestment Act legislation passed by the United States Congress in 2009) as well as electronic reporting

requirements for quality measures from the Centers for Medicare and Medicaid Services (CMS).[2-4] These programs resulted in the creation of a certification program that is used to test the functionality of various EHR components. Over 75% of physicians (2015 data) and 96% (2016 data) of non-federal acute care hospitals within the United States report the adoption of a certified EHR.[5] Whereas health systems would previously adopt "best of breed" clinical information systems to handle different components of the care process (ambulatory, inpatient, e-prescribing, laboratory results, etc.), the burden of attempting to integrate different systems has caused the industry to move towards enterprise solutions provided by a single vendor.

EHRs can be used to capture many different types of data – vital signs, patient history, diagnoses/conditions, treatments and therapies, laboratory results, surveys and questionnaires, etc.[6] As such, they contain a wealth of potentially relevant information for a registry. Data in the EHR reflect the practice of medicine or healthcare within a health system and specialty. The use of standard medical practice data can be useful when looking at treatments and outcomes in the real world, including all of the confounders that affect the measurement of effectiveness (as distinguished from efficacy) and safety outside of the controlled conditions of a clinical trial. Documentation within the EHR is variable, and patients who are seen at multiple health systems will have multiple records. While there have been efforts to promote the interoperability of EHRs, there is still a wide variation in coding practices. In addition, while EHRs support the capture of structured text or coded fields, a large percentage of documentation still occurs as free text, which limits reuse without additional processing.

It is worth noting that, within the registry context, an EHR may function as both a primary and a secondary source. An EHR system may include condition-specific data collection forms that can be used to capture standard-of-care data elements that are equivalent to those collected in a patient registry.[7] Completion of this form would constitute primary data collection, while electronically transferring laboratory results from the EHR to the registry would constitute use of secondary data.

Ancillary clinical information systems—Even with the widespread adoption of EHRs, many health systems still use ancillary clinical information systems to manage specialized workflows. Examples include radiology or other imaging systems, genomic repositories, pharmacy systems, and patient monitors. These systems may have an interface with the EHR, but they typically only transmit a small fraction of the information that they collect (e.g., interpretation of an echocardiogram vs. all of the data generated during the procedure). Due to their specialized nature, these systems may not be used for reporting or analysis to the same degree as enterprise systems like the EHR, making it more difficult to obtain the underlying source data.

Clinical data warehouses or integrated data repositories—Institutions or health systems also typically maintain one more integrated data repositories that pull together data from the EHR and other systems into a common, standardized data model. Such systems may also be called a data warehouse or a data lake, depending on the level of standardization and harmonization. Institutions or health systems may develop their own data model, purchase one from a vendor, or adopt one of several common data models (CDMs) that have been developed to support clinical, observational and comparative effectiveness research. These models include i2b2, Sentinel, PCORnet, HSCRN and OMOP/OHDSI. While it can be appealing to obtain data that has been standardized into a common model, particularly if that model is utilized by many health systems,

it is necessary to understand how the model relates to the way the information was captured in the source system and whether that representation changes the meaning of the data for the purpose of the registry. For instance, an EHR may have an encounter type of "Social Work," while a CDM may only allow a handful of values for encounter type (Ambulatory visit, ED encounter, Inpatient stay, Other). Collapsing these values can make it difficult to separate out the relevant information, so additional steps must be taken in order to ensure that the information incorporated into the registry is correct. More information on CDMs and data repositories can be found in the eBook on Tools and Technologies for Registry Interoperability.[1]

Administrative (claims) databases—Private and public medical insurers collect a wealth of information in the process of tracking healthcare, evaluating coverage, and managing billing and payment. Information in the databases includes patient-specific information (e.g., insurance coverage and copays; identifiers such as name, demographics, SSN or plan number, and date of birth) and healthcare provider descriptive data (e.g., identifiers, specialty characteristics, locations). Typically, private insurance companies organize healthcare data by physician care (e.g., physician office visits) and hospital care (e.g., emergency room visits, hospital stays). Data include procedures and associated dates, as well as costs charged by the provider and paid by the insurers. Amounts paid by insurers are often considered proprietary and unavailable. Standard coding conventions are used in the reporting of diagnoses, procedures, and other information. Coding conventions include the Current Procedure Terminology (CPT) for physician services and International Classification of Diseases (ICD) for diagnoses and hospital inpatient procedures. The databases serve the primary function of managing and implementing insurance coverage, processing, and payment. (See Case Example 13.)

Medicare and Medicaid claims files are commonly used administrative databases in the United States. Together, the programs cover nearly 133 million people in the United States. The Medicare program covers some 59 million individuals ages 65 and older, as well as younger individuals with end-stage renal disease or who qualify for Social Security Disability.[8] Medicaid and Children's Health Insurance Program (CHIP) together cover an additional 73.8 million individuals.[9] Both programs are administered by the Centers for Medicare and Medicaid Services (CMS). Claim files for these programs can be obtained for inpatient and outpatient visits, skilled nursing facility stays, durable medical equipment, hospital services, and prescription drugs. These data, which are subject to privacy rules and regulations, can be linked to other databases with appropriate permissions. The Research Data Assistance Center (ResDAC) is a CMS contractor that supports researchers interested in using Medicare and/or Medicaid data for research purposes.[10]

Death and birth records—Death indexes are national databases tracking population death data (e.g., the NDI[11] and the Death Master File [DMF] of the Social Security Administration [SSA][12]). Data include patient identifiers, date of death, and attributed causes of death. These indexes are populated through a variety of sources. For example, the DMF includes death information on individuals who had an SSN and whose death was reported to the SSA. Reports may come in to the SSA by different paths, including from survivors or family members requesting benefits or from funeral homes. Because of the importance of tracking Social Security benefits, all States, nursing homes, and mortuaries are required to report all deaths to the SSA. Prior to 2011, the DMF contained virtually 100-percent complete mortality ascertainment for those eligible for SSA benefits. As of November 2011, however, the DMF no longer includes

protected State death records. In practical terms, this means that approximately 4.2 million records were removed from the historical public DMF (which contained 89 million records), and some 1 million fewer records will be added to the DMF each year.[13] The NDI can be used to provide both fact of death and cause of death, as recorded on the death certificate. Cause-of-death data in the NDI are relatively reliable (93–96 percent) compared with death certificates.[14,15] Time delays in death reporting should be considered when using these sources, and vital status should not be assumed to be "alive" by the absence of information at a recent point in time. These indexes are valuable sources of data for death tracking. Of course, mortality data can be accessed directly through queries of State vital statistics offices and health departments when targeting information on a specific patient or within a State. Likewise, birth certificates are available through State departments and may be useful in registries of children or births.

Aggregate/non-patient-level databases—Databases that provide aggregate, non-patient-level statistics may be valuable resources to augment an existing registry. These databases may contain area or population-level statistics, details about providers or medical facilities, or de-identified encounter details. The frequency with which these databases are updated varies by source. Depending on the level of aggregation, it may be possible to link these data to a registry database (i.e., generating neighborhood-level socioeconomic information via geocoding). Two sources of area-level data are the U.S. Census and the Area Health Resources Files (AHRF). The U.S. Census Bureau databases[16] provide population-level data utilizing survey sampling methodology. The Census Bureau conducts many different surveys, the main one being the population census. The primary use of the data is to determine the number of seats assigned to each State in the House of Representatives, although the data are used for many other purposes. These surveys calculate estimates through statistical processing of the sampled data. Estimates can be provided with a broad range of granularity, from population numbers for large regions (e.g., specific States), to ZIP Codes, all the way down to a household level (e.g., neighborhoods identified by street addresses). Information collected includes demographic, gender, age, education, economic, housing, and work data. The data are not collected at an individual level but may serve other registry purposes, such as understanding population numbers in a specific region or by specific demographics. The AHRF is maintained by the Health Resources and Services Administration, which is part of the Department of Health and Human Services. The AHRF includes county-level data on health facilities, health professions, measures of resource scarcity, health status, economic activity, health training programs, and socioeconomic and environmental characteristics.[17] The Environmental Protection Agency maintains datasets of air quality and other measures and has developed a number of methods for estimating exposure.[18,19]

Data on medical facilities and physicians may be important for categorizing registry data or conducting subanalyses. Two sources of such data are the American Hospital Association's Annual Survey Data and the American Medical Association's Physician Masterfile Data Collection. The Annual Survey Data is a longitudinal database that collects 700 data elements, covering organizational structure, personnel, hospital facilities and services, and financial performance, from more than 6,000 hospitals in the United States.[20] Each hospital in the database has a unique ID, allowing the data to be linked to other sources; however, there is a data lag of about 2 years, and the data may not provide enough nuanced detail to support some analyses of cost or quality of care. The Physician Masterfile Data Collection contains current and historic

data on nearly one million physicians and residents in the United States. Data on physician professional medical activities, hospital and group affiliations, and practice specialties are collected each year. The National Plan and Provider Enumeration System (NPPES) also contains information about healthcare providers and can be used to provide additional details if the registry captures National Provider Identifiers (NPIs).[21]

Databases of individual patient encounters (e.g., physician office visits, emergency department visits, hospital inpatient stays), generally do not contain individual patient identifiers and thus may not be linkable to patient registries, but nevertheless provide valuable insight into the makeup of the registry's target population. This is particularly true for data from nationally representative surveys, such as AHRQ's Health Care Utilization Project (H-CUP), Nationwide Inpatient Sample (NIS), and the suite of surveys by the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS), including the National Ambulatory Medical Care Survey (NAMCS), the National Hospital Ambulatory Medicare Care Survey (NHAMCS), and the National Hospital Discharge Survey (NHDS).

Existing registries and other databases—There are numerous national and regional registries and other databases that may be leveraged for incorporation into other registries (e.g., disease-specific registries managed by nonprofit organizations, professional societies, or other entities). An example is the National Marrow Donor Program (NMDP),[22] a global database of cord blood units and volunteers who have consented to donate marrow and blood cells. Databases maintained by the NMDP include identifiers and locators in addition to information on the transplants, such as samples from the donor and recipient, histocompatibility, and outcomes. NMDP actively encourages research and utilization of registry data through a data application process and submission of research proposals. Other registries may also be valuable sources of data. Resources such as ClinicalTrials.gov and HSRProj are useful for searching for and identifying relevant registries to contact about data sharing or research collaborations.

Distributed research networks—Distributed research networks (DRNs) may be a possible way to obtain EHR or claims data on a registry population. A number of established DRNs exist in the United States, including the Sentinel Initiative (formerly Mini-Sentinel),[23,24] the Health Care Systems Research Network (HCSRN) (formerly the Health Maintenance Organization Research Network (HMORN)),[25-27] and the National Patient-Centered Clinical Research Network (PCORnet).[28,29] These networks are used to support a range of research activities, including pharmacovigilance studies, pragmatic clinical trials, and studies of treatment effectiveness.[30-34] Within a DRN, partners (sites) typically standardize their data into a CDM, with the data refreshed at a specified frequency (i.e., quarterly). After each refresh, partners will usually execute a data curation package to assess the underlying data quality.[35] Partners whose data pass the required checks can then respond to network queries. Data in a DRN typically remain at the local level (behind the network partner's firewall), with analyses done at the local level and only results, in the form of aggregate counts or summary statistics, returned to the requestor. However, many DRNs have provisions to allow the exchange of patient-level data in some contexts.[36,37] Registries that maintain patient identifiers may be able to link to DRN data to obtain greater detail on their population than can reasonably collected within the registry itself, but technical and governance issues must be resolved before any linkage can actually occur.

5. Other Considerations for Secondary Data Sources

The discussion below focuses on logistical and data issues to consider when incorporating data from other sources. Chapter 11 fully explores data collection, management, and quality assurance for registries.

In accessing data from one registry for the purposes of another, it is important to recognize that data may have changed during the course of the source registry, and this may or may not have been well documented by the providers of the data. For example, in the United States Renal Data System (USRDS),[38] a vital part of personal identification is CMS 2728, an enrollment form that identifies the incident data for each patient as well as other pertinent information, such as the cause of renal failure, initial therapy, and comorbid conditions. Originally created in 1973, this form is in its third version, having been revised in 1995 and again in 2005. Consequently, there are data elements that exist in some versions and not others. In addition, the coding for some variables has changed over time. For example, race has been redefined to correspond with Office of Management and Budget directives and Census Bureau categories. Furthermore, form CMS 2728 was optional in the early years of the registry, so until 1983 it was filled out for only about one-half of the subjects. Since 1995, it has been mandatory for all people with end-stage renal disease. These changes in form content, data coding, and completeness would not be evident to most researchers trying to access the data.

Before incorporating a secondary data source into a registry, it is critical to consider the potential impact of the data quality of the secondary data source on the overall data quality of the registry. The potential impact of quality issues in the secondary data sources depends on how the data are used in the primary registry. For example, quality would be significant for secondary data that are intended to be populated throughout the registry (i.e., used to populate specific data elements in the entire registry over time), particularly if these populated data elements are critical to determining a primary outcome. Quality of the secondary data will have less effect on overall registry quality if the secondary data are to be linked to registry data only for a specific analytic study. For more information on data quality, see Chapter 11.

The importance of patient identifiers for linking to secondary data sources cannot be overstated. Multiple patient identifiers should be used, and primary data for these identifiers should not be entered into the registry unless the identifying information is complete and clear. While an SSN is very useful, high-quality probabilistic linkages can be made to secondary data sources using various combinations of such information as name (last, middle initial, and first), date of birth, and gender. For example, the NDI will make possible matches when at least one of seven matching conditions is met (e.g., one matching condition is “exact month and day of birth, first name, and last name”). However, the degree of success in such probabilistic and deterministic matching generally is enhanced by having many identifiers to facilitate matching. As noted earlier, the various types of data (e.g., personal history, adverse events, hospitalization, and drug use) have to be linked through a common identifier.

The best identifier is one that is not only unique but has no embedded personal identification, unless that information is scrambled and the key for unscrambling it is stored remotely and securely. The group operating the registry should have a process by which each new entry to the

registry is assigned a unique code and there is a crosswalk file to enable the system to append this identifier to all new data as they are accrued. The crosswalk file should not be accessible by people or entities outside the management group.

In addition, consideration should be given to the fact that a registry may need to accept and link datasets from more than one outside organization. Each institution contributing data to the registry will have unique requirements for patient data, access, privacy, and duration of use. While having identical agreements with all institutions would be ideal, this may not always be possible from a practical perspective. Yet all registries have resource constraints, and decisions about including certain institutions have to be determined based on the resources available in order to negotiate specialized agreements or to maintain specialized requirements. Agreements should be coordinated as much as possible so that the function of the registry is not greatly impaired by variability among agreements. All organizations participating in the registry should have a common understanding of the rules regarding access to the data. Although exceptions can be made, it should be agreed that access to data will be based on independent assessment of research protocols and that participating organizations will not have individual veto power over access.

When data from secondary sources are used, agreements should specify ownership of the source data and clearly permit data use by the recipient registry. The agreements should also specify the roles of each institution, its legal responsibilities, and any oversight issues. It is critical that these issues and agreements be put in place before data are transferred so that there are no ambiguities or unforeseen restrictions on the recipient registry later on.

Some registries may wish to incorporate data from more than one country. In these cases, it is important to ensure that the data are being collected in the same manner in each country or to plan for any necessary conversion. For example, height and weight data collected from sites in Europe will likely be in different units than height and weight data collected from sites in the United States. Laboratory test results may also be reported in different units, and there may be variations in the types of pharmaceutical products and medical devices that are approved for use in the participating countries. Understanding these issues prior to incorporating secondary data sources from other countries is extremely important to maintain the integrity and usefulness of the registry database.

When incorporating other data sources, consideration should also be given to the registry update schedule. A mature registry will usually have a mix of data update schedules. The registry may receive an annual update of large amounts of data, or there could be monthly, weekly, or even daily transfers of data. Regardless of the schedule of data transfer, routine data checks should be in place to ensure proper transfer of data. These should include simple counts of records as well as predefined distributions of key variables. Conference calls or even routine meetings to go over recent transfers will help avoid mistakes that might not otherwise be picked up until much later.

An example of the need for regular communication is a situation that arose with the United States Renal Data System a few years ago. The United Network for Organ Sharing (UNOS) changed the coding for donor type in their transplant records. This resulted in an apparent 100-percent loss of living donors in a calendar year. The change was not conveyed to USRDS and

was not detected by USRDS staff. After USRDS learned about the change, standard analysis files that had been sent to researchers with the errors had to be replaced.

6. Summary

In summary, a registry is not a static enterprise. The management of registry data sources requires attention to detail, constant feedback to all participants, and a willingness to make adjustments to the operation as dictated by changing times and needs.

References for Chapter 6

1. Gliklich RE, Leavy MB, Dreyer NA, (eds). Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide. 3rd Edition, Addendum II. (Prepared by L&M Policy Research, LLC under Contract No. 290-2014-00004-C.) AHRQ Publication No. 17(18)-EHC017-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2019. https://effectivehealthcare.ahrq.gov/.

2. Blumenthal D. Launching HITECH. N Engl J Med. 2010;362(5):382-5. PMID: 20042745. DOI: 10.1056/NEJMp0912825.

3. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363(6):501-4. PMID: 20647183. DOI: 10.1056/NEJMp1006114.

4. Centers for Medicare and Medicaid Services. Physician Quality Reporting System (PQRS). https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/PQRS_OverviewFactSheet_2013_08_06.pdf. Accessed June 20, 2019.

5. The Office of National Coordinator of Health Information Technology. U.S. Department of Health and Human Services. Health IT Dashboard. https://dashboard.healthit.gov/index.php. Accessed June 10, 2019.

6. Marsolo K, Spooner SA. Clinical genomics in the world of the electronic health record. Genet Med. 2013;15(10):786-91. PMID: 23846403. DOI: 10.1038/gim.2013.88.

7. Marsolo K, Margolis PA, Forrest CB, et al. A Digital Architecture for a Network-Based Learning Health System: Integrating Chronic Care Management, Quality Improvement, and Research. EGEMS (Wash DC). 2015;3(1):1168. PMID: 26357665. DOI: 10.13063/2327-9214.1168.

8. Centers for Medicare & Medicaid Services. Medicare Enrollment Dashboard. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Dashboard/Medicare-Enrollment/Enrollment%20Dashboard.html. Accessed June 10, 2019.

9. Centers for Medicare and Medicaid Services. March 2019 Medicaid & CHIP Enrollment Data Highlights. https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/report-highlights/index.html. Accessed June 10, 2019.

10. Research Data Assistance Center (ResDAC). https://www.resdac.org/cms-data. Accessed June 10, 2019.

11. Centers for Disease Control and Prevention. National Death Index. https://www.cdc.gov/nchs/ndi/index.htm. Accessed June 20, 2019.

12. Social Security Administration. Death Master File. National Technical Information Service. https://dmf.ntis.gov/. Accessed June 20, 2019.

13. National Technical Information Service. Important Notice: Change in Public Death Master File Records. https://classic.ntis.gov/assets/pdf/import-change-dmf.pdf. Accessed June 20, 2019.

14. Doody MM, Hayes HM, Bilgrad R. Comparability of national death index plus and standard procedures for determining causes of death in epidemiologic studies. Ann Epidemiol. 2001;11(1):46-50. PMID: 11164119.

15. Sathiakumar N, Delzell E, Abdalla O. Using the National Death Index to obtain underlying cause of death codes. J Occup Environ Med. 1998;40(9):808-13. PMID: 9777565.

16. U.S. Bureau of the Census. http://www.census.gov. Accessed June 20, 2019.

17. Health Resources and Services Administration. Area Health Resources Files (AHRF). https://data.hrsa.gov/topics/health-workforce/ahrf. Accessed June 20, 2019.

18. U.S. Environmental Protection Agency. RSIG Data Inventory. https://www.epa.gov/hesc/rsig-data-inventory. Accessed June 20, 2019.

19. McMillan NJ, Holland DM, Morara M, et al. Combining numerical model output and particulate data using Bayesian space–time modeling. Environmetrics. 2010;21(1):48-65. DOI: 10.1002/env.984.

20. American Hospital Association. AHA Data and Directories. https://www.aha.org/other-resources/2018-01-08-aha-data-and-directories. Accessed June 20, 2019.

21. U.S. Department of Health and Human Services. Centers for Medicare & Medicaid Services. Data Dissemination. https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/DataDissemination.html. Accessed June 20, 2019.

22. National Marrow Donor Program. http://www.marrow.org. Accessed June 20, 2019.

23. Carnahan RM, Moores KG. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative and claims data: methods and lessons learned. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 1:82-9. PMID: 22262596. DOI: 10.1002/pds.2321.

24. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 1:23-31. PMID: 22262590. DOI: 10.1002/pds.2336.

25. Ross TR, Ng D, Brown JS, et al. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. EGEMS (Wash DC). 2014;2(1):1049. PMID: 25848584. DOI: 10.13063/2327-9214.1049.

26. Steiner JF, Paolino AR, Thompson EE, et al. Sustaining Research Networks: the Twenty-Year Experience of the HMO Research Network. EGEMS (Wash DC). 2014;2(2):1067. PMID: 25848605.

27. Vogt TM, Lafata JE, Tolsma DD, et al. The Role of Research in Integrated Health Care Systems: The HMO Research Network. Perm J. 2004;8(4):10-7. PMID: 26705313.

28. Califf RM. The Patient-Centered Outcomes Research Network: a national infrastructure for comparative effectiveness research. N C Med J. 2014;75(3):204-10. PMID: 24830497. DOI: http://www.ncmedicaljournal.com/content/75/3/204.long.

29. Fleurence RL, Curtis LH, Califf RM, et al. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc. 2014;21(4):578-82. PMID: 24821743. DOI: 10.1136/amiajnl-2014-002747.

30. Toh S, Platt R, Steiner JF, et al. Comparative-effectiveness research in distributed health data networks. Clin Pharmacol Ther. 2011;90(6):883-7. PMID: 22030567. DOI: 10.1038/clpt.2011.236.

31. Hernandez AF, Fleurence RL, Rothman RL. The ADAPTABLE Trial and PCORnet: Shining Light on a New Research Paradigm. Ann Intern Med. 2015;163(8):635-6. PMID: 26301537. DOI: 10.7326/m15-1460.

32. Johnston A, Jones WS, Hernandez AF. The ADAPTABLE Trial and Aspirin Dosing in Secondary Prevention for Patients with Coronary Artery Disease. Curr Cardiol Rep. 2016;18(8):81. PMID: 27423939. DOI: 10.1007/s11886-016-0749-2.

33. Block JP, Bailey LC, Gillman MW, et al. PCORnet Antibiotics and Childhood Growth Study: Process for Cohort Creation and Cohort Description. Acad Pediatr. 2018;18(5):569-76. PMID: 29477481. DOI: 10.1016/j.acap.2018.02.008.

34. Toh S, Rasmussen-Torvik LJ, Harmata EE, et al. The National Patient-Centered Clinical Research Network (PCORnet) Bariatric Study Cohort: Rationale, Methods, and Baseline Characteristics. JMIR Res Protoc. 2017;6(12):e222. PMID: 29208590. DOI: 10.2196/resprot.8323.

35. Qualls LG, Phillips TA, Hammill BG, et al. Evaluating Foundational Data Quality in the National Patient-Centered Clinical Research Network (PCORnet(R)). EGEMS (Wash DC). 2018;6(1):3. PMID: 29881761. DOI: 10.5334/egems.199.

36. Brown JS, Holmes JH, Shah K, et al. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Med Care. 2010;48(6 Suppl):S45-51. PMID: 20473204. DOI: 10.1097/MLR.0b013e3181d9919f.

37. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. Ann Intern Med. 2009;151(5):341-4. PMID: 19638403.

38. United States Renal Database. https://www.usrds.org/. Accessed June 20, 2019.

Case Examples for Chapter 6

Case Example 13. Using Claims Data Along With Patient-Reported Data To Identify Patients

Description	The National Amyotrophic Lateral Sclerosis (ALS) Registry is a rare disease registry created by the Agency for Toxic Substances and Disease Registry (ATSDR) within the U.S. Centers for Disease Control and Prevention (CDC), Department of Health and Human Services (HHS). The purpose of the registry is to quantify the incidence and prevalence of ALS in the United States, describe the demographics of people with ALS, and examine potential risk factors for the disease.
Sponsor	U.S. Department of Health and Human Services and Agency for Toxic Substances and Disease Registry, through funding from the "ALS Registry Act" (U.S. Congress Public Law 110-373).
Year Started	2010
Year Ended	Ongoing
No. of Sites	All 50 States, including U.S. territories; data from national administrative databases are combined with patient self-enrollment data.
No. of Patients	16,583 as of 2015; prevalence estimates are released annually for the successive calendar year

Challenge

Amyotrophic lateral sclerosis (ALS) is a progressive, fatal neurodegenerative disorder of both the upper and lower motor neurons. Many knowledge gaps exist in the understanding of ALS, including uncertainty about the disease's incidence and prevalence, misdiagnosis of ALS in patients with other motor neuron disorders, and the role of environmental exposures in the etiology of ALS. Because ALS is a non-reportable disease in the United States (except for the Commonwealth of Massachusetts), previous attempts to estimate ALS incidence and prevalence using nonspecific mortality data have faced many challenges and at best overestimated disease frequency. Identifying patients through site recruitment for research purposes poses additional challenges, as access to patient medical records can be limited, costly, and time-consuming to obtain. Patient recruitment issues are compounded by the complexities of this rare disease, in which the average timeframe from diagnosis to death is 2–5 years. U.S. governmental agencies acknowledged that a national, structured data collection program for ALS was greatly needed, and that alternative data sources and recruitment strategies would need to be identified.

Proposed Solution

In 2008, President Bush signed the ALS Registry Act into law, allowing ATSDR to create the National ALS Registry. The registry is the only Congressionally mandated population-based ALS registry in the United States. As a first step in developing the registry, a workshop of international experts in neurological and autoimmune conditions was convened to discuss approaches to creating a national database. Based on feedback from these experts, the registry

uses a two-pronged approach to identify all U.S. cases of ALS. The first approach uses national administrative databases, including those of Medicare, Medicaid, the Veterans Health Administration, and the Veterans Benefit Administration, to identify prevalent cases based on an algorithm developed through pilot projects. These administrative databases cover approximately 90 million Americans, and the algorithm identifies 80 to 85 percent of all true ALS cases when applied to these databases. The second approach uses a secure Web portal to allow patients to self-enroll voluntarily. Data from the two approaches are combined into the registry database, and duplicate patients are identified and removed so that each person with ALS is counted only once in the registry.

Results

The National ALS Registry has funded over 15 research projects, such as evaluating environmental risk factors and possible etiologies for ALS. In addition, the Registry has published 65 articles and more than 50 abstracts in peer-reviewed publications. The Web portal for self-enrolled participants contains 17 brief surveys that collect information on potential risk factors, such as socio-demographic characteristics, occupational history, military history, cigarette smoking, alcohol consumption, physical activity, family history of neurodegenerative diseases, and disease progression. To date, over 80,000 surveys have been completed by Registry enrollees. ATSDR also funded active surveillance projects that allowed population-based case estimates of ALS in certain smaller geographic areas (i.e., at the State and metropolitan levels) to help ATSDR evaluate the completeness of the registry. In addition, ATSDR has developed a system to inform people with ALS about new research (e.g., clinical trials, epidemiological studies) for which they may be eligible. To date, the Registry has help to recruit for over 45 research projects. Lastly, the Registry now includes a national biorepository that is designed to help researchers better understand the disease by pairing biospecimens (e.g., blood, brain tissue) with existing risk-factor data from patients. Thousands of biospecimens are currently available to researchers for analysis.

Key Point

Combining multiple data sources, such as administrative databases and patient-reported information, is a novel and effective way to successfully identify patients with a rare disease and to better understand the prevalence, incidence, and etiology of the disease. However, using alternative approaches requires a strong understanding of the nuances of the individual data sources; pilot testing is also helpful to identify potential issues with data sources prior to registry launch.

For More Information

- http://wwwn.cdc.gov/als
- Mehta P, Kaye W, Raymond J, et al. Prevalence of Amyotrophic Lateral Sclerosis — United States, 2015. MMWR Morb Mortal Wkly Rep 2018;67:1285–1289. DOI: http://dx.doi.org/10.15585/mmwr.mm6746a1.
- Mehta P, Horton DK, Kasarskis EJ, et al. CDC Grand Rounds: National Amyotrophic Lateral Sclerosis (ALS) Registry Impact, Challenges, and Future Directions. MMWR Morb Mortal Wkly Rep 2017;66:1379–1382. DOI: http://dx.doi.org/10.15585/mmwr.mm6650a3.

- Kaye W, Wagner L, Wu R, Mehta P. Evaluating the completeness of the national ALS registry, United States. Amyotroph Lateral Scler Frontotemporal Degener; 2018 Feb;19(1-2):112-117. PMID: 29020837. DOI: 10.1080/21678421.2017.1384021.
- Horton DK, Kaye W, Wagner L. Integrating a Biorepository into the National Amyotrophic Lateral Sclerosis Registry. J Environ Health. 2016 Nov;79(4):38-40. PMID: 28935999.
- Horton DK, Graham S, Punjani R, et al. A spatial analysis of amyotrophic lateral sclerosis (ALS) cases in the United States and their proximity to multidisciplinary ALS clinics. Amyotroph Lateral Scler Frontotemporal Degener. 2018 Feb;19(1-2):126-133. PMID: 29262737. DOI: 10.1080/21678421.2017.1406953.

Case Example 14. Using a Patient-Centered Study Design To Collect Informed Consent, Maximize Recruitment and Retention, and Provide Meaningful Clinical Data

Description	Function and Outcomes Research for Comparative Effectiveness in Total Joint Replacement (FORCE-TJR) is a prospective research registry tracking and studying long-term outcomes of elective total joint replacement (TJR) surgery. The registry seeks to understand patient-reported and clinical outcomes by collecting data on baseline patient attributes, procedure approach and technology, inpatient hospital stay, surgeon and institutional characteristics, longitudinal patient pain and function, and post-procedure complications and revisions. A diverse patient cohort allows the generation of aggregate severity-adjusted national and regional data against which participating surgeons can compare their own practice.
Sponsor	Funded in part by grants from the Agency for Healthcare Research and Quality and the Patient-Centered Outcomes Research Institute to the University of Massachusetts Medical School.
Year Started	2011
Year Ended	Ongoing
No. of Sites	Over 200 orthopedic surgeons in 28 states
No. of Patients	>50,000

Challenge

Total joint replacement (TJR) is a common procedure, with more than 700,000 primary hip and knee replacements performed in the United States each year. Although TJR can result in significant pain relief, physical function and activity levels can vary widely after surgery. FORCE-TJR collects data to track patient, provider, and site characteristics in order to evaluate their contributions to patient-reported and clinical outcomes of TJR over time.

TJR patients often have limited contact with their surgeons immediately after making the decision to have surgery, instead interacting with office and hospital staff to complete insurance or anesthesia pre-operative paperwork. Administrative site staff often do not have the time or

training to effectively inform patients about the risks and benefits of participating in patient-centered studies. Further, clinical information that may contain important data about adverse events resides in various, disconnected points of care. Patients may return to the hospital in which TJR was performed or they may present at another hospital, urgent care center, or doctor's office. Often these disparate sites of care are not linked with the same electronic medical record, making data difficult to collect. Collecting informed consent, patient reported outcomes, and other followup data from TJR patients can therefore be challenging and requires an innovative approach.

Proposed Solution

Successful approaches to maximizing patient participation in research are based in creating a relationship with each patient and minimizing the burden on site staff. Patients who schedule a TJR are asked by administrative staff at the participating site to sign a short study contact form, giving permission for registry staff to contact them. Site staff give the patient an informational packet and send the signed contact forms to the registry. To collect informed consent, registry research staff contact patients at their convenience via telephone to review the study procedures, informed consent form, and medical release forms in the informational packet. At this point, patients have the opportunity to ask questions of registry staff and discuss with them any concerns, facilitating a deep understanding of the registry and their role in its success. Patients return the signed informed consent and medical release forms to registry staff via U.S. mail. At the same time, they complete a standardized, patient-reported outcome (PRO) to quantify pain and function before surgery. PROs are repeated at annual intervals after surgery to quantify pain relief and functional gain. Patients also answer brief questions to screen for post-surgical events, including revision surgery or other complications. Collecting clinical data that does not reside in a single medical record also relies upon patient engagement. Upon enrollment in the registry, patients are asked to authorize release of their medical records; at each contact following surgery, patients are asked if they sought medical care since their last contact with the registry. If registry staff determine the medical care could be related to the TJR, the related medical records are obtained and analyzed.

Results

The model described above uses registry staff to enroll patients, obtain informed consent, and deliver longitudinal information and motivation, enhancing participant enrollment and commitment over the long term. This procedure facilitates the longitudinal collection of patient-reported outcomes and medical records data, thus enabling more precise severity adjustment than relying on administrative data. Sites report high satisfaction with the model, contributing to an 80 percent overall patient recruitment ratio in the registry.

Key Point

Registries and other patient-centered research can benefit from a study design that engages patients at enrollment, thereby increasing their participation over the life of the study. For registries that require clinical data from patients who may not access all their care within one system, or that require patient-reported outcome measures, an approach that follows the patient across settings can be beneficial. Contacting patients at their convenience rather than in a healthcare setting can allow them more time to have their questions answered, increasing patient commitment.

For More Information

- https://force-tjr.org/

Case Example 15. Linking a Procedure-Based Registry With Claims Data To Study Long-Term Outcomes

Description	The CathPCI Registry measures the quality of care delivered to patients receiving diagnostic cardiac catheterizations and percutaneous coronary interventions (PCI) in both inpatient and outpatient settings. The primary outcomes evaluated by the registry include the quality of care delivered, outcome evaluation, comparative effectiveness, and postmarketing surveillance.
Sponsor	American College of Cardiology Foundation through the National Cardiovascular Data Registry. Funded by participation dues from catheterization laboratories.
Year Started	1998
Year Ended	Ongoing
No. of Sites	1,698 catheterization laboratories
No. of Patients	23.3 million patient records; 9.6 million PCI procedures

Challenge

The registry sponsor was interested in studying long-term patient outcomes for diagnostic cardiac catheterizations and PCI, but longitudinal data were not collected as part of the registry. Rather than create an additional registry, it was determined that the most feasible option was linking the registry data with available third-party databases such as Medicare.

Before the linkage could occur, however, several legal questions needed to be addressed, including what identifiers could be used for the linkage and whether institutional review board (IRB) approval was necessary.

Proposed Solution

The registry developers explored potential issues relating to the use of protected health information (under the Federal HIPAA [Health Insurance Portability and Accountability Act] law) to perform the linkage; the applicability of the Common Rule (protection of human subjects) to the linkage; and the contractual obligations of the individual legal agreement with each participating hospital with regard to patient privacy. The registry gathers existing data, including direct patient identifiers collected as part of routine healthcare activities. Informed consent is not required. The registry sponsor has business associate agreements in place with participating catheterization laboratories for which the registry conducts the outcomes evaluations.

After additional consultation with legal counsel, the registry sponsor concluded that the linkage of data could occur under two conditions: (1) that the datasets used in the merging process must be in the form of a *limited dataset* (see Chapter 7), and (2) that an IRB must evaluate such

linkage. The decision to implement the linkage was based on two key factors. First, the registry participant agreement includes a data use agreement, which permits the registry sponsor to perform research on a limited dataset but also requires that no attempt be made to identify the patient. Second, since there was uncertainty as to whether the proposed data linkage would meet the definition of research on human subjects, the registry sponsor chose to seek IRB approval, along with a waiver of informed consent. The registry sponsor has a policy that requires that all registry research be conducted consistent with the requirements of the Common Rule.

Results

The registry data were linked with Medicare data, using probabilistic matching techniques to link the limited datasets. A research protocol describing the need for linkage, the linking techniques, and the research questions to be addressed was approved by an IRB. Researchers must reapply for IRB approval for any new research questions they wish to study in the linked data.

Results of the linkage analyses were used to develop a new measure, "Readmission following PCI," for the Centers for Medicare & Medicaid Services' hospital inpatient quality pay-for-reporting program, and researchers have used the linked data to address other questions.

Key Point

There are many possible interpretations of the legal requirements for linking registry data with other data sources. The interpretation of legal requirements should include careful consideration of the unique aspects of the registry, its data, and its participants. In addition, clear documentation of the way the interpretation occurred and the reasoning behind it will help to educate others about such decisions and may allay anxieties among participating institutions.

For More Information

- https://cvquality.acc.org/NCDR-Home/registries/hospital-registries/cathpci-registry

Section II

Legal and Ethical Considerations for Registries

Chapter 7. Principles of Registry Ethics, Data Ownership, and Privacy

1. Introduction

This chapter covers the ethical and legal considerations that are relevant to the development and use of all health information registries, including patient registries as defined in this document, for the purposes of public health activities, governmental health program oversight, quality assurance/improvement (A/I), and research. These considerations include generally accepted ethical principles for the collection and use of health information in connection with research as applied to the establishment and use of registries. Where relevant, this chapter also discusses notable emerging and evolving ethical and legal considerations. Related topics include issues of transparency in the operation of registries, oversight of registry activities, and property rights in healthcare information and registries.

Section 2.1. of this chapter discusses the ethical concerns and considerations involved with obtaining and using confidential health information in registries. Section 2.2. describes the transformation of ethical concerns into the legal regulation of human subjects research and the privacy of individually identifiable health information and other personally identifiable information. In Section 3, an overview is presented of these regulatory requirements and their interactions as they specifically relate to registries. Section 4 makes recommendations about registry transparency and oversight, based on the need to ensure the independence, integrity, and credibility of biomedical research, while preserving and improving the utility of registry data. Finally, property rights in health information and registries are briefly discussed.

Table 7-1 provides an overview of the applicable regulatory requirements based on the type of registry developer and the extent to which registry data are identifiable. The table summarizes key considerations relating to the applicability of and pathways under the Privacy Rule, Common Rule, and FDA GCP regulations. Note that the information in the table is a high-level summary of such considerations, and is not intended to address all of the requirements or considerations that may apply to the development or use of a registry, as such analysis is highly fact-specific. In addition, there may be other laws and individual institutional policies that apply. Each registry is unique. Therefore, this table is not intended to provide answers to specific questions that arise in the context of a given registry. This table is no substitute for consultation with institutional officials and others about the regulatory requirements that apply to a particular registry project.

Table 7-1. Summary of Privacy Rule and Common Rule requirements

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
1A. Federal or State public health agency: Registry for public health practice within agency's legal authority **not** involving research.	The Privacy Rule would not apply to the disclosure to or use by the registry of the de-identified information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. The Common Rule is not applicable. FDA GCP is not applicable.	The Privacy Rule permits use or disclosure to a public health authority for public health activities. The Common Rule is not applicable. FDA GCP is not applicable.	The Privacy Rule permits use or disclosure to a public health authority for public health activities. The Common Rule is not applicable. FDA GCP is not applicable.	Waivers are not applicable as the activity is not research.
1B. Federal or State public health agency: Registry is an agency research project.	The Privacy Rule would not apply to the disclosure to or use by the registry of the de-identified information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. The Common Rule is not applicable. FDA GCP is likely applicable if FDA is the public health agency. IRB review and documented	The Privacy Rule permits the use or disclosure of a limited dataset, provided the data source and registry developer enter into a DUA and downstream users of the registry enter into a DUA with the registry developer. The Common Rule may apply if the agency is a Common Rule Agency.** In addition, determine if a "human subject" is involved. FDA GCP is likely applicable if FDA is the public health agency. IRB review and	The Privacy Rule permits use or disclosure with patient authorization or IRB or privacy board waiver of authorization. If the Common Rule applies,** IRB review and documented consent are required unless an IRB grants a waiver of documentation or waiver for the consent process. FDA GCP is likely applicable if FDA is the public health agency. IRB review and documented consent would be required unless an IRB grants	See applicable regulatory criteria for waiver under HIPAA, Common Rule, and FDA requirements.

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
	consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	a waiver of documentation or waiver for the consent process.	
2. Registry producing evidence in support of labeling for an FDA-regulated product.	The Privacy Rule would not apply to the disclosure to or use by the registry of the de-identified information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. The Common Rule is not applicable. FDA GCP applies to registry use and may apply to registry development. Where applicable, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	The Privacy Rule permits the use or disclosure of a limited dataset for purposes of research, provided the data source and registry developer enter into a DUA and downstream users of the registry enter into a DUA with the registry developer. The Common Rule may apply.** In addition, determine if a "human subject" is involved. FDA GCP applies to registry use and may apply to registry development. Where applicable, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	The Privacy Rule permits use or disclosure with patient authorization or IRB or privacy board waiver of authorization. If the Common Rule applies,** IRB review and documented consent are required unless an IRB grants a waiver of documentation or waiver for the consent process. FDA GCP applies to registry use and may apply to registry development. Where applicable, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	See applicable regulatory criteria for waiver under HIPAA, Common Rule, and FDA requirements.
3. Health oversight agency registry to perform a health oversight activity **not** involving research.	The Privacy Rule would not apply to the disclosure to or use by the registry of the de-identified	The Privacy Rule permits use or disclosure for health oversight activities authorized by law.	The Privacy Rule permits use or disclosure for health oversight activities authorized by law.	Waiver of authorization is not applicable under the Privacy Rule as the activity is not research.

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
	information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. The Common Rule is not applicable. FDA GCP is likely not applicable as the activity is not research (even if FDA is the health oversight agency).	Institutional policy may apply the Common Rule or require IRB review. If so, determine if a "human subject" is involved. FDA GCP is likely not applicable as the activity is not research (even if FDA is the health oversight agency).	Institutional policy may apply the Common Rule or require IRB review. If so, IRB review and documented consent are required unless an IRB grants a waiver of documentation or waiver for the consent process. FDA GCP is likely not applicable as the activity is not research (even if FDA is the health oversight agency).	If institutional policy applies the Common Rule, IRB approval of a waiver of consent documentation or process depends on satisfaction of specific regulatory criteria. FDA GCP is likely not applicable as the activity is not research (even if FDA is the health oversight agency).
4. Quality A/I registry where a secondary purpose of the registry involves research.	The Privacy Rule would not apply to the disclosure to or use by the registry of the de-identified information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. The Common Rule is not applicable. FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of	The Privacy Rule permits the use or disclosure of a limited dataset for healthcare operations, provided the data source and registry developer enter into a data use agreement and downstream users of the registry enter into a DUA with the registry developer. A healthcare operations pathway may apply to the creation of the registry insofar as the primary purpose of the registry is not the obtaining of generalizable knowledge. Secondary uses of the registry may require either a research or healthcare	The Privacy Rule permits use or disclosure for the "healthcare operations" of the data source and in certain circumstances, of another covered entity. A healthcare operations pathway may apply to the creation of the registry insofar as the primary purpose of the registry is not the obtaining of generalizable knowledge. Secondary uses of the registry may require either a research or healthcare operations pathway, depending on the purpose. The Common Rule may apply as one purpose includes research.** If so, IRB review	Waiver of authorization for creation of the registry is not applicable under the Privacy Rule insofar as the registry constitutes healthcare operations because the primary purpose is not to obtain generalizable knowledge. See applicable regulatory criteria for waiver under HIPAA for secondary uses that constitute research. See applicable regulatory criteria for waiver under Common Rule and FDA requirements.

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
	documentation or waiver for the consent process.	operations pathway, depending on the purpose. The Common Rule may apply as one purpose includes research.** In addition, determine if a "human subject" is involved. FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	and documented consent are required unless an IRB grants a waiver of documentation or waiver for the consent process. FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	
5. Quality A/I registry not involving any research.	The Privacy Rule would not apply to the disclosure to or use by the registry of the de-identified information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. The Common Rule is not applicable. FDA GCP is likely not applicable as the activity is not research.	The Privacy Rule permits the use or disclosure of a limited dataset for healthcare operations, provided the data source and registry developer enter into a data use agreement and downstream users of the registry enter into a DUA with the registry developer. The Common Rule is not applicable. FDA GCP is likely not applicable as the activity is not research.	The Privacy Rule permits use or disclosure for the "healthcare operations" of the data source and in certain circumstances, of another covered entity. The Common Rule is not applicable. FDA GCP is likely not applicable as the activity is not research.	Waivers are not applicable as the activity is not research.

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
6. Research registry developed by organization that is not a healthcare provider or insurance plan and is not subject to the Common Rule, using health information obtained from a healthcare provider or insurance plan or a business associate thereof.	The Privacy Rule would not apply to the disclosure to the registry of the de-identified information. A business associate should ensure it has the right to create the de-identified datasets and share the data for such purpose. FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process. Note potential limitations on uses and disclosures under the FTC Act and state law.	The Privacy Rule permits the disclosure of a limited dataset, provided the data source and registry developer enter into a DUA, and downstream users of the registry enter into a DUA with the registry developer. FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process. Note potential limitations on uses and disclosures under the FTC Act and state law.	The Privacy Rule permits disclosure for research with individual patient authorization or waiver of authorization. FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process. Note potential limitations on uses and disclosures under the FTC Act and state law.	See applicable regulatory criteria for waiver under Privacy Rule and FDA requirements.
7. Research registry developed by organization that is not a healthcare provider or insurance plan and is not subject to the Common Rule, using health information collected from entities not subject to the Privacy Rule (e.g., direct-to-consumer mobile medical app or survey of patients to collect patient-reported outcomes).	FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	FDA GCP may apply to registry development or use depending on research purpose. If so, IRB review and documented consent would be required unless an IRB grants a waiver of documentation or waiver for the consent process.	See applicable regulatory criteria for waiver under FDA requirements.

Registry Developer or Purpose of Registry	Health Information Is De-identified*	Health Information Excludes Direct Identifiers	Health Information Includes Direct Identifiers	Waiver of Authorization, Documentation of Consent, or Consent Process
	Note potential limitations on uses and disclosures under the FTC Act and state law.	Note potential limitations on uses and disclosures under the FTC Act and state law.	Note potential limitations on uses and disclosures under the FTC Act and state law.	

DUA = data use agreement; FDA = U.S. Food and Drug Administration; IRB = institutional review board; GCP = Good Clinical Practice requirements; FTC Act = Federal Trade Commission Act.

Note: Reference to this table is not a substitute for consultation with appropriate institutional officials about the regulatory requirements that may apply to a particular registry project.

*Information lacks the data elements specified in the Privacy Rule standard for de-identification.

**The Common Rule would apply to the development and maintenance or the use of a research registry if the registry or use thereof is funded or supported by a Common Rule Agency. In addition, even where the Common Rule is not applicable by operation of law, institutional policy may apply the Common Rule, and the Common Rule may be incorporated into certain state laws. In addition, note that IRBs may differ as to whether a limited dataset under HIPAA constitutes a "human subject" under the Common Rule, although the Secretary's Advisory Committee for the Protection of Human Subjects has recommended that OHRP issue guidance clarifying that a limited dataset is not a human subject. See Secretary's Advisory Committee on Human Research Protections, Attachment C – Updated FAQs on Informed Consent for Use of Biospecimens and Data (April 11, 2018), available at https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-faqs-recommendations-and-glossary-informed-consent-and-research-use-of-biospecimens-and-associated-data/index.html.

As healthcare and, more broadly, consumer life become increasingly digitized, the extent and variety of information that can be leveraged for registries continue to expand. In the context of this chapter, health information is broadly construed to include any (i) information created or used by or on behalf of healthcare providers and insurance plans that relates to an individual's health condition, the provision of healthcare services to an individual, or payment for healthcare services provided to an individual, as well as (ii) health, wellness, and other lifestyle-related information collected through devices, mobile applications, and other interfaces or initiatives that engage directly with individuals as consumers and are not provided or initiated on behalf of a healthcare provider.[1] This definition is designed to reflect the growth of the "patient-as-consumer" construct. As a result, health information may include a broad range of information relating to the provision and payment of healthcare, such as medical history, prescription history, provider notes, test results and reports, genomic sequencing data, demographic information, and claims data, as well as self-reported data, metrics, and other information collected from wearables, mobile medical apps, and other platforms that may include information on mental health, lifestyle habits, medication adherence, socioeconomic status, the environment, and other factors that may affect health status or health risks. Certain types of genomic information[2] and other health information includes information about family members, so it also can have an impact on the privacy of third parties. Individuals widely regard health information as private and thus expect confidentiality to be maintained, although such expectations may vary depending on the nature of the information and the context (such as the involvement of commercial entities versus academic institutions).

Concerns about potential risks to individual privacy have led to federal legal requirements for prospective review of research projects and conditions on the use or disclosure of health information for research and other purposes. The creation and use of patient registries for a research purpose ordinarily constitute "research involving human subjects" as defined by regulations applicable to research activities funded by the U.S. Department of Health and Human Services[3] (HHS) and certain other Federal agencies.[4] Moreover, federal privacy regulations promulgated under the Health Insurance Portability and Accountability Act of 1996 (HIPAA)[5] and modified by the Health Information Technology for Economic and Clinical Health Act (HITECH – part of the American Recovery and Reinvestment Act of 2009) specifically apply to the use and disclosure of certain individually identifiable health information, known as protected health information (PHI) under the HIPAA Rules, for research and other purposes. The Federal Food, Drug, and Cosmetic Act (FDCA) and implementing regulations of the U.S. Food and Drug Administration (FDA) also include requirements for the protection of human subjects in connection with clinical investigations, which may apply to certain activities involving the collection and use of health information in registries that fall under the FDA's jurisdiction.

The term *human subjects* is used throughout this chapter for consistency with applicable Federal law. Some may prefer the term *research participants*.

This chapter provides a general guide to Federal legal requirements in the United States. (Legal requirements in other countries may also be relevant and may be different from those in this country, but a discussion of any applicable international rules is beyond the scope of this document.) These legal requirements may influence registry decisions involving the selection of data elements and data verification procedures, and may also affect subsequent uses of registry data for secondary research purposes. State laws also may apply to the use of health information

for research purposes. The purpose of a registry, the status of its developer, and the nature and source of the registry data, including the extent to which such data are identifiable, largely determine applicable regulatory requirements. This chapter reviews the most common of these arrangements. The complexity and sophistication of registry structures and operations vary widely, with considerable variability also observed in the processes registry stewards use to obtain data. Nonetheless, common ethical and legal principles are associated with the creation and use of registries. These commonalities are the focus of this chapter.

Ethical concerns about the conduct of biomedical research, especially research involving the interaction of the clinical research community with its patients and commercial funding agencies, have produced an impetus to make financial and other arrangements more public. The discussion of transparency in this chapter includes recommendations for the public disclosure of registry operations as a means of maintaining public trust and confidence in the use of health information for registry purposes, particularly as questions and concerns about privacy intensify as a result of widely-reported cybersecurity breaches and reports of alleged violations of privacy by well-known companies and other entities. Reliance on a standing advisory committee is recommended to registry developers as a way to provide expert technical guidance for registry operations and to firmly establish the independence of the registry from committed or conflicted interests, as described in Chapter 9. This discussion of transparency in methods is not intended to discourage private investments in registries that produce proprietary information in some circumstances. Neither the funding source nor the generation of proprietary information from a registry determines whether a registry exercises and adheres to the good practices described in this guide.

Healthcare providers and health insurance plans have plausible claims to the exclusive use of health and claims information, although the public perspective on these claims has not been tested. Registry developers should anticipate negotiating access to health and claims information, especially when it is maintained in electronic form. Registry developers also are likely to encounter licensing requirements, including processing and use fees, in obtaining health and claims information. The processes for use of registry datasets, especially in multiple analyses by different investigators, should be publicly disclosed to assure the public that registries are appropriately protecting the confidentiality of health information.

2. Ethical Concerns Relating to Health Information Registries

2.1 Application of Ethical Principles

The Belmont Report[6] is a summary of the basic principles and guidelines developed to assist in resolving ethical problems in conducting research using human subjects. It was the work product of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, which was created by the National Research Act of 1974.[7]

The Belmont Report identifies three fundamental principles for the ethical conduct of scientific research that involves human subjects. These principles are respect for persons as autonomous agents (self-determination), beneficence (do good, do no harm, protect from harm), and justice (fairness, equitable distribution of benefits and burdens, equal treatment). Together, they provide a foundation for the ethical analysis of human subjects research, including the use of health

information in registries developed for scientific purposes with a prospect of producing social benefits. These principles are substantively the same as those identified by the Council for International Organizations of Medical Sciences in its international guidelines for the ethical review of epidemiologic studies.[8]

Nevertheless, the application of these principles to specific research activities can result in different conclusions about what comprises ethical design and conduct of the research in question. These different conclusions frequently occur because the principles are assigned different values and relative importance when more than one person performs the ethical analysis. In most of these situations, however, a generally supported consensus position on the ethical design and conduct of the research is a desired and achievable goal. This goal does not preclude re-analysis as social norms or concerns about research activities change over time in response to new information, new technologies or persistent ethical questioning.

The ethical principle of *respect for persons* supports the practice of obtaining individuals' consent to the use of their health information for research purposes related or unrelated to the clinical and insurance reasons for creating the information. In connection with research registries, consent may have multiple components: (1) consent to registry creation by the compilation of patient information; (2) consent to the initial research purpose and uses of registry data; and (3) consent to subsequent use of registry data by the registry developer or others for the same or different research purposes. The consent process should adequately describe registry purposes and operations to inform potential subjects' decisions about participation in a research registry. In some defined circumstances, the principle of *respect for persons* may be subordinate to other ethical principles and values, with the result that an explicit consent process for participation in the registry may not be necessary. A waiver of informed consent requirements may apply to the registry and be ethically acceptable. (See discussion of waivers of informed consent and authorization requirements below.) In these situations, alternatives to an explicit consent process for each individual contributing health information to the registry may be adequate. For example, the registry might provide readily accessible, publicly available information about its activities as an alternative to individual informed consent, or use an opt-out approach for collecting health information through the registry.

A general ethical requirement for consent clearly implies that human subjects voluntarily permit the use of their health information in a registry, unless a specific exception to voluntary participation applies to the registry. One such exception is a legally mandated, public health justification for the compilation of health information (e.g., certain infectious disease reporting). Voluntary agreement to the use of health information in a registry necessarily allows a subsequent decision to discontinue participation. Any limitation on an individual's ability to withdraw information from the registry (e.g., once incorporation into aggregated data has occurred) should be clearly communicated in the consent process as a condition of initial participation. The consent process should also include instructions about the procedures for withdrawal at any time from participation in the registry unless a waiver of consent applies to the registry. Incentives for registry use of health information (e.g., insurance coverage of payments for healthcare services) should be carefully evaluated for undue influence both on the individuals whose health information is sought for registry projects and on the healthcare providers of those services.[9,10]

Conflicts of interest also may result in undue influence on patients and may compromise voluntary participation. One potential source of conflict widely identified within clinical research is the use of recruitment incentives paid by funding agencies to healthcare providers.[11] Some professional societies and research organizations have established policy on the use of recruitment incentives. Many entities have characterized as unethical incentives that are significantly beyond fair market value for the work performed by the healthcare provider; others require disclosure to research subjects of any conflicting interest, financial or nonfinancial.[12] Federal law now requires manufacturers of certain drugs, devices, or medical supplies to report, for public display, the amounts of remuneration paid to physicians for research purposes.[13] Some States, including Massachusetts, have similar laws in effect.[14] Research organizations, particularly grantees of Federal research funding, may have systematic policies and procedures in place that registry developers can rely on for managing employee conflicts of interest. Nonetheless, in their planning, registry developers should specify and implement recruitment practices that protect patients against inappropriate influences.

Applying the principle of *respect for persons* to the research use of health information generates additional ethical concerns about preserving the privacy and dignity of patients, protecting the confidentiality of health information, and minimizing potential harms. These concerns have intensified as healthcare services and third-party payment systems have become more complex and as technology continues to transform healthcare and contribute to the proliferation of data and the ease with which such data can be collected and shared. Legal standards for the use and disclosure of health information create a baseline of required privacy protections for individually identifiable health information. However, depending on the particular health condition, population of interest, or nature of the health information, safeguards for the confidentiality of registry data beyond applicable legal requirements may be ethically necessary or appropriate to protect the privacy and dignity of those individuals contributing health information to the registry. For example, certain institutions may determine that it would be prudent to use an informed consent model to collect genetic sequencing data that, while considered de-identified under current standards under HIPAA and the Common Rule, are obtained from members of a community that tends to be more disenfranchised or that experienced historical ethical transgressions in connection with human subjects research.

The principle of *beneficence* ethically obligates developers of health information registries for research purposes to minimize potential harms to the individuals or groups[15] whose health information is included in the registry. There are usually no apparent benefits to offset potential harm to the individuals or groups whose health information is used in the registry. Exceptions to this arise when a registry is designed to provide benefits to the human subjects as individuals, such as longitudinal reports on treatment effects or health status or quality-of-care reports. Risks to privacy and dignity are minimized by conscientious protection of the confidentiality of the health information included in the registry[16] through the use of appropriate physical, technical, and administrative safeguards for data in the operations of the registry. These safeguards should include controls on access to registry data, including access to individual identifiers that may be included in registry data. Minimization of risks also requires a precise determination of what information is necessary for the research purposes of the registry and limiting the information collected accordingly. Further, in considering the principle of beneficence, developers of health information registries should assess whether a proposed registry promotes the efficient use of

resources, including whether individuals will be asked to contribute data that is duplicative of existing data sources.

Certain populations of patients may be vulnerable to social, economic, or psychological harms as a result of a stigmatizing health condition. This concern will likely become more pronounced as genetic testing and sequencing continues to play a bigger role in medicine and society, including in the direct-to-consumer context, and contributes to the proliferation of genomic data that can be used for registries. There has been much debate surrounding the notion of genetic exceptionalism (i.e., the proposition that genetic or genomic information should be treated and protected differently from other types of health information) and different institutions, IRBs, patient populations, and others will differ as to their position on the issue.[17] Developers of registries compiling this health information must consider these challenging issues and determine whether additional efforts to protect the identities of the human subjects contributing data to the registry or other ethical safeguards are necessary or appropriate given the particular patient population and related contextual considerations. Additional protections also apply to populations such as pregnant women, human fetuses, neonates, prisoners, and children, who are considered vulnerable to undue influence and coercion during the consent process. In particular, data obtained from pediatric and adolescent populations may lead to ethical concerns if there is the potential for lifelong discrimination that may effectively exclude them from educational opportunities and other social benefits[18] (e.g., health insurance, although under the Affordable Care Act health insurers may not discriminate against individuals on the basis of pre-existing conditions).

In an analysis applying the principle of *beneficence*, research involving human subjects that is unlikely to produce valid scientific information is unethical. This conclusion is based on the lack of social benefit to offset even minimal risks imposed by the research on participating individuals. Health information registries should incorporate an appropriate design (including, where appropriate, calculation of the patient sample as described in Chapter 3) and data elements, written operating procedures, and documented methodologies, as necessary, to ensure the fulfillment of a valid scientific purpose.[19]

An ethical analysis employing the principle of *justice* also yields candid recognition of the potential risks to those who contribute health information to a registry, and the probable lack of benefit to those individuals (except in the cases where registries are specifically constructed to provide benefit to those individuals). The imbalance of burden and benefit to individuals reinforces the need to minimize the risks from registry use of health information. Precise and well-developed scientific reasons for inclusion (or exclusion) of defined health information in a registry help ensure that the burden placed on individuals as a result of their participation is fair and equitable.

The above analysis refers to research activities. However, the ethical concerns expressed may also apply to other activities involving the use or disclosure of individuals' health information for nonresearch purposes. Public health, oversight of the delivery of healthcare services through government programs, and quality A/I activities all can evoke the same set of ethical concerns as research activities about the protection of patient self-determination, privacy, and dignity; the maintenance of the confidentiality of individually identifiable health information to avoid potential harms; and the imposition of a risk of harm on some individuals to the benefit of others

not at risk. In the past, different assignments of social value to these activities and different potential for the social benefits and harms they produce have created different levels of social acceptance and formal oversight for these activities compared with research activities. Nonetheless, these activities may include a research component in addition to their primary stated objectives, a circumstance that implicates the ethical concerns discussed above and produces additional concerns about compliance with the legal requirements for research activities. In addition, in an era of "big data," registries may be leveraged for multiple purposes and intersecting activities that may make it challenging to properly categorize the nature of the registry and any distinct use, as well as the legal and ethical standards that should apply. Registry developers should prospectively apply careful scrutiny to the proposed purposes for and activities of a registry, in consultation with appropriate institutional officials, to avoid both ethical and compliance issues that may undermine achievement of the registry's objectives.

Registry developers also must consider confidentiality and/or proprietary concerns with regard to the identity of the healthcare providers, at the level of both individual professionals and institutions, and the healthcare insurance plans from which they obtain registry data. Information about healthcare providers and insurance plans can also identify certain patient populations and, in rare circumstances, individual patients. Moreover, the objectives of any registry, broadly speaking, are to enhance the value of the healthcare services received, not to undermine the credibility and thus the effectiveness of healthcare providers and insurance plans in their communities. Developers of registries created for public health investigations, health system oversight activities, and quality A/I initiatives to monitor compliance with recognized clinical standards must consider whether safeguards for the identity of service professionals and institutions are appropriate. At the same time, however, any confidentiality safeguards should permit certain disclosures, as permitted by applicable law and designated by the service professionals and institutions, for the reporting of performance data, which are increasingly associated with payment from payers.

2.2 Transformation of Ethical Concerns Into Legal Requirements

Important ethical concerns about the creation, maintenance, and use of patient registries for research purposes include risks of harm to human subjects resulting from unauthorized access to registry data and inappropriate use of the compiled health information. These concerns about harms arise from public expectations of confidentiality for health information and the importance of that confidentiality in preserving the privacy and dignity of individual patients as well as the clinician/patient relationship.

Over the last decade, two rapid technological developments have intensified these ethical concerns. One of these advances was DNA sequencing, replication, recombination, and the concomitant application of this technology to facilitate the delivery of and research into precision medicine. Despite the potential that genomics holds in the quest to find cures and more effective treatments, its proliferation has also raised new ethical questions and prompted debate about whether it creates unique considerations and risks from an individual and group privacy standpoint.

Another contributing factor is the rapid digitization of healthcare and daily consumer life, as exemplified by the advance of health information systems technology, electronic information processing, and development of connected devices and platforms that enable the rapid and extensive collection, generation, and sharing of electronic information that can be leveraged for research and other purposes. In some circumstances, such information is being gathered by entities that are not regulated under the traditional Federal healthcare privacy framework because they operate under a direct-to-consumer paradigm and not on behalf of a HIPAA covered entity.[20] The emergence of these stakeholders, and the growing ease with which information can be harnessed in the digital age, raise new questions surrounding the protections that should be incorporated into health information registries. The discussion below about legal protections for the privacy of health information focuses solely on U.S. law.

2.2.1 The Common Rule

International and domestic concerns about the protection, respect, and privacy of human subjects resulted in a uniform set of regulations from the Federal agencies that fund such research known as the "Common Rule."[21,22] The legal requirements of the Common Rule apply to research involving human subjects conducted or supported by the 20 Federal departments and agencies, including HHS, that intend to follow the revised Common Rule (Common Rule Agencies). Some of these agencies may require additional legal protections for human subjects. In addition, under the revised Common Rule, institutions that receive funding from a Common Rule Agency may also voluntarily elect through their Federalwide Assurance (as discussed further below) to apply the Common Rule to all human subjects research, irrespective of funding source, conducted by their employees or agents.[23]

Significant amendments were made to the Common Rule in 2017. As of January 21, 2019, institutions are expected to be compliant with the revised Common Rule.

Each institution engaged in human subjects research conducted or supported by a Common Rule Agency must enter into a formal written agreement to comply with the Common Rule. For human subjects research conducted or supported by most of the Common Rule Agencies, the required agreement is called a Federalwide Assurance (FWA).[24] The Office for Human Research Protections (OHRP) administers the Common Rule as it applies to human subjects research conducted or supported by HHS. The application of Common Rule requirements to a particular registry depends on the institutional context of the registry developer, relevant institutional policies, and whether the health information contributed to the registry maintains patient identifiers. Of particular note is that while the pre-2018 Common Rule allowed institutions to voluntarily subject all of their human subjects research, irrespective of the source of funding or support, to oversight by OHRP for compliance with the Common Rule, the revised Common Rule eliminated this option. Therefore, institutions that seek to conduct non-federally funded or supported human subjects research in accordance with the Common Rule will need to rely on internal oversight and compliance mechanisms to facilitate adherence to Common Rule standards.

Guidance documents published by OHRP, such as the 2008 guidance entitled "Coded Private Information or Specimens Use in Research" and the guidance entitled "Issues to Consider in the

Research Use of Stored Data or Tissues" (last updated in 1997), discuss the research use of identifiable private health information.[25] The latter guidance makes clear that OHRP considers the creation of health information registries—containing individually identifiable, private information—for research purposes to be human subjects research for the institutions subject to its jurisdiction.[26] The applicability of the Common Rule to research registries is discussed in more detail in Section 3.

OHRP regulations for human subject protection require prospective review and approval of human subjects research by an institutional review board (IRB) and the informed consent (usually written) of each of the human subjects involved in the research, unless an IRB expressly grants a waiver of informed consent requirements.[27] A research project must satisfy certain regulatory conditions to obtain IRB approval of a waiver of the informed consent requirements. (See below for discussion of waivers of informed consent requirements.) A registry plan is the research "protocol" reviewed by the IRB. At a minimum, the protocol should identify (1) the research purpose of a health information registry, (2) detailed arrangements for obtaining informed consent, or detailed justifications for not obtaining informed consent, to collect health information, and (3) appropriate safeguards for protecting the confidentiality of registry data, in addition to any other information required by the IRB on the risks and benefits of the research.[28]

As noted previously, for human subjects research conducted or supported by most Common Rule Agencies, an FWA satisfies the requirement for an approved assurance of compliance. In addition, irrespective of requirements applicable by operation of the revised Common Rule, some research organizations have explicit institutional policies and procedures that require IRB review and approval of all human subjects research.

2.2.2 The Privacy Rule

In the United States, HIPAA and its implementing regulations (namely, the Privacy, Security, and Breach Notification Rules, which are collectively referred to here as the HIPAA Rules) created legal protections for the privacy of individually identifiable health information created and maintained by "covered entities" and their "business associates." "Individually identifiable health information" is information, including demographic data, created or received by a healthcare provider, health plan, employer, or healthcare clearinghouse, that identifies an individual or could reasonably be used to identify an individual, and relates to (1) an individual's past, present, or future physical or mental health or condition; (2) the provision of healthcare to an individual; or (3) the past, present, or future payment for healthcare to an individual. With certain exceptions, "individually identifiable health information" is "protected health information" (PHI) under the HIPAA Rules when it is transmitted or maintained by a covered entity or a business associate on behalf of a covered entity.[29] Because registries may exist over long periods of time, it is important to note that the individually identifiable information of persons who have been deceased for more than 50 years is not considered PHI.

Covered entities are healthcare providers that engage in certain standard financial or administrative healthcare transactions electronically, health plans, and healthcare clearinghouses.[30] Business associates generally are persons or organizations, other than a member of a covered entity's workforce, that perform certain functions or services (e.g., claims

processing, data analysis, data aggregation, patient safety activities) on the covered entity's behalf that involve access to PHI.[31] Covered entities and business associates are subject to civil—and in some cases criminal—liability for violations of the HIPAA Rules. This chapter will focus on covered entity healthcare providers and healthcare plans, as well as their respective business associates.

Generally, the Privacy Rule defines the circumstances under which covered entities and their business associates may use and disclose PHI for a variety of purposes, including research.[32] The Privacy Rule establishes a Federal baseline of protections, and it does not preempt State laws that provide even greater, more stringent privacy protections for PHI.[33] For example, the Privacy Rule requires covered entities to include certain information in patient authorizations for the use or disclosure of PHI, including an expiration date or event that can be many years in the future. The laws of the State of Maryland, however, specifically require that, absent certain exceptions, a patient's authorization may only be valid for a maximum period of 1 year.[34] In this case, a covered entity located in Maryland can and should satisfy both the Privacy Rule and State law requirements by complying with the State's one-year maximum expiration deadline on its patient authorization forms.

The HIPAA Rules may apply to either or both the registry developer (as a covered entity or a business associate developing the registry in a business associate capacity) and the registry's data sources. A registry's initial collection of health information from a covered entity or business associate requires a disclosure pathway under the Privacy Rule. Thus, registry developers that are not themselves subject to the HIPAA Rules should nonetheless be knowledgeable about the HIPAA Rules to facilitate the necessary processes for any of their data sources that are covered entities or business associates. In developing a registry, they should expect to interact with clinicians, the privacy officer, the IRB or privacy board staff, health information system representatives, legal counsel, compliance officials, and contracting personnel. Registry developers should also maintain awareness of regulatory modifications or amendments to, or new guidance on how to comply with, the HIPAA Rules, which can be expected as the use of electronic PHI becomes more prevalent. For example, on January 25, 2013, HHS issued significant modifications to the HIPAA Rules, many of which implemented HITECH Act requirements.[35] One of the most relevant modifications for registry developers, as mentioned above and discussed more fully below, is the extension of certain requirements of the HIPAA Rules and liability for noncompliance directly to business associates.[36]

The HIPAA Rules would also apply where the registry developer is a covered entity or business associate (creating the registry in its business associate capacity) and collects health information from other covered entities or business associates, or from data sources that are not subject to the HIPAA Rules. Examples of the latter may include developers of wearable devices and mobile applications that are provided directly, and not on behalf of a covered entity or health plan, to individuals to track their health and fitness. Although HIPAA does not apply to such data sources, a pathway under the Privacy Rule would still be necessary for the HIPAA covered registry developer to use and disclose the data once in its possession, as the data would then constitute PHI in its possession. Note also that the data source itself may be subject to other laws, such as the Federal Trade Commission Act's (FTC Act), which prohibits unfair or deceptive trade practices, and State laws that include requirements for the protection of personal information from a privacy or consumer protection standpoint. Registry developers should

anticipate that such non-HIPAA regulated data sources may become more prevalent with the proliferation of digital health solutions and consider the regulatory and legal requirements and other limitations (such as in privacy policies and terms of use applicable to the data collection platform) that may apply to the registry's collection and use of data from such sources.

Under certain circumstances, registry developers and the associated institutions where the registry will reside may not be subject to the HIPAA Rules. Notably, the HIPAA Rules do not apply to registries that reside outside of a covered entity or business associate. Within academic medical centers, for example, registry developers may be associated with units that are outside of the institutional healthcare component to which the HIPAA Rules apply, such as a biostatistics or economics department. The FTC Act and similar state laws referenced above may apply to such registry developers if they are not a public agency or non-profit institution.[37] To avoid running afoul of the FTC Act, such registry developers should be transparent in privacy notices provided to participants through any website or application used for information collection process about the intended uses and disclosures of the information. The registry developer should also limit uses and disclosures of information collected through the registry to those described in the privacy notice and maintain safeguards commensurate with the representations made in the privacy notice and the sensitivity of the information collected by the registry.

Ultimately, however, many potential data sources for registries will be covered entities or business associates, such that registry developers are likely to find themselves deeply enmeshed in the HIPAA Rules. As noted above, a registry may have direct liability under HIPAA if the registry is considered a business associate of a data source that is a covered entity (see the discussion below of the HITECH Act, which extended direct liability for compliance with certain requirements of the HIPAA Rules to business associates of covered entities, where before business associates were required and liable to protect the information to which they had access only through their business associate agreements with covered entities). Under such circumstances, the registry developer must enter into a business associate agreement with the covered entity that meets the requirements under the Privacy Rule before the registry developer can use or disclose PHI in connection with the development or deployment of the registry. Therefore, registry developers should be cognizant of the patient privacy considerations confronting their likely data sources—as well as themselves, if they are performing functions or services on behalf of their data sources as business associates—and should consider implementing certain HIPAA protections whether or not they are required to do so. In addition, the HIPAA Rules require that covered entities enter into formal agreements, known as data use agreements, with any recipient of PHI that constitutes a limited dataset before the recipient may use the limited dataset for permitted purposes (i.e., research, healthcare operations, or public health activities). Recipients of limited datasets may be subject to legally enforceable obligations under contract law by virtue of the data use agreement in addition to any regulatory obligations that apply in the event the recipient is a covered entity or business associate.

A registry developer that is not a covered entity or business associate and that seeks to collect information directly from individuals may also still encounter Privacy Rule requirements if patient information from a healthcare provider or insurance plan for purposes of is needed to recruit registry participants. For example, a patient authorization or waiver of authorization (discussed below) may be necessary for the disclosure of patient contact information by a healthcare provider or insurance plan (or their business associate) to a registry developer, even if

the actual information to be collected by the registry will be provided from the patient him or herself. Note that the strategy of requesting data directly from individuals can be useful for collecting data on mobile populations, such as elderly retirees who occupy different residences in winter and summer, and for collecting the health records of school children. A Federal privacy law[38] protects the health records of children that are held by schools from disclosure without explicit parental consent; thus, parents can often obtain copies of these records more easily than investigators.

Following the registry's collection of data, its subsequent use and sharing of registry data will be informed by the regulatory conditions that applied to the initial collection of the registry data, as well as by other ethical and legal considerations. The Privacy Rule created multiple pathways by which registries can compile and use patient information. For instance, a registry within a covered entity may obtain a HIPAA authorization from each patient contributing PHI to a registry for a particular research project, such as the relationship between hypertension and Alzheimer's disease. If the registry subsequently seeks to use the PHI for a different research purpose, it may do so if it obtains new authorizations or the use otherwise satisfies the Privacy Rule. For example, the registry may de-identify the PHI in accordance with the Privacy Rule's de-identification standards, at which point the data would no longer be considered as PHI. Alternatively, the registry can obtain authorizations for use and disclosure of individuals' PHI for future research purposes at the same time that it obtains authorization to place the information in the registry, as long as the authorization adequately describes the purposes of the future research such that it would be reasonable for the individual to expect that his or her information could be used or disclosed in connection with the future research activity.[39]

The authors recommend that registry developers establish a detailed tracking system, based on the extent to which registry data remain identifiable for individual patients, for the collection, uses, and disclosures of registry data. The tracking system should produce comprehensive documentation of compliance with both Privacy Rule requirements, including requirements to obtain authorizations, and any legally binding contractual obligations to data sources.

With regard to registries developed for research purposes, the Privacy Rule defines research as "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge."[40] Commentary by HHS on the Privacy Rule explicitly includes within this definition of research the development (building and maintenance) of a repository or database for future research purposes.[41] The definition of research in the Privacy Rule partially restates the definition of research in the Common Rule for the protection of human subjects, adopted by HHS and certain other Federal agencies.[42] Some implications of this partial restatement of the definition of research are discussed later in this chapter.

The National Institutes of Health (NIH) has published guidance, in collaboration with the Office for Civil Rights and other HHS offices and agencies, on the impact of the Privacy Rule on health services research and research databases and repositories. The NIH guidance identifies the options available to investigators under the Privacy Rule to gain access to PHI held by healthcare providers and insurance plans.[43] For example, in addition to provisions for the use or disclosure of identifiable patient information for research, the Privacy Rule permits healthcare providers and insurance plans (and business associates on their behalf) to use or disclose patient information for certain defined public health activities.[44] The Privacy Rule defines a public

health authority as "an agency or authority of the United States, a State, a territory, a political subdivision of a State or territory, or an Indian tribe, or a person or entity acting under a grant of authority from or contract with such public agency… that is responsible for public health matters as part of its official mandate." The Centers for Disease Control and Prevention and HHS have jointly published specific guidance on the Privacy Rule requirements related to public health activities.[45] Other Privacy Rule provisions permit uses or disclosures of PHI that are required by law, including State laws.[46]

The protections for patient information created by the Privacy Rule that are generally relevant to registries developed for research purposes include explicit individual patient authorization for the use or disclosure of PHI,[47] legally binding data use agreements for the release of "limited datasets" between health information sources and users,[48] the removal of specified identifiers or statistical certification to achieve de-identification of health information,[49] an accounting of disclosures to be made available to patients at their request,[50] and notification in the event of a breach of unsecured PHI to affected individuals who may be affected by the breach, as well as HHS and, in some cases, the media. In addition, if certain criteria required by the Privacy Rule are satisfied, an IRB or privacy board may grant a waiver of individual patient authorization for the use or disclosure of health information in research.[51]

2.2.3 FDA Regulations

Depending on the circumstances, FDA regulatory requirements may apply to the development or use of a health information registry. In particular, FDA requirements for the protection of human subjects (also referred to as "Good Clinical Practice" or "GCP" requirements) apply to activities that constitute a "clinical investigation," which generally means any experiment that involves an FDA-regulated test article and one or more human subjects, and that either is subject to FDA requirements for prior submission, or the results of which are intended to be submitted later to, or held for inspection by, the FDA as part of an application for a research or marketing permit.[52] Where a health information registry is developed or used to collect data for supporting a marketing permit for a new product, a labeling update for a currently marketed product, or other submission to the FDA relating to an FDA-regulated drug, device, or biologic, FDA human subject protection regulations may apply.

Similar to the Common Rule, FDA human subject protection regulations include requirements for obtaining IRB review of proposed clinical investigations and the informed consent (or an IRB waiver thereof[53]) of participating human subjects.[54] Informed consent must be documented by a signed written consent form unless an IRB waives the documentation requirement.[55] Also of relevance to the establishment and use of such health information registries are the FDA regulatory requirements for the use of electronic records and electronic signatures.[56] Commonly referred to as "Part 11," these requirements include standards for access controls, audit trails, and other safeguards to protect the integrity and validity of electronic records and electronic signatures that are used to satisfy FDA statutory or regulatory records requirements.[57]

Unlike the Common Rule and HIPAA, FDA human subject protection regulations do not explicitly exclude clinical investigations that are limited to the use of de-identified information.[58] Rather, FDA regulations define "human subject" as any "individual who is or becomes a

participant in research, either as a recipient of the test article or as a control. A subject may be either a healthy human or a participant."[59] This paradigm is likely a reflection of the nature of much research involving FDA-regulated investigational products, which often involves interventions or interactions with living human subjects rather than constitute research limited to the retrospective analysis of data. Until July 2017, the challenge in the lack of a carve-out for research using de-identified data had been compounded by the inability of IRBs to waive FDA's informed consent requirements for certain minimal risk research (as is permitted under the Common Rule and HIPAA).[60] As real world data sources proliferate and technological advances make it increasingly easier to capture, share, and learn from electronic information, however, stakeholders are recognizing the potential that data holds in research and development, post-market surveillance, and other FDA-regulated activities.

In that regard, and driven in part by changes mandated under the 21st Century Cures Act,[61] the FDA has taken steps to facilitate and provide more regulatory guidance regarding the secondary use of data. First, investigators may now seek IRB waiver of the informed consent requirement under FDA regulations according to criteria that are comparable to those under the Common Rule. The parameters for seeking IRB waiver of FDA informed consent requirements are discussed in more detail in Section 3.3.5. Additionally, developers of health information registries should note that the FDA has issued final guidance on the use of real world evidence to support FDA regulatory decision making for medical devices (RWE Guidance).[62] The RWE Guidance includes the FDA's recommendations for when a proposed collection of real world data that constitutes a clinical investigation would require an investigational device exemption (IDE). The FDA states that such a determination is fact- and context-specific, but that generally, if the collection of real world data involves using a device in the normal course of medical practice or routine clinical care under the authority of a healthcare practitioner, an IDE would likely not be required.[63] In contrast, if the goal is to generate data on the safety and efficacy of a device and the process influences treatment decisions, an IDE may be required.[64] The RWE Guidance also describes the characteristics of real world data that the FDA may assess to consider whether the data is suitable for regulatory decision making, including the relevance and reliability of the data.[65]

2.2.4 Applicability of Regulations to Research; Multiple-Purpose Registries

At many institutions, the IRB or the office that provides administrative support for the IRB interprets the regulations to determine which activities at that institution constitute human subjects research, and thus may itself determine what activities require IRB review. A registry developer is strongly encouraged to consult his or her organization's IRB or a central IRB, as applicable, early in the registry planning process to avoid delays and lessen the need for multiple revisions of documentation submitted to the IRB. Distinctions between research and other activities that apply scientific methodologies are frequently unclear. Such other activities include both public health practice[66] and quality-related investigations.[67] Both the primary and secondary purposes of an activity are factors considered in the determination of whether registry activities constitute research. For purposes of the Common Rule, as interpreted by OHRP, an activity is considered research even if research is only a secondary purpose of the activity.[68] This OHRP interpretation of research purpose differs from that of the Privacy Rule with respect to quality-related studies performed by healthcare providers and insurance plans. Under the Privacy Rule,

only if the primary purpose of a quality-related activity is to obtain generalizable knowledge do the research provisions of the Privacy Rule apply; otherwise, the Privacy Rule defines the activity as a "healthcare operation."[69]

Additionally, registry developers should be mindful of the distinctions between the scope of FDA GCP requirements, on the one hand, and the Common Rule and HIPAA, on the other hand. With the steps the FDA has taken to facilitate secondary use of data as described previously, the scope of potential FDA-regulated use cases for health information registries may expand. In that regard, while a research registry that is limited to collecting and using de-identified information would generally fall outside the scope of the Privacy Rule and the Common Rule,[70] as noted above, FDA regulations do not explicitly exempt or exclude clinical investigations that use only de-identified data from FDA human subject protection requirements.[71] Thus, compliance with FDA GCP requirements may be required for certain registry activities that are not subject to the Common Rule or HIPAA.

Registry developers should rely on their privacy officer's and IRB's experience and resources in defining research and other activities for their institutions and determining which activities require IRB review as research. In meeting accreditation standards, inpatient facilities typically maintain standing departmental (e.g., pediatrics) or service (e.g., pharmacy or nursing) committees to direct, review, and analyze quality-related activities. Some physician groups also establish and maintain quality-related programs, because good clinical practice includes ongoing evaluation of any substantive changes to the standard of care. These institutional quality committees can provide guidance on the activities that usually fall within their purview. Similarly, public health agencies typically maintain systematic review processes for identifying the activities that fit within their legal authority.

Standard confidentiality protections for registry data include requirements for physical, technical, and administrative safeguards to be incorporated into plans for a registry. In some instances, an IRB may not consider legally required protections for the research use of patient information sufficient to address relevant confidentiality concerns, including the Privacy Rule protections that may be applicable to registries created by or maintained within covered entities, such as healthcare providers and insurance plans, or business associates. Some IRBs and institutions, for example, may take the position that more stringent privacy protections should apply (if not required under applicable federal or state law) to genomic information. The potential bases for this position include the immutable nature of genetic traits, the potential stigma and discrimination that may result if the information were to be disclosed, and the notion that genomic information may be used to identify an individual or his or her genetic relatives, even if the genomic information does not contain any information that is currently considered PHI under HIPAA.[72] Likewise, information about certain conditions (such as alcoholism or HIV-positive status) and certain populations (such as children) may be associated with a greater potential for harm from social stigma and discrimination. Under these circumstances, the IRB can make approval of a registry plan contingent on implementation of additional safeguards that it determines are necessary to minimize the risks to the individuals contributing health information to the registry.

3. Applicable Regulations

This section discusses the specific applicability of the Common Rule[73], the Privacy Rule[74], and FDA human subject protection regulations[75] to the creation and use of health information registries. Registry developers are strongly encouraged to consult with their organization's privacy officer and IRB or privacy board early in the planning process to clarify applicable regulatory requirements and the probable effect of those requirements on registry design and development.

This discussion assumes four general models for health information registries. One model is the creation of a registry containing the contact, demographic, and diagnostic or exposure information of potential research subjects who will be individually notified about projects in which they may be eligible to participate. The notification process permits the registry to shield registry participants from an inordinate number of invitations to participate in research projects, as well as to protect privacy and confidentiality. This model is particularly applicable to patients with unusual conditions, patients who constitute a vulnerable population,[76] or both (e.g., children with a rare condition). A second model is the creation of a registry and the conduct of all subsequent research using registry data by the same group of investigators. No disclosures of registry data will occur and all research activities have the same scientific purpose. This model applies, in general, to quality improvement registries and other quality-related investigations of a clinical procedure or service. Note, however, that some quality improvement registries may involve confidential feedback to providers as well as public reporting of provider performance in a patient de-identified format. These activities may or may not constitute research as defined by the Common Rule. Under the Privacy Rule, these activities may be regulated as the healthcare operations of the covered entity that provides the data to the registry, rather than research, provided the obtaining of generalizable knowledge is not the primary purpose of the activities. A third model is the creation of a registry for an initial, specific purpose by a group of investigators with the express intent to use registry data themselves, as well as to disclose registry data to other investigators for additional related or unrelated scientific purposes. An example of this last model is a registry of health information from patients diagnosed with a condition that has multiple known comorbidities to which registry data can be applied. This third model is most directly applicable to industry-sponsored registries. The American College of Epidemiology encourages the data sharing contemplated in this last registry model.[77] A fourth model, which is a variation of the third, is the creation of a registry to support multiple purposes and endeavors at the outset (such as for research and quality A/I activities), which may be for a specific organization, and to disclose registry data to other investigators and organizations for myriad scientific and other permitted purposes. This model underscores the trend toward "big data" initiatives that are designed to leverage central repositories of standardized, normalized, and curated data for many different purposes.

The extent to which the regulations will apply to each of these registry models will depend on factors such as the registry developer, purpose of the registry, potential for individual patient identification, consent process, and inclusion of genetic information. These factors are discussed further below.

3.1 Public Health, FDA-Regulated Products, Health Oversight

When Federal, State, or municipal public health agencies create registries in the course of public health practice, specific legislation typically authorizes the creation of the registries and regulates data acquisition, maintenance, security, use, and disclosures of registry data for research. Ethical considerations and concerns about maintaining the confidentiality of patient information used by public health authorities are similar to those for research use, but they generally are explicitly balanced against potential social benefits during the legislative process. Nonetheless, if the registry supports human subjects research activities as well as its public health purposes, Common Rule requirements for IRB review may apply to the creation and maintenance of the registry. Further, depending on the nature and structure of the activity, FDA GCP requirements may apply, such that IRB review and compliance with FDA informed consent requirements may be necessary.

Cancer registries performing public health surveillance activities mandated by State law are well-known exceptions to Common Rule regulation. However, secondary uses of public health registry data for research and the creation of registries funded by public health agencies, such as the Centers for Disease Control and Prevention and the Agency for Healthcare Research and Quality, may be subject to the Common Rule as sponsored research activities. The Common Rule's definitions of human subjects research[78] may encompass these activities, which are discussed in the next subsections of this chapter. Not all cancer registries support public health practice alone, even though the registries are the result of governmental programs. For example, the Surveillance Epidemiology and End Results (SEER) program, funded by the National Cancer Institute, operates and maintains a population-based cancer reporting system of multiple registries, including public use datasets with public domain software. SEER program data are used for many research purposes in addition to aiding public health practices. These latter research activities may be subject to the Common Rule.[79]

Disclosures of health information by healthcare providers and insurance plans (and their business associates on their behalf) for certain defined public health activities are expressly permitted by the Privacy Rule without patient authorization.[80] An example of a public health activity is the practice of surveillance, in which the distributions and trends of designated risk factors, injuries, or diseases in populations are monitored and disseminated.[81] Healthcare providers or insurance plans are likely to insist upon documentation of public health authority for legal review before making any disclosures of health information. Registry developers should obtain this documentation from the agency that funds or enters into a contract for the registry, and present it to the healthcare provider or insurance plan well in advance of data collection efforts.

The Privacy Rule characterizes responsibilities related to the quality, safety, or effectiveness of a product or activity regulated by the FDA as public health activities. This public health exception allows uses and disclosures of patient information to a person subject to FDA jurisdiction with respect to FDA-regulated products or activities for which the person has responsibility, such as for adverse event reporting; product tracking; product recalls, repairs, replacement, or look-back; and postmarketing surveillance (e.g., as part of a risk management program that is a condition for approval of an FDA-regulated product).[82] Nonetheless, while the use and disclosure of PHI in connection with such public health activities may not necessarily require a research pathway under HIPAA, it is possible such activities may require compliance with FDA GCP requirements

and an IDE or investigational new drug (IND) application.[83] Determining the applicability of such FDA requirements is a fact-intensive analysis, and the FDA continues to consider the appropriate parameters and requirements and the value proposition for the use of real world data to support regulatory decision making.[84] Registry developers and the users of the registry data, as applicable, should consult with the FDA early in the planning process should they anticipate that the registry may be used to support FDA-regulated activities, such as postmarket surveillance studies.[85]

The Privacy Rule also permits uses and disclosures by healthcare providers and insurance plans (and their business associates on their behalf) for "health oversight activities" authorized by law.[86] These activities include audits and investigations necessary for oversight of the "healthcare system" and other entities subject to government regulatory programs for which health information is relevant to determining compliance with program standards.[87] The collection of patient information, such as occurrences of decubitus ulceration, from nursing homes that are operating under a compliance or corporate integrity agreement with a Federal or State healthcare program, is an example of a health oversight activity.

3.2 Distinguishing Research Activities From Quality Assurance/Improvement Activities

Ascertaining whether the development and use of a health information registry would be regulated as research can be difficult. This is particularly the case as registries seek to maximize the utility of the data once it is collected and rendered usable – a potentially time- and resource-intensive task given the vast amounts of data that is often collected – by making the data available for myriad purposes. Some registries may support multiple activities, all of which are ultimately related in some way to the objective of improving quality in healthcare, but the nuances of each distinct activity may render some activities research (e.g., analyzing data to learn how to improve quality) while others constitute quality A/I (e.g., reviewing data to assess adherence to quality metrics). In creating and making the registry available for use, registry developers should consider independently the applicability of the various federal standards.

Under the Common Rule, research means "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge. Activities that meet this definition constitute research for purposes of the Common Rule, whether or not they are conducted or supported under a program that is considered research for other purposes. For example, some demonstration and service programs may include research activities."[88] OHRP would consider an activity to constitute research if it meets the above definition, irrespective of whether the expressly stated objectives or characterizations of the activity include references to research.[89]

The Privacy Rule's definition of research[40] restates the first sentence of the Common Rule definition set forth above. However, the Privacy Rule distinguishes between research and quality A/I conducted by covered entities or their business associates[90] that meet the definition of "healthcare operations." Under the Privacy Rule, if the primary purpose of a quality-related registry maintained by a covered entity is to support a research activity (i.e., to create generalizable knowledge), Privacy Rule requirements for research apply to the use or disclosure

of the patient information to create the registry and to subsequent research use of registry data. If, however, the primary purpose is other than to create generalizable knowledge, the study is considered a healthcare operation of the covered entity and is not subject to Privacy Rule requirements for research activities, including the requirement to obtain patient authorization or a waiver of authorization from an IRB or privacy board for the uses or disclosures.

As noted earlier, both public health practice and quality A/I can be difficult to distinguish from research activities, and this may become increasingly so as healthcare industry stakeholders harness big data initiatives to support the need for insights and metrics in value-based purchasing contexts.[91] The determination of whether a particular registry should be considered as or include a research activity under the Common Rule or the Privacy Rule depends on a number of different factors, including but not limited to:

- the nature of the organization where the registry will reside;
- the employment duties of the individuals performing the activities associated with the registry;
- the source of funding for the registry;
- the sources of registry data; and
- whether the creation or subsequent use of the registry:
    - is designed to test a theory, hypothesis, or answer a question;
    - entails the collection of data specifically for purposes of the registry or any interventions or interactions not necessary to deliver healthcare or assess the quality of healthcare provided;
    - is intended to evaluate the safety or efficacy of a product, intervention, process, or activity that is not considered standard of care or supported by current medical evidence and literature;
    - involves overriding or directing the treatment decisions of healthcare providers;
    - will inform process or delivery changes immediately, or if changes would be delayed until the end of the activity;
    - .is designed to help patients broadly and in the future, rather than specific patients in the present;
    - requires adherence to a protocol and does not allow for procedures to be adapted to findings during the course of the project;
    - involves the use of a product or a product for an indication not yet approved or cleared by the FDA; or
    - involves evaluating the impact of an activity, process, or other intervention on the behavior of healthcare providers or organizations.

Quality A/I activities entail many of the same ethical concerns about protecting the confidentiality of health information as research activities do. Obtaining express patient consent to participate in quality A/I activities is not the usual practice; instead, the professional and

cultural norms of healthcare providers, both individual and institutional, regulate these activities. Registry developers should consider whether the ethical concerns associated with a proposed quality A/I or patient safety registry require independent review and the use of special procedures such as notice to patients or providers. Registry advisory committee members, quality A/I and patient safety literature,[92] hospital ethics committees, IRB members, and clinical ethicists can make valuable contributions to these decisions.

To avoid surprises and delays, the decision about the nature of the activity that the registry is intended to support should be made prospectively, in consultation with appropriate officials of the funding agency and officials of the organization where the registry will reside and, in the event of contemplated use of the registry for FDA-regulated purposes, the FDA. Some research institutions may have policies that either require IRB review for quality A/I, especially if publication of the activity is likely, or exclude them from IRB review. Frequently, IRBs make this determination on a case-by-case basis.

3.3 Potential for Individual Patient Identification

The specific regulatory requirements applicable to the use or disclosure of patient information for the creation of a registry to support human subjects research depend in part on the extent to which patient information received and maintained by the registry can be attributed to a particular person. Various categories of information, each with a variable potential for identifying individuals, are distinguished in the Privacy Rule: *individually identifiable* health information, *de-identified* information (all identifying elements removed), and a *limited dataset* of information (specified direct identifiers removed).[93] The latter two categories of information may include codes that are assigned to each registry entry and could permit the re-identification of the entry by someone with the legend for the code, provided certain conditions are met.

Common Rule requirements would apply to any human subjects research involving information that is individually identifiable and obtained by the investigator conducting the research, and that is supported or funded by a Common Rule Agency or an FWA in which the institution has "checked the box."[94] The definition of "human subject" in the Common Rule is "a living individual about whom an investigator (whether professional or student) conducting research: (1) [o]btains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (2) [o]btains, uses, studies, analyzes, or generates identifiable information or identifiable biospecimens."

Private information includes information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.[95]

Registry developers should be mindful of the ways in which notions of identifiability are being scrutinized under the Common Rule, particularly as relating to genomic data. The Common Rule Final Rule, which has a compliance date of January 21, 2019, would require the Common Rule Agencies to periodically assess: (i) what constitutes identifiable private information and

identifiable biospecimens, and (ii) whether there are analytic technologies and techniques that should be considered to generate identifiable private information and identifiable biospecimens.[96] OHRP has indicated that it expects whole genome sequencing will be one of the first such technologies to be evaluated as part of this assessment.[97] It is prudent to consider whether, under the applicable facts and circumstances, genomic data proposed to be used in connection with the registry would constitute identifiable private information or whether certain safeguards (such as robust data minimization protocols and access restrictions) may be appropriate to protect the interests of the individuals whose data is being used. Registry developers should consult an IRB early in the process of selecting data elements to obtain guidance about whether registry activities constitute human subjects research or may be exempt from Common Rule requirements.

Also among the criteria specified by the Common Rule for IRB approval of research involving human subjects are provisions to protect the privacy of subjects and to maintain the confidentiality of data.[98] In addition, the consent process for research subjects should include explicit information about the confidentiality protections in place when records containing identifiers are going to be used.[99]

Data collection frequently requires patient identifiers, especially in prospective registries with ongoing data collection, revision, and updates. Secondary or subsequent research use by outside investigators (i.e., those not involved in the original data collection) of patient information containing direct identifiers is complicated, however, because ethical principles and regulatory criteria for the conduct of human subjects research require that risks, including risks to confidentiality of patient identifiable information, be minimized.[100] Under the Privacy Rule, covered entities are permitted to use and disclose protected health information for research with individual authorization, or without individual authorization under limited circumstances set forth in the Privacy Rule, as described above. In order for a HIPAA authorization to allow for secondary research uses of the patient information, the authorization must adequately describe the purpose of the future research. According to guidance that OCR issued pursuant to Section 2063(b) of the 21st Century Cures Act (which reiterates language in the preamble to the January 25, 2013 omnibus HIPAA final rule), a HIPAA authorization for the use or disclosure of PHI for future purposes (such as research) need not "specify each specific future study if the particular studies to be conducted are not yet determined."[101] Nonetheless, the authorization must adequately describe the purpose of any future use "such that it would be reasonable for the individual to expect that the protected health information could be used or disclosed for such future research."[102] Thus, unless the registry developer has anticipated and adequately described the purposes of the secondary research in the initial HIPAA authorization received from a patient, the initial authorization to contribute PHI to the registry may not have contemplated the use of the PHI for secondary research purposes. Although the HIPAA Rules may not apply directly to researchers that receive PHI pursuant to a HIPAA authorization (unless they are also covered entities or business associates), covered entities have a responsibility to ensure that the registries to which they disclose PHI are limiting their uses and disclosures of PHI to those stated in the original HIPAA authorization. Additionally, if the registry intends to seek IRB approval to conduct a secondary use study, the IRB is likely to consider the scope of the original consent and HIPAA authorization for the initial collection of the PHI.

In cases where there is no authorization for the secondary research, there may be other options under the Privacy Rule for secondary use of the data collected, such as de-identification of the information or the creation of a limited dataset.[103] Chapter 16 of the third edition of the User's Guide provides a discussion of the technical and legal considerations related to linking registry data for secondary research purposes.

The Privacy Rule also addresses potential pathways for the use and disclosure of a limited dataset. A limited dataset, which is still considered PHI under HIPAA, does not include specified direct identifiers of the patient, or the patient's relatives, employer, or household members.[104] In order to create a limited dataset, the covered entity or business associate (if permitted by the applicable business associate agreement) must remove the same identifiers listed under the Privacy Rule's de-identification standard[105] except for dates and certain geographic information such as city, state, and/or zip code.[106]

In an electronic environment, operationalizing de-identification or the removal of identifiers to create a limited dataset can be a complex task. Data suppression limits the utility of the information from the registry. Linkage or triangulation of information can enable the re-identification of individuals, even if unintended. A technical assessment of electronic records for their uniqueness within any dataset may be necessary to minimize the potential for re-identification. When publishing aggregated data for public use or to demonstrate results from research studies, researchers often take safety precautions to prevent the inadvertent disclosure of PHI within the results. For example, aggregated results revealing that only one person participating in the study suffered a rare adverse event could potentially reveal the identity of the individual, depending on the nature of the data. As a result, some study publishers institute policies requiring the suppression of aggregated results that reveal a small subgroup.[107] An evaluation for uniqueness should be performed to ensure that the electronic format does not produce a potential for identification greater than this standard practice, including when the information is triangulated within a record or linked with other data files.

If a registry for research, public health, or other purposes will use any of the categories of health information discussed below, a registry developer should establish the purpose and contemplated uses of the registry and determine the applicability of the Common Rule and Privacy Rule requirements to the collection and use of registry data. Note that certain institutional policies or other considerations may require or warrant consultation with an IRB at the outset to determine whether the development or use of a registry constitutes human subjects research or an exemption determination. Further, as noted above, FDA GCP requirements may apply depending on the purpose of the development or use of the registry, even only de-identified health information or a limited dataset is involved. In addition, the registry developer may need to consult a representative of the information technology or health information system office of each healthcare provider or insurance plan that will be a source of data for the registry, so as to obtain feasibility estimates of data availability and formats.

3.3.1 De-Identified Health Information

The Privacy Rule describes two methods for de-identifying health information.[108] The "Safe Harbor" method of de-identification requires the removal of 18 specific identifiers related to the

individual and the individual's relatives, household members, and employers. In addition, this method requires the removal of any information that "could be used alone or in combination with other information to identify an individual."[109] The "Expert Determination" method requires that a qualified expert certify that the potential for identifying an individual from the data elements is very small.

A qualified expert should have "appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable" in order to make this determination.[110] De-identified information may include a code permitting re-identification of the original record by the data source (covered entity), provided the code is not derived from information about an individual, including hash codes,[111] and resists translation. In addition, the decoding key must remain solely with the healthcare provider or plan that is the source of the patient information, and the covered entity cannot use or disclose the code for any other purpose.

Research using existing data in which individual patients cannot be identified directly or indirectly through linked identifiers does not involve human subjects as defined by the Common Rule, and thus is not subject to the requirements of the Rule.[95] Refer to the discussion later in this chapter. On the other hand, registry developers and users should note that FDA regulations do not explicitly exclude the use of de-identified information from the definition of a clinical investigation.[112]

As a prudent business practice, each healthcare provider or insurance plan or its respective business associate that is a source of de-identified information is likely to require an enforceable legal agreement with the registry developer. It should be signed by an appropriate institutional official on behalf of the registry developer. At a minimum, this agreement will likely contain the following terms, some of which may be negotiable: the identification of the content of the data and the medium for the data; a requirement that the data recipient, and perhaps the healthcare provider or insurance plan or their business associate providing the data, make no attempt to identify individual patients; the setting of fees for data processing and data use; limitations on disclosure or further use of the data, if any; and an allocation of the risks of legal liability for any improper use of the data.

In some cases, the registry developer may receive fully identifiable data from healthcare providers or insurance plans and then de-identify the data prior to including it in the registry. In these scenarios, the registry developer plays what is known as an "honest broker" function – which means that the registry developer must not only de-identify the data prior to including it in the registry, but must ensure that any identifiers that the registry developer receives remain inaccessible to the researchers that access the de-identified registry data. Registry developers that provide these "honest broker" services typically use administrative, physical, and/or technical safeguards to establish a "firewall" between the members of its workforce that have access to the fully identifiable data provided by healthcare providers and insurance plans and the workforce members that are performing analyses on the de-identified data. The firewall prevents the researcher from being able to access the identifiable information provided by health plans and insurance companies. Health providers or insurance plans that receive honest broker services from registry developers must enter into business associate agreements with the registry developer in order to disclose PHI to them for this purpose. Most healthcare providers and

insurance plans have developed a standard business associate agreement in response to the Privacy Rule and will likely insist on using it, although some modifications may need to be negotiated in order to produce registry data. A registry developer hired to create a de-identified dataset must return or destroy the direct identifiers once the business associate relationship formed for purposes of creating the de-identified dataset terminates.

3.3.2 Limited Datasets of Health Information

De-identified health information may not suffice to carry out the purposes of a registry, especially if the registry is designed to receive followup information as a result of monitoring patients over time or information from multiple sources in order to compile information on a health event (e.g., cancer incidence). Dates of service and geographic location may be crucial to achieving the purposes of the registry or to the integrity and use of the data. Health information provided to the registry that is not fully de-identified but excludes direct identifiers may constitute a *limited dataset* as defined by the Privacy Rule. A healthcare provider or insurance plan (or business associate on behalf of a provider or plan if permitted by the terms of the business associate agreement) may disclose a limited dataset of health information for research, public health, or healthcare operations purposes, provided it enters into a data use agreement (DUA) with the recipient. The terms of the DUA must satisfy specific Privacy Rule requirements.[113] Officials for both the data source and the registry developer should sign the DUA so that a legal contract results. The DUA establishes the permitted uses of the limited dataset by the registry developer (i.e., the creation of the registry and subsequent use of registry data for specified research purposes). The DUA may not authorize the registry developer to use or disclose the information in a way that would result in a violation of the Privacy Rule if done by the data source.[114] Furthermore, the DUA for a limited dataset of health information must provide that the data recipient will appropriately safeguard the information and not attempt to identify individual patients or to contact those patients.[115] Certain other requirements also apply.

An investigator who works for a healthcare provider or insurance plan to which the Privacy Rule applies and that is the source of the health information for a registry may use a limited dataset to develop a registry for its own research purpose. In these circumstances, the Privacy Rule still requires a DUA that satisfies the requirements of the Privacy Rule between the healthcare provider or insurance plan and the investigator. This agreement may be in the form of a written confidentiality agreement.[116]

As with de-identified data, a registry developer may assist a healthcare provider or insurance plan or their business associate by creating the limited dataset as an honest broker.[106] In some situations, this assistance may be crucial to ensuring that data are accessible and available to the registry. In order for the registry developer to perform the honest broker function, the Privacy Rule requires that the data source (the covered entity or their business associate) and the registry developer (in this instance acting as a business associate) enter into a business associate agreement.[117]

The registry populated with a limited dataset may include a coded link that connects the data back to patient records, provided the link does not replicate part of a direct identifier. The key to

the code (e.g., encryption key) may allow health information obtained from patients over time to supplement existing registry data or allow the combination of information from multiple sources.

If the registry data obtained by investigators constitutes a limited dataset, the investigator may elect to consult an IRB or an institutional official knowledgeable about the Common Rule requirements to determine whether the registry involves human subjects, as there is no clear crosswalk between the concept of a limited dataset under the Privacy Rule and the definition of a "human subject" under the Common Rule. Frequently, a special form for this purpose is available from the IRB. The IRB (or institutional official) should provide the registry developer with documentation of its decision. While the Secretary's Advisory Committee for the Protection of Human Subjects has issued recommended guidance for OHRP's consideration that would clarify that a limited dataset does not constitute a human subject for purposes of the Common Rule, OHRP has yet to adopt the recommendation as of the date of publication of the fourth edition of the User's Guide.[118]

3.3.3 Direct Identifiers: Authorization and Consent

As discussed above, the Privacy Rule permits the use or disclosure of patient information for research with a valid, written authorization from each patient whose information is used or disclosed.[119] The Privacy Rule specifies the content of this authorization, which gives permission for a specified use or disclosure of the health information.[120] Healthcare providers and insurance plans frequently insist on using the specific authorization forms that they have developed in order to avoid additional legal review and minimize any potential liability that they believe might be associated with use of other forms.

One exception to the requirement for an authorization occurs when a healthcare provider or insurance plan creates a registry to support its "healthcare operations."[121] The Privacy Rule's definition of "healthcare operations" explicitly includes quality improvement and quality assurance activities, outcomes evaluation, and the development of clinical guidelines, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities.[40] For example, a hospital registry created to track its patient outcomes against a recognized clinical care standard as a quality improvement initiative has a healthcare operations purpose. The hospital would not be required to obtain authorizations from its patients to use or disclose health information in a registry for this purpose.

Research use of health information containing identifiable information constitutes human subjects research as defined by the Common Rule.[95] In general, the Common Rule requires documented, legally effective, voluntary, and informed consent of each research subject.[122] The applicability of FDA GCP requirements, as noted previously, does not turn on the identifiability of the information used. Like the Common Rule, where applicable, FDA GCP regulations require documented, legally effective, voluntary, and informed consent of each research subject.[123]

Documentation of the consent process required by the Common Rule or FDA regulations may be combined with the authorization required by the Privacy Rule for disclosure and use of health information.[124] However, registry developers should be aware that a healthcare provider or

insurance plan may not immediately accept the combined form as a valid authorization and may insist on legal review of a combined form before disclosing any health information.

In addition to being voluntary and legally effective, an individual's consent should contain meaningful information about the research, including what activities are involved, and the expected risks and potential benefits from participation. The Common Rule and FDA regulations require the consent process to include specific elements of information.[125] Registry developers should provide non–English-speaking patients with appropriate resources to ensure that the communication of these elements during the consent process is comprehensible. All written information for patients should be translated, or else arrangements should be made for qualified translators to assist in the consent process.

IRBs may approve waivers or alterations of both authorization (for use or disclosure of patient information for registry purposes) and consent (to registry participation), provided the research use of health information satisfies certain regulatory conditions. In addition, the Privacy Rule provides for the ability of privacy boards to approve waivers of authorization for the research use of health information where an organization does not have an IRB.[126] Waivers are discussed in detail below.

In certain limited circumstances, research subjects can consent to future unspecified research using their identifiable private information. The Common Rule permits an IRB-approved consent process to be broader than a specific research project[124] and to include information about research that may be done in the future. In its review of such future research, an IRB can subsequently determine that the previously obtained consent (1) satisfies or (2) does not satisfy the regulatory requirements for informed consent. If the previously obtained consent is not satisfactory, an additional consent process may be required; alternatively, the IRB may grant a waiver of consent, provided the regulatory criteria for a waiver are satisfied.

Investigators may seek informed consent for future unspecified research through one of two pathways under the revised Common Rule. First, as permitted under the pre-2018 Common Rule and now under the revised Common Rule, research subjects may provide consent to future research using their identifiable private information through a consent process that meets the elements set forth in 45 C.F.R. § 46.116(a)-(c), with the purpose of any future research described in accordance with such required elements (to the extent not waived or altered by the IRB).[127] In addition, under the revised Common Rule, individuals could consent to such research through the new "broad consent" pathway, where the investigator maintains documentation of the broad consent, an IRB conducts a limited review of the proposed use of the information and the scope of the consent, and the investigator does not include returning individual research results to subjects as part of the study plan.[128] This new consent pathway could allow registry developers to collect data in registries for secondary use studies under a more limited IRB review than previously required. The broad consent provided by individuals under this new consent pathway must include several required elements, including a description of reasonably foreseeable risks and benefits, information about how the confidentiality of the information will be maintained, a description of the type of information that might be used for research purposes and the types of research that will be conducted, a description of the time period that the information will be used, and, if appropriate, a statement that the information may be used for commercial profit (and if the individual will share in the research profit).[129] OHRP has made clear that this new consent

pathway does not replace the pathways noted above for conducting research – i.e., (1) full IRB review of a consent process that is designed to meet the "study-specific" informed consent framework as applied to future use and future studies involving existing datasets; and (2) obtaining a waiver of the consent and authorization requirements from an IRB. In addition, note that certain restrictions and requirements apply in the event an investigator seeks to use the broad consent pathway. For example, an IRB may not waive the informed consent requirement for the storage, maintenance, or secondary research use of identifiable private information or identifiable biospecimens if an individual was asked but declined to provide broad consent for the applicable research activity.[130]

Additionally, note that the elements of informed consent that are required under FDA regulations are comparable to those under the Common Rule. In particular, FDA has issued guidance indicating that the new basic and additional elements of informed consent under the revised Common Rule "are not inconsistent with FDA's current policies and guidances," thereby helping to avoid the need to develop separate informed consent forms to comply with FDA requirements and the revised Common Rule.[131]

Consistent with the parameters discussed above, an IRB-approved consent process for the creation of a research registry should include a description of the specific types of research to be conducted using registry data. For any future research that involves private identifiable information maintained by the registry, the IRB may determine that the original consent process (for the creation of the research registry) satisfies the applicable regulatory requirements because the prospect of future research and future research projects were adequately described. The specific details of that future research using registry data may not have been known when data were collected to create the registry, but that research may have been sufficiently anticipated and described to satisfy the regulatory requirements for informed consent. For consent to be informed as demanded by the ethical principle of respect for persons, however, any description of the nature and purposes of the research should be as specific as possible.

If a registry developer anticipates subsequent research use of identifiable private registry data, he or she should request an assessment by the IRB of the description of the research that will be used in the consent process for potential subjects at the time the data are initially collected. Nonetheless, in its review of any subsequent research, an IRB may require an additional consent process for each research subject or may grant a waiver for obtaining further consent.

With respect to HIPAA, historically, HHS rejected broadening the description of purpose in authorizations under the Privacy Rule to allow for future unspecified research.[132] As a result, an authorization for the use or disclosure of health information to create a research registry could not also authorize the future research uses of the information if the specific details of the future studies were not known when the authorization was obtained.[133] However, under the modified HIPAA Privacy Rule released on January 25, 2013, HHS modified its prior interpretation and guidance that research authorizations must be research study specific.[134] While this modification does not make any changes to the authorization requirements at 45 CFR § 164.508, HHS no longer interprets the "purpose" provision for authorizations as permitting only study-specific descriptions. This change now allows future research to be authorized provided the authorization adequately describes the purposes of any future research such that it would be reasonable for the individual to expect that his or her health information could be used or disclosed for such future

research.[135] Where an authorization for the use or disclosure of registry data for the future research does not exist, a healthcare provider or health insurance plan maintaining the registry may need to obtain an additional authorization for the research from individuals or seek a waiver of authorization from an IRB or privacy board. Alternatively, the use or disclosure of a limited dataset or de-identified registry data can occur, provided regulatory criteria are satisfied. Registries maintained by organizations to which the Privacy Rule does not apply (e.g., funding agencies for research that are not healthcare providers or insurance plans, professional societies, or non-healthcare components of hybrid entities such as in many universities) are not legally bound by the limited purpose of any original authorization that was obtained to permit data sources to disclose identifiable patient information to the registry. However, data sources or their business associates that are subject to the Privacy Rule are unlikely to be willing to provide patient information without a written agreement with the registry developer that includes legally enforceable protections against redisclosure of identifiable patient information. Regardless of whether such a written agreement is in place, a valid authorization must contain a warning to patients that their health information may not be protected by Privacy Rule protections once disclosed to recipient organizations.[136]

3.3.4 Certificates of Confidentiality and Other Privacy Protections

Certificates of confidentiality (CoCs) granted by the NIH permanently protect identifiable, sensitive information about research subjects from legally compelled disclosure. For the purposes of CoCs, identifiable, sensitive information is broadly defined to include any information "through which an individual is identified," or "for which there is at least a very small risk, as determined by current scientific practices or statistical methods, that some combination of the information, a request for the information, or other available data sources could be used to deduce the identity of an individual."[137]

Prior to the enactment of the 21st Century Cures Act, the issuance of CoCs was discretionary by the NIH or other HHS agencies to which a researcher may submit a request for a CoC. With the passage of the 21st Century Cures Act, the Secretary of HHS is required to issue a CoC to all ongoing or new research funded in whole or in part by the federal government as of December 13, 2016, that collects or uses identifiable, sensitive information.[138] For NIH-funded research, this includes any non-exempt human subjects research as defined under the Common Rule, as well as any research involving generating individual level, human genomic data from biospecimens or using such data, whether the data is recorded in a manner that enables the identification of human subjects as defined under the Common Rule.[139] That research involving non-identifiable genomic data is required to be protected by a CoC further underscores the diverse and shifting views on genomic privacy and the safeguards that are deemed necessary or appropriate for research involving genomic information. Separately, researchers may continue to seek a CoC for non-federally funded research involving the collection or use of identifiable, sensitive information, which HHS may issue in its discretion.[140] Registry developers and investigators should determine whether a registry or research conducted using the registry would be subject to the amended NIH CoC policy described above and thus be automatically issued a CoC.

An investigator whose research project has been granted a CoC may refuse to disclose identifying information collected for that research even though a valid subpoena demands that information for a civil, criminal, administrative, or legislative proceeding at the Federal, State, or local level. The protection provided by a CoC is intended to prevent the disclosure of personal information that could result in adverse effects on the social, economic, employment, or insurance status of a research subject.[141] Detailed information about CoCs is available on the NIH website.[142]

The grant of a CoC to a research project, however, is not intended to affect State laws requiring healthcare and other professionals to report certain conditions to State officials; for example, designated communicable diseases, neglect and abuse of children and the elderly, or threats of violent harm. If investigators are mandatory reporters under State law, in general, they continue to have a legal obligation to make these reports.[143] In addition, other limitations to the privacy protection provided by CoCs exist and may be relevant to particular research projects. Information on the NIH website describes some of these other legal limitations.[142]

Registry developers should also be aware that Federal law provides specific confidentiality protections for the identifiable information of patients in drug abuse and alcoholism treatment programs that receive Federal funding.[144] These programs may disclose identifiable information about their patients for research activities only with the documented approval of the program director or authorization of the patient.[145] The basis for the director's approval is receipt of written assurances about the qualifications of the investigator to conduct the research and the confidentiality safeguards incorporated into the research protocol, and an assurance that there will be no further disclosure of identifying information by the investigator. Moreover, an independent review of the research project should determine and verify in writing that the protocol provides adequate protection of the rights and welfare of the patients and that the benefits of the research outweigh any risks to patients.[145] Prior to submitting proposed consent documentation to an IRB, registry developers should consult legal counsel about the limitations of these confidentiality protections.

As a condition of approval, IRBs frequently require investigators to obtain a CoC for research involving information about substance use disorder or other illegal activities (e.g., underage purchase of tobacco products) and sexual attitudes and practices. Registry developers should consult legal counsel to determine if and how the limitations of a CoC may affect privacy protection planning for registry data. In all circumstances, the consent process should ensure that clear notice is given to research subjects about the extent of privacy protections they may expect for their health information when it is incorporated into a registry.

In the absence of a CoC, a valid subpoena or court order for registry data will usually compel disclosure of the data unless State law specifically protects the confidentiality of data. For example, Louisiana's laws specifically protect the collection of information related to tobacco use from subpoena.[146] On the other hand, a subpoena or court order may supersede State law confidentiality protections. These legal instruments can be challenged in the court having jurisdiction for the underlying legal proceeding. In some circumstances, research institutions may be willing to pursue such a challenge. The remote yet definite possibility of this sort of disclosure should be clearly communicated to research subjects as a limitation on confidentiality

protections, both during the consent process and in an authorization for use or disclosure of patient information.

State law may assure the confidentiality of certain quality A/I activities performed by healthcare providers as peer review activities.[147] When State law protects the confidentiality of peer review activities, generally, it is implementing public policy that encourages internal activities and initiatives by healthcare providers to improve healthcare services by reducing the risks of medical errors and systematic failures. Protection by peer review statutes may limit the use of data generated by quality A/I activities for any other purposes.

3.3.5 Waivers and Alterations of Authorization and Consent

Waiver or alteration of authorization and informed consent is a key potential pathway for registry developers in the creation and use of health information registries. This is particularly true where the creation of the registry is limited to the collection of data originally created for other purposes, and does not involve the collection of information, such as patient-reported outcomes, specifically for purposes of the registry. Where the needs of a registry may be satisfied through the secondary use of data, including the data of a large number of individuals, waiver of authorization and informed consent is a pathway that many registry developers may pursue.

The Privacy Rule, the Common Rule, and FDA guidance[148] all provide for the ability of IRBs (and, in the case of the Privacy Rule, privacy boards) to waive or alter the authorization requirement (in the case of the Privacy Rule) and informed consent requirement (in the case of the Common Rule and FDCA) for the disclosure or use of health information for research purposes, provided applicable criteria are satisfied. It is important for registry developers to keep distinct the terms “informed consent” and “authorization,” as they are not interchangeable with respect to the Privacy Rule, the Common Rule, and FDA regulations. As described above, authorization is the term used to describe individuals’ written agreement to the use or disclosure of their PHI under the Privacy Rule, while informed consent is the term used to describe research subjects’ agreement to participate in research or a clinical investigation, as required by the Common Rule and FDA regulations, respectively. There are separate and distinct requirements for obtaining and waiving or altering each of these permissions.

As technological advances enable and facilitate the collection and sharing of vast amounts of data regarding thousands or even millions of individuals, the use of waivers should be considered in view of the potential risks to the patients participating in the registry. A waiver of authorization potentially imposes the risk of a loss of confidentiality and consequent invasion of privacy. A waiver of consent potentially imposes risks of harm from the loss of self-determination, dignity, and privacy expected under the ethical principles of respect for persons and beneficence. Acknowledging these potential risks, regulatory criteria for waiver and alterations require an IRB or privacy board to determine that risks are minimal, in addition to other criteria. This determination is a necessary condition for approval of an investigator’s request for a waiver or alteration of these permissions.

The following discussion refers only to waivers; registry developers should note that privacy boards and IRBs may approve alterations to authorizations or the consent process, provided a requested alteration satisfies all of the criteria required for a waiver by the Privacy Rule,

Common Rule, or FDA regulations and guidance, as applicable. While alterations are arguably more aligned with the principle of respect for persons because they entail obtaining some (albeit altered) form of authorization or informed consent from the subject, they may be less practicable where very large numbers of individuals are contemplated as participants.

The Privacy Rule permits a covered entity to obtain the approval of an IRB or privacy board for a waiver of authorization if the following criteria are met: (1) the use or disclosure involves no more than minimal risk to the privacy of individuals; (2) the research cannot practicably be conducted without the waiver; and (3) the research cannot be practicably conducted without access to, and use of, the health information. The determination of minimal risk to privacy includes several elements: an adequate plan to protect identifiers from improper use or disclosure; an adequate plan to destroy identifiers at the earliest opportunity, unless a health or research justification exists to retain them; and adequate written assurances that the health information will not be reused or disclosed to others, except as required by law, as necessary for oversight of the research, or as permitted by the Privacy Rule for other research.[149] The registry developer should provide detailed documentation of the IRB or privacy board's decision to the healthcare provider or insurance plan (covered entity) that is the source of the health information for registry data. The documentation should clearly communicate that each of the criteria for a waiver required by the Privacy Rule has been satisfied.[150] The privacy board or IRB documentation should also provide a description of the health information it determined to be necessary to the conduct of the research and the procedure it used to approve the waiver.[151] A healthcare provider or insurance plan might insist on legal review of this documentation before disclosing any health information.

The criteria for waiver of consent under the Common Rule are similar to those for a waiver of authorization under the Privacy Rule. Specifically, an IRB may waive the requirement to obtain informed consent under the Common Rule if the following criteria are met: (1) the research involves no more than minimal risk to subjects; (2) the research could not practicably be carried out without the requested waiver; (3) if the research involves using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format; (4) the waiver will not adversely affect the rights and welfare of the subjects; and (5) whenever appropriate, the subjects or legally authorized representatives will be provided with additional information after participation.[152] The criterion for additional information can be satisfied at least in part by public disclosure of the purposes, procedures, and operations of a registry, as discussed below.

While the FDA has not always permitted IRBs to waive the informed consent requirement under FDA regulations, it has initiated notice and comment rulemaking to allow for such IRB waiver.[153] Prior to publishing its proposed rule, FDA issued guidance indicating that it does not intend to object to clinical investigations for which an IRB has waived consent pursuant to the criteria set forth in the guidance.[154] Specifically, the guidance states that an IRB may waive the FDA's consent requirement if it determines that: (1) the clinical investigation involves no more than minimal risk to subjects (as defined under 21 CFR 50.3(k) or 56.102(i)); (2) the waiver will not adversely affect the rights and welfare of subjects; (3) the clinical investigation cannot practicably be carried out without a waiver; and (4) whenever appropriate, subjects will be provided with additional information after participation.[155] It remains to be seen whether FDA will align its waiver criteria with those under the revised Common Rule as set forth above, which

added a fifth criterion requiring that the research could not practicably be carried out without the use of information in an identifiable format.

Some IRBs produce guidance about what constitutes "not practicable" justifications and the circumstances in which justifications are applicable. For population-based research projects, registry developers may also present the scientific justification of avoiding selection bias. A waiver permits the registry to include the health information of all patients who are eligible. As patient portals, email, and other electronic platforms make it easier to communicate remotely with digitally connected patients, registry developers should be mindful of whether the circumstances support a finding of "not practicable."[156] Likewise, where a registry involves the collection of particularly sensitive information, or use for research that may be potentially stigmatizing, registry developers and users should consider whether the "minimal risk" criterion is satisfied.[157]

Separately, the Common Rule and FDA regulations also permit an IRB to waive documentation of the consent process under certain circumstances. Waiver of documentation may be particularly useful where registry developers have a platform that enables electronic consent but not the ability to obtain an electronic signature that meets requirements under applicable law. For purposes of the Common Rule, one set of conditions for approval of this limited waiver requires that the only record linking an individual subject to the research is the consent document, and that the principal risk to subjects is the potential harm from a breach of confidentiality. Each subject individually determines whether his or her consent should be documented.[158] Alternatively, as permitted under the Common Rule and FDA regulations, an IRB can waive documentation of consent if the research involves no more than minimal risk of harm to subjects and entails no procedures for which written consent is normally obtained outside of a research context.[159] For any of these circumstances, the IRB may require the investigator to provide subjects with written information about the research activities in which they participate.[160] The written information may be as simple as a statement of research purposes and activities, or it may be more elaborate, such as a website for regularly updated information describing the progress of the research project.

3.3.6 Patient Safety Organizations

This section provides basic information about the Patient Safety and Quality Improvement Act of 2005 (PSQIA) and an overview of some considerations for registries that are considering becoming or working with a Federally-listed Patient Safety Organization (PSO). The PSQIA was enacted in response to a 1999 report by the Institute of Medicine (now the National Academy of Medicine) that identified medical errors as a leading cause of hospital deaths in the United States, with many such errors being preventable.[161] It creates Federal confidentiality and privilege protections for certain information that meets the criteria in the statute to qualify as patient safety work product (PSWP). The Patient Safety Rule at 42 CFR Part 3, the implementing regulation for the PSQIA, became effective on January 19, 2009.

The PSQIA authorized the U.S. Department of Health and Human Services (HHS) to list PSOs, entities with patient safety expertise, to which providers can voluntarily report patient safety events with the aim of improving patient safety and the quality of care. PSOs provide feedback

to providers to assist them with improving patient safety, encouraging a culture of safety and minimizing patient risks. The HHS Agency for Healthcare Research and Quality (AHRQ) is responsible for listing PSOs.

Generally, PSWP is any data, reports, records, memoranda, analyses, or written or oral statements which: (1) could improve patient safety, healthcare quality, or healthcare outcomes and are assembled or developed by a provider to be reported to, and are reported to a PSO, or are developed by a PSO to conduct patient safety activities; or (2) identify or constitute the deliberations or analysis of, or fact of reporting to, a patient safety evaluation system. Information collected, maintained, or developed separately from, or existing separately from, a patient safety evaluation system as defined in the statute is excluded from the definition of PSWP. Thus, individual patient medical records, billing and discharge information, and any other original patient or provider records are not confidential and privileged under the PSQIA.

With specified exceptions, PSWP is privileged and confidential and is not subject to subpoena, order or discovery in connection with a Federal, State, or local civil, criminal, or administrative proceeding, including a Federal, State, or local civil or administrative disciplinary proceeding; may not be admitted as evidence in any Federal, State, or local governmental civil, criminal or administrative rulemaking proceeding, or administrative adjudicatory proceeding; is not subject to disclosure under the Federal Freedom of Information Act or any other similar Federal, State, or local law; and may not be admitted in a professional disciplinary proceeding of a professional disciplinary body established or specifically authorized under State law.

Once PSWP is received by a PSO, it may be aggregated and analyzed by the PSO to assist a provider in determining and addressing underlying factors that contribute to patient safety risks. Under the PSQIA, PSWP may not be disclosed unless specified requirements in the Patient Safety Rule are met. Civil money penalties may be imposed for confidentiality violations.[162] In addition, the HIPAA Privacy Rule's limitations on uses and disclosures may apply where PSWP includes protected health information (PHI).

Information collected by a provider to comply with external reporting requirements (for example, State incident reporting requirements) is not PSWP, and PSWP generally may not be used to comply with such obligations. Thus, a significant amount of data in its original form remains outside the PSWP definition. This includes registry data that is not developed by a provider or an AHRQ-listed PSO in a manner that meets the definition of PSWP. The statute and regulation provide no protection for information in registries acting outside the protected scope of the PSO arena. For example, if a PSO also operates a registry that submits data to the Centers for Medicare and Medicaid Services (CMS) on behalf of individual clinicians or groups to meet the requirements for their participation in a quality payment incentive program, neither the data submitted by the provider to the registry for this purpose nor the data submitted by the registry to CMS is PSWP. A provider may submit a copy of these registry data to the PSO, and the copy may be protected as PSWP, as long as the original exists outside of the patient safety evaluation system.

A registry may choose to seek listing by AHRQ as a PSO so long as it meets the requirements to become listed as a PSO. However, the registry should carefully assess whether it would be able to conduct all of its intended operations as a PSO. The disclosure of PSWP is limited to specific

permissible disclosures enumerated in 42 CFR § 3.206(b) (and is also subject to the HIPAA Privacy Rule and other requirements, if applicable). The limitations on disclosure in the Patient Safety Rule might not allow a registry to conduct research, publish or disseminate PSWP to the same extent or in the same way it would do so with registry information that is not PSWP.

There are some additional points an entity operating a registry may want to consider before seeking listing as a PSO:

- Information received by a registry from providers prior to the registry becoming an AHRQ-listed PSO would not be PSWP unless the information was PSWP prior to being provided to the registry. However, a provider could submit or authorize the submission of a copy of non-PSWP it previously reported to the registry to the PSO, as long the original is maintained separately. The copy could become PSWP, provided the copy satisfies all other requirements of the definition of PSWP. Before preparing copies for reporting to the PSO of information originally developed for a registry, the provider will want to ensure that making such copies is permissible; that disclosing the information to a PSO does not affect other confidentiality or privilege protections, such as those under State law; and that the disclosure is not prohibited by other laws, regulations or contractual obligations.
- When PHI is contained in PSWP, both the Patient Safety Rule and HIPAA Privacy and Security Rules must be taken into account. The Patient Safety Rule permits PSWP containing PHI to be disclosed to a PSO for the conduct of patient safety activities, which are considered healthcare operations under the HIPAA Privacy Rule.
- The Patient Safety Rule definition of provider primarily relates to individuals or entities licensed or otherwise authorized under U.S. State law to provide healthcare services and individuals or entities that deliver healthcare as part of U.S. Federal, State, local, or tribal governments. Data reported to a PSO from healthcare providers who do not satisfy this definition is not confidential or privileged PSWP. Additionally, the PSQIA privilege protections only apply in U.S. tribunals.
- A registry should also consider the long-term implications for any data that are PSWP if its PSO is subsequently delisted. The regulations do not permit a former PSO to retain any of the PSWP it collected during its period of listing, even if the PSO delists voluntarily. All PSOs must meet the PSWP disposition requirements set forth at 42 CFR § 3.108(b)(3) at the time of delisting.

Finally, it is important for registry developers to know that, instead of becoming a PSO itself, a registry may elect to form a separate unit or division of a legal entity, or a separate legal entity that can become listed as a PSO. The Patient Safety Rule defines this type of PSO as a "component PSO." The Patient Safety Rule prohibits health insurance issuers from becoming listed as or forming a PSO. However, the other excluded entities may create or designate a component organization to seek listing as a PSO. (see 42 CFR § 3.102(a)(2)) Whether the component PSO is part of an excluded entity or not, the Patient Safety Rule's additional requirements regarding component PSOs must be satisfied. A component PSO must maintain PSWP separately from non-PSWP and from the rest of the parent organization.

3.4. Developments Affecting the HIPAA Privacy Rule

3.4.1 The Institute of Medicine Report

On February 4, 2009, the Institute of Medicine (IOM) published a report that examined how research was being conducted within the framework of the Privacy Rule. The IOM Report presented findings and recommendations of an IOM Committee tasked with assessing the impact of the HIPAA Privacy Rule on health research. This group had proposed recommendations to ensure that important health research might be conducted while maintaining or strengthening privacy protections for research subjects' health information.[163] The IOM Report stated that certain Privacy Rule requirements were difficult to reconcile with other regulations governing the conduct of research, including the Common Rule and the FDA regulations, and it noted a number of inconsistencies among applicable regulations related to the de-identification of data and the ability to obtain informed consent for future research studies, among other differences.

Citing more uniform regulations in other countries, the IOM Report affirmed that "a new direction is needed, with a more uniform approach to patient protections, including privacy, in health research."[164] As its primary recommendation, the IOM Committee held that Congress should authorize HHS and other Federal agencies to develop a new approach to protecting privacy that would apply uniformly to all health research and to exempt health research from the Privacy Rule when this new approach was implemented. Until such an overhaul could be accomplished, the IOM Committee called upon HHS to revise the Privacy Rule and associated guidance to address certain issues. HHS addressed some of these issues in the January 25, 2013, modifications to the Privacy Rule, such as by allowing HIPAA authorizations to encompass future research and removing prohibitions on combining certain HIPAA authorizations for multiple research studies, thereby harmonizing these HIPAA Privacy Rule requirements with the Common Rule. Nevertheless, registry operators should be aware that additional clarifications or modifications to the Privacy Rule as it relates to research activities may continue to be made in the future.

3.4.2 The Genetic Information Nondiscrimination Act of 2008

The Genetic Information Nondiscrimination Act of 2008 (GINA) was signed into law on May 21, 2008. In general, GINA prohibits discrimination in health insurance coverage (Title I) and employment (Title II) based on genetic information. GINA defines genetic information as, with respect to any individual, information about the individual's genetic tests, the genetic tests of the individual's family members, and the manifestation of a disease or disorder in the individual's family members (e.g., family health history). Title I of GINA took effect for most health insurance plans on May 22, 2009, and Title II became effective for employers on November 21, 2009. GINA also specifies that the definition of genetic information includes the genetic information of a fetus carried by a pregnant woman and an embryo legally held by an individual or family member utilizing an assisted reproductive technology. Pursuant to GINA, health insurers are prohibited from using the genetic information of individuals for underwriting purposes (e.g., determining health insurance eligibility and coverage, or premium setting), and employers are prohibited from using genetic information in making employment-related decisions.

In addition to its nondiscrimination requirements, GINA also required related amendments to the Privacy Rule to clarify that genetic information is PHI for purposes of the Privacy Rule, and to prohibit certain health plans from using or disclosing genetic information for underwriting purposes.[165]

3.4.3 The HITECH Act

The American Recovery and Reinvestment Act of 2009 (ARRA) was signed into law on February 17, 2009. Funds appropriated as a result of passage of ARRA are supporting new registries developed to study comparative effectiveness of treatments and protocols. It should be noted that there are no specific exceptions to regulatory or ethical requirements for such comparative effectiveness registries. Title XIII of division A and Title IV of division B of ARRA, the Health Information Technology for Economic and Clinical Health Act (HITECH Act) significantly modified the obligations of HIPAA covered entities and their business associates.

Perhaps most significantly, the HITECH Act extends to business associates direct liability for compliance with many of the key privacy and security obligations contained in the HIPAA Rules, whereas business associates were previously only contractually liable for failing to protect PHI in accordance with the terms of their business associate agreement with covered entities. Specifically, the HITECH Act imposed direct liability on business associates for compliance with the HIPAA Security Rule's requirements for implementing administrative, physical, and technical safeguards to protect electronic PHI, as well as for compliance with the use and disclosure provisions of the Privacy Rule and the terms of business associate agreement into which they enter. While many business associate agreements previously contained general safeguarding requirements (e.g., requiring the business associate to maintain appropriate technical safeguards), these agreements often had not imposed specific security requirements (e.g., a requirement that the business associate establish a security management process, which includes conducting security risk assessments and developing risk management plans). The HITECH provisions now subject business associates to civil and criminal penalties once reserved only for covered entities under the HIPAA Rules. The HITECH Act obligations imposed on business associates were finalized through HHS rulemaking on January 25, 2013, with compliance required by September 23, 2013.

The HITECH Act also created new breach notification requirements for covered entities and business associates. The Breach Notification Rule requires covered entities to notify the affected individuals, HHS, and in some cases, the media, of a breach of unsecured PHI. Notification must be provided without unreasonable delay but in no case later than 60 calendar days after the breach is discovered (except in cases of reports to HHS of breaches affecting less than 500 individuals, in which case, notification to HHS is required within 60 days after the end of the calendar year in which the breach was discovered). Depending on the circumstances, individual notification may include both direct, written notification to affected individuals via first-class mail or email, as well as substitute notice via conspicuous posting on the entity's website or in major print or broadcast media.

If a business associate experiences a breach of any unsecured PHI it maintains, the business associate must provide notification to the applicable covered entity without unreasonable delay, and in no case later than 60 calendar days after the breach is discovered, so that the covered entity can provide the notifications described above with respect to the breach or delegate that responsibility to the business associate. Any notification by a business associate must include the identification of any individual(s) whose information was accessed, acquired, or disclosed during the breach. The Department issued an interim final rule to implement the breach notification requirements on August 24, 2009, which became effective on September 23, 2009. On January 25, 2013, the Department published modifications to and made permanent these breach notification requirements.[166]

3.4.4 Summary of Regulatory Requirements

The use and disclosure of health information by healthcare providers and insurance plans for research purposes, including for registries, *are assumed* by the authors of this chapter to be subject to regulation under the HIPAA Rules and *may* be subject to the Common Rule and/or FDA regulations.

In general, the Privacy Rule permits a covered entity (or business associate on its behalf) to use or disclose patient information for registry purposes, subject to specific conditions, where the use or disclosure is: (1) for a registry supporting certain public health activities, including registries developed in connection with FDA-regulated products; (2) for a registry supporting the healthcare operations of a healthcare provider or insurance plan (covered entities), such as for quality A/I; (3) for a registry created by health oversight authorities for health system oversight activities authorized by law; (4) limited to de-identified health information; (5) limited to a "limited dataset" of patient information that lacks specified direct identifiers, and a data use agreement is in place with the recipient; (6) pursuant to patient authorizations; or (7) consistent with a waiver or alteration of authorization by an IRB or privacy board.

The Common Rule will apply to the creation and use of registry data if (1) the registry is funded by a Common Rule Agency or the organization responsible for the registry has an FWA that encompasses the registry project, regardless of funding,[167] *and* (2) the creation of the registry and subsequent research use of the registry data constitute nonexempt human subjects research as defined by the Common Rule. As interpreted by OHRP, human subjects research includes the creation or use of a registry that has any research purpose, even if the main purpose of the registry or use of the registry is not research. Registry developers are strongly encouraged to consult the IRB, not only about the applicability of the Common Rule, but also about the selection of data elements, the content of the consent process or the regulatory criteria for waiver, and any anticipated future research involving identifiable registry data.

FDA GCP requirements may apply to the creation and/or use of a registry if the registry involves the collection of data specifically for purposes of gathering information regarding the safety or effectiveness of a product in a manner that influences treatment decisions, or if the registry data relates to the use of an FDA-regulated product and the data or analyses thereof are intended to be submitted to or held for inspection by the FDA. The scope and requirements of FDA regulations with respect to the use of real world data continue to evolve, and registry developers should be

mindful of the purposes for which the registry is developed and/or used to assess the potential applicability of FDA requirements for the protection of human subjects (even if the registry or downstream use of the registry is limited to de-identified data). With the FDA's harmonization of its consent waiver and alteration requirements with those under the Common Rule and HIPAA, it is possible that more registries will be used to support FDA-regulated research.

State laws regulate public health activities and may also apply in various ways to the research use of health information. NIH may issue – either as a matter of course or upon request, depending on the funding source of the study – CoCs to particular research projects for the protection of identifiable, sensitive information from most legally compelled disclosures. Federal law provides specific privacy protections to the health information of patients in substance use disorder programs that receive Federal funding (42 CFR Part 2). Under recent revisions to 42 CFR Part 2 that became effective in 2017, HIPAA covered entities and business associates may conduct research on information from a federally supported substance use disorder program if the HIPAA covered entity or business associate has obtained and documented authorization from the patient, or a waiver of authorization consistent with the Privacy Rule and complied with the Common Rule in obtaining informed consent (if applicable to it). Researchers subject to the Common Rule and not HIPAA may also conduct research substance use disorder information covered by 42 CFR Part 2 if they conduct the research in compliance with the Common Rule. Note, the Coronavirus Aid, Relief and Economic Security Act (CARES Act), which was signed into law on March 27, 2020, amended certain confidentiality protections and requirements for substance-use disorder records under 42 CFR Part 2. HHS is expected to issue the regulations amending 42 CFR Part 2 later in 2020 or early 2021. Although there will remain some discrepancies, the new regulations are expected to more closely align 42 CFR Part 2 and the HIPAA Privacy Rule.

The institutional policies of healthcare providers and insurance plans may also affect the use and disclosure of the health information of their patient or insured populations. Legal requirements applying to use or disclosure of health information for research are evolving and can significantly influence the planning decisions of registry developers and investigators. It is prudent to obtain early and frequent consultation, as necessary, with institutional privacy officers, privacy board or IRB staff and members, information system representatives of healthcare providers and insurance plans, funding agencies and the FDA (as applicable), the developers of any device or application from which the registry will directly collect data, and technology transfer representatives and legal counsel.

4. Registry Transparency, Oversight, and Data Ownership

4.1 Registry Transparency

Efforts to make registry operations transparent (i.e., to make information about registry operations public and readily accessible to anyone who is interested) are desirable and may even be required under certain circumstances involving waiver or alteration of consent. Registry transparency can educate registry participants and the public about scientific processes. Transparency also contributes to public and professional confidence in the scientific integrity and validity of registry processes, and therefore in the conclusions reached as a result of registry

activities. Public information about registry operations may also increase the scientific utility of registry data by promoting inquiries from scientists with interests to which registry data may apply. Registry participants who are more informed as to the objectives of the registry and how their data helps achieve those objectives may have an enhanced sense of investment in the registry's success and be less likely to decline or revoke permission for, or object to, the use their data. This may be particularly valuable in response to overarching concerns and anxieties among patients, consumers, or the public regarding the extent to which their privacy is being adequately protected.

Registry developers can promote transparency by making the registry's scientific objectives, governance, eligibility criteria, sampling and recruitment strategies, general operating protocol, and sources of data available to anyone who is interested. Proprietary interests of funding agencies, contractual obligations, and licensing terms for the use of patient or claims information may limit, to some extent, the information available to the public about the registry. It is important to stress that, while transparency and access to information are to be encouraged, the intent is not to discourage or criticize investments in patient registries that produce proprietary information. Neither the funding source nor the generation of proprietary information from a registry determines whether a registry adheres to the good practices described in this handbook. Funding agencies, healthcare providers, and insurance plans do, however, have an important stake in maintaining public confidence in how health information is managed. The extent of registry transparency should be prospectively negotiated with these entities.

Creating a website of information about registry objectives and operations is one method of achieving transparency; ideally, registry information should be available in various media. One example of registry transparency can currently be found on an international transplant registry website.[168] Also instructive may be the accounting of disclosures construct under HIPAA, even if HIPAA does not apply to a particular activity by operation of law. In particular, registry developers may consider the elements of information that must be included in an abbreviated accounting that a covered entity may provide for certain research-related disclosures involving 50 or more people.[169] Including similar information regarding a registry on a publicly accessible website or other platform may be helpful in promoting transparency regarding how the registry is maintained and the various ways in which it is used. Particularly where a registry collects information that is considered by many to be more sensitive (such as genomic data), the provision of additional meaningful information regarding the objectives the registry and how confidentiality is protected may be a key tool to enhance participant buy-in and mitigate the risk of participant declinations of permission, opt-outs, or revocations of permission. Note also that an IRB may require registry transparency as a condition of approval to satisfy one of the regulatory criteria for granting a waiver of consent, which is to provide "additional pertinent information after participation."[170]

4.2 Registry Oversight

Registry governance must reflect the nature and extent of registry operations. As described in Chapter 9, governing structures can vary widely, from one in which the registry developer is the sole decision-maker to a system of governance by committee(s) comprised of representatives of

all stakeholders in the registry, including investigators, the funding agency, patients, clinicians, biostatisticians, information technology specialists, and government agencies.

Registry developers should also consider appointing an independent advisory board to provide oversight of registry operations. An advisory board can assist registry operations in two important ways: (1) providing guidance for the technical aspects of the registry operations and (2) establishing the scientific independence of the registry. The latter function can be valuable when controversies arise, especially those related to patient safety and treatment, or resulting from actions by a regulatory agency. Advisory boards collectively should have relevant technical expertise, but should also include representatives of other registry stakeholders, including patients. Advisory board actions should be limited to making recommendations to the ultimate decision-maker, whether an executive committee or the registry developer.

Registry developers may also appoint other types of oversight committees to resolve specific recurring problems, such as verifying diagnoses of patient conditions or adjudicating data inconsistencies.

4.3 Data Ownership

4.3.1 Health Information Ownership and Value Proposition

There is no general consensus with respect to the ownership of health information, and multiple stakeholders assert ownership claims to health information in various forms. Certain States have enacted laws that seek to clarify ownership of health records, aiming to strike the balance between healthcare provider's rights to these records and patients' rights to maintain confidentiality of, and have access to, the information in their records. However, there is much inconsistency among these laws,[171] and, accordingly, these laws are helpful to outline providers' and patients' rights, but do not definitely answer the ownership question. In addition to healthcare providers and individuals, any number of stakeholders could claim ownership of health-related information, including insurance plans, funding agencies for registry projects, research institutions, government agencies, registry developers, and investigators. The basis for these claims is typically possession or control of the tangible expression of the health information or an interest in controlling its use.

While many entities claim ownership of healthcare data, the central question should be who has the possession of the data and the right to control its use. As a general matter, there is no legal basis for assertions of ownership of facts or raw data elements. In fact, long-established public policy supports the free exchange of ideas and wide dissemination of facts as fundamental to innovation and social progress.[172] However, an entity in possession or control of health information in the form of raw data elements has the ability to control it by maintaining it as confidential, thereby acting as if it is the owner of the information because it can control how such information is shared and used. For example, an entity in possession or control of health information may transfer it to another party under contractual conditions that restrict the recipient's use and disclosure. In addition, entities that are in possession or control may allocate "ownership" of health information by contract as between them – without reference to the true owner. This may permit, for example, the funding agency for a registry to assert claims to

ownership as a matter of contract law in their sponsorship agreements with research organizations.

U.S. copyright law allows an individual or entity to claim ownership of compilations of facts if the facts are selected, coordinated or arranged in a manner that meets a degree of originality.[173] Electronic health records systems and data warehouses contain vast amounts of heath information that can be aggregated or complied in unique way, leading to datasets that may be protected under U.S. copyright law. Accordingly, it is possible to create a registry that is selected, coordinated or arranged in such a way to be protect the registry under U.S. copyright law; however, stakeholders developing registries should be aware that protection of registries as a compilation under U.S. copyright law will be thin, and will not protect the underlying data. Therefore, it is advisable to develop a protection strategy for registries that includes a combination of confidentiality requirements and copyright protection.

Notwithstanding any copyright protection that may be available, stakeholders need to bear in mind that registry data constitutes legally protected, confidential information about individual patients to which independent and varied legal protections apply. Copyright protections may marginally enhance, but do not diminish, other legal restrictions on access to and use of health information and registry data.

For more information on copyright law, see Appendix B.

4.3.2 Publications

For academic institutions, publication rights are an important component of the value proposition of healthcare data, and any publication itself (but not the data therein) constitutes a literary work under U.S. copyright law. Formal institutional policies may address publication rights resulting from faculty educational and research activities. Moreover, the social utility and benefit of any registry is evaluated on the basis of its publicly known findings and any conclusions based on them. The authors strongly encourage registry developers to maximize public communication of registry findings through the customary channels of scientific conferences and peer-reviewed journals. The goal of public communication for scientific findings and conclusions applies equally to registries operated outside of academic institutions (i.e., directly by industry or professional societies). For further discussion of developing data access and publication policies for registries, see Chapters 2 and 8.

4.3.3 Intersection of Ownership Rights With Other Considerations

The concept of ownership does not fit comfortably in the context of health information, because it largely fails to acknowledge individual patient privacy interests in health information. An inescapable personal nexus exists between individuals and information about their health. Considerations for rights with respect to health information may derive from applicable Federal or State law. The Privacy Rule, for example, provides individuals with the right to access, amend, and obtain an accounting of disclosures of their PHI contained in a designated record set, although it does not directly affect existing laws, if any, regarding property rights in health information.[174] A designated record set means a group of records maintained by or for a covered

entity that is: (i) the medical records and billing records maintained by or for a covered entity healthcare provider; (ii) the enrollment, payment, claims adjudication, and case or medical management record systems maintained by or for a health plan, or (iii) used by or for the covered entity to make decisions about individuals.[175] The assessment of whether a registry would contain data that constitutes a designated record set is a fact-specific one. Where applicable, such individual rights with respect to health information in a registry may inform the extent to which an individual has ownership rights in the information.

At the time of this writing, health information sources and other users privately reach agreement to manage access and control. The question of the extent to which individuals have rights in their own health information will likely become more pronounced insofar as registries seek to expand the collection of patient-reported outcomes. Likewise, as more health information becomes generated through connected devices, mobile applications, and other digital platforms that are deployed in a consumer context, it will be important for registry developers and the entities that collect and make available to the registry such data to consider the terms and conditions and privacy policies that apply to the platforms, and the attendant implications for the ability of individuals to access or otherwise assert rights with respect to such data. These contractual terms, the legal and regulatory requirements described above, as well as claims of property rights and concerns about legal liability, will inform the formal written agreements for the establishment and use of registries. Registry developers should also consider what, if any, rights they are willing to or should provide to the individuals whose data is being used, such as by giving individuals the ability to access or request the sharing of their own data, receive reports or updates regarding the use of their data, or other means of engagement that may enhance transparency or buy-in from the individuals with respect to the registry's mission or activities.

5. Conclusions

Ethical considerations arise in many of the essential aspects of planning and operating a registry. These considerations can affect the scientific, logistical, and regulatory components of registry development, as well as claims of property rights in health information. The guiding ethical principles for these considerations are *beneficence*, *justice*, and *respect for persons and avoidance of harm*.

At the most fundamental level, investigations that involve human subjects and that are not capable of achieving their scientific purpose are unethical. The risk-benefit ratio of such studies is unacceptable in an analysis based on the principle of *beneficence*, which obligates investigators to avoid harming subjects, as well as maximize the benefits and minimize the harms of research projects. Ethical scientific design must be robust, must be based on an important question, and must ensure sufficient statistical power, precise eligibility criteria, appropriately selected data elements, and adequately documented operating procedures and methodologies.

In addition, an ethical obligation to minimize harms requires planning for and establishing adequate protections to ensure the confidentiality of the health information disclosed to a registry, taking into consideration the evolving and diversifying nature and rapidly proliferating amounts of data that is becoming available to registry developers. Such planning should include

developing policies and procedures for the appropriate use and disclosure of registry data, and implementing physical, technical, and administrative safeguards to limit access to and use of registry data accordingly. Reducing the potential harms associated with the use of health information in a registry is particularly important (for example, where genomic or particularly sensitive health information is involved), because generally no directly offsetting benefit from participation in a registry accrues to individuals whose health information is used in the registry. According to an analysis applying the principle of *justice*, research activities that produce a significant imbalance of potential risks and benefits to participating individuals are unethical.

Protection of the confidentiality of the health information used to populate a registry reflects the ethical principle of *respect for persons and avoidance of harm*. Health information intimately engages the privacy and dignity of patients. Registry developers should acknowledge public expectations of protection for patient privacy and dignity with clear and consistent communications to patients about protections in place to prevent inappropriate access to and use of registry data.

The regulatory requirements of the Privacy Rule and Common Rule and FDA GCP regulations reflect past ethical concerns about research involving human subjects, a recognition that the use of solely data in connection with research nonetheless warrants protections for the individuals whose data is used, as well as general social anxiety about potential loss of privacy associated with rapid advances in health information systems technology and communications and biomedical developments in human genetics. Compliance with these regulatory requirements not only is a cost of doing business for a registry project, but also demonstrates recognition of the ethical considerations accompanying use of health information for scientific purposes. Compliance efforts by registry developers also acknowledge the important public relations and liability concerns of healthcare providers and insurance plans, public health agencies, health oversight agencies, and research organizations. Regulatory compliance contributes to, and generally supports, the credibility of scientific research activities and research organizations, as well as that of particular projects.

These and other Federal and State privacy laws may affect registry development, especially registries created for public health purposes. Such laws express an explicit, legislatively determined balance of individual patient interests in health information against the potential social benefits from various uses of that information, including in research. Consultation with legal counsel is strongly recommended to determine the possible effect of these laws on a particular registry project.

Additional ethical considerations also affect the operational aspects of registries, including governance, transparency, and data ownership. Registry governance, discussed in Chapter 9, should reflect both appropriate expertise and representation of stakeholders, including patients. An independent advisory committee can provide useful guidance to registry developers and managers, especially on controversial issues. Transparency involves making information about registry governance and operations publicly available. Registry transparency improves the credibility of the scientific endeavors of a registry, the use of health information for scientific purposes, and the results based on analyses of registry data. In short, registry transparency promotes public trust.

There is no general consensus with respect to the ownership of health information, and, as a general matter, there is no legal basis for assertions of ownership of facts or raw data elements. Nonetheless, in theory, copyright protections for compilations may be applied to the patient information held by healthcare providers and insurance plans, as well as to registries. In general, property rights related to health information are likely to be negotiated privately under the terms and conditions of formal agreements between registry developers, funding agencies, and healthcare providers or insurance plans. As a practical matter, "ownership" implies operational control of registry data and publication rights, and the more relevant question may be which stakeholder have the right to control use of registry data.

In summary, careful attention to the ethical considerations associated with the design and operation of a registry, and fulfillment of the applicable legal requirements, are critical to the success of registry projects and to the realization of their social and scientific benefits.

References for Chapter 7

1. Subsection (i) is based to a certain extent on 45 CFR 160.103: definition of individually identifiable health information. Subsection (ii) refers broadly to other types of personal information that is not regulated as "protected health information" under HIPAA but that may be protected under other applicable federal and state law.

2. Genomic information may relate to somatic mutations and/or germline mutations. Germline mutations are those that are inherited and present in every cell in the individual's body; importantly, they can be passed on from generation to generation. Conversely, somatic mutations are not inherited but acquired (such as through exposure to sunlight or other external causes), affect only the cells that grow from the mutated cell, and are not passed on to future generations. Griffiths AJF, Miller JH, Suzuki DT, et al., An Introduction to Genetic Analysis (7th Ed.) (2000), available at https://www.ncbi.nlm.nih.gov/books/NBK21894/.

3. 45 CFR Part 46.

4. These regulations, also known as the "Common Rule," currently also apply to institutions that receive funding for human subjects research from such federal agencies and that voluntarily elect to conduct all of its human subjects research in accordance with the Common Rule, irrespective of funding source.

5. Part C of Title XI of the Social Security Act, 42 USC §§ 1320d to 1320d-9, and section 264 of the Health Insurance Portability and Accountability Act of 1996, 42 USC §1320d-2 note ; 45 CFR Parts 160 and 164.

6. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Apr 18, 1979. [March 4, 2014].http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.

7. Public Law 93-348 (1974), Title II.

8. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. [March 4, 2014]. Noted to be under revision. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. See especially sections entitled General Ethical Principles and Informed Consent.

9. Grant RW, Sugarman J. Ethics in human subjects research: do incentives matter? J Med Philos. 2004 Dec;29(6):717–38. [PubMed]

10. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies. 1991. [March 4, 2014]. Note 6, at paragraphs 11 and 12. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm.

11. U.S. Department of Health and Human Services; Office of the Inspector General. Recruiting Human Subjects: Sample Guidelines for Practice. Jun, 2000. p. 5. OEI-01-97-00196. http://oig.hhs.gov/oei/reports/oei-01-97-00196.pdf.

12. U.S. Department of Health and Human Services; Office of the Inspector General. Recruiting Human Subjects: Sample Guidelines for Practice. Jun, 2000. Appendix A. OEI-01-97-00196. http://oig.hhs.gov/oei/reports/oei-01-97-00196.pdf.

13. Patient Protection and Affordable Care Act. Section 6002, Transparency Reports and Reporting of Physician Ownership or Investment Interests. Public Law 111-148. Mar 23, 2010. http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf.

14. Massachusetts regulation 105 CMR 970.000 implementing Massachusetts General Law, Chapter 111N, Pharmaceutical and Medical Device Manufacturer Conduct, as enacted under Chapter 305 of the Acts of 2008, An Act To Promote Cost Containment, Transparency and Efficiency in the Delivery of Quality Health Care. http://www.lawlib.state.ma.us/source/mass/cmr/cmrtext/105CMR970.pdf.

15. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies.1991. [March 4, 2014]. Note 6, at paragraphs 18–21. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm.

16. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies.1991. [March 4, 2014]. Note 6, at paragraph 26. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm.

17. Amy L. McGuire et al, Confidentiality, Privacy, and Security of Genetic and Genomic Test Information in Electronic Health Records: Points to Consider, Genetics in Medicine, July 2008; Witt, Magdalena M., and Michał P. Witt. “Privacy and Confidentiality Measures in Genetic Testing and Counselling: Arguing on Genetic Exceptionalism Again?” Journal of Applied Genetics 57.4 (2016): 483–485. PMC. Web. 26 July 2018.

18. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies.1991. [March 4, 2014]. See generally. Note 6, at paragraph 43.http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Patient Protection and Affordable Care Act of 2010, Sec. 1101. See also the. http://www.gpo.gov/fdsys/pkg/PLAW-111publ148/pdf/PLAW-111publ148.pdf.

19. Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies.1991. [March 4, 2014]. Note 6, at paragraph 40. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm.

20. For a discussion of health information privacy and security considerations for entities that do not fall within the purview of HIPAA, see U.S. Department of Health and Human Services; Office of the National Coordinator for Health Information Technology. Examining Oversight of the Privacy & Security of Health Data Collected by Entities Not Regulated by HIPAA. June 17, 2016. https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf

21. See, for example, U.S. Department of Health and Human Services regulations at 45 CFR Part 46; 21 CFR Parts 50 and 56 for research conducted in support of products regulated by the U.S. Food and Drug Administration (FDA); and Williams ED. Federal Protection for Human Research Subjects: An Analysis of the Common Rule and Its Interactions with FDA Regulations and the HIPAA Privacy Rule. Congressional Research Service, Library of Congress. Updated June 2, 2005. http://www .fas.org/sgp/crs/misc/RL32909.pdf

22. Regulations identical to 45 CFR 46 Subpart A apply to research funded or conducted by a total of 17 Federal agencies, some of which may also require additional legal protections for human subjects.

23. The final rule amending the Common Rule, published January 19, 2017, removes the option for institutions to "check the box" in their Federalwide Assurance with OHRP. 82 Fed. Reg. 7149, 7156 (Jan. 19, 2017). The compliance date of the aforementioned final rule has been delayed to January 21, 2019. Federal Policy for the Protection of Human Subjects: Six Month Delay of the General Compliance Date of Revisions While Allowing the Use of Three Burden-Reducing Provisions During the Delay Period, 83 Fed. Reg. 28,497 (June 19, 2018).

24. U.S. Department of Health and Human Services; Office for Human Research Protections. Terms of the Model Federalwide Assurance (FWA) for the Protection of Human Subjects. [March 4, 2014].http://www.hhs.gov/ohrp /assurances/assurances/filasurt.html.

25. U.S. Department of Health and Human Services; Office for Human Research Protections. Guidance on Research Involving Coded Private Information or Biological Specimens. Oct 16, 2008. http://www .hhs.gov/ohrp/policy/cdebiol.html; Issues to Consider in the Research Use of Stored Data or Tissues. Nov 7, 1997. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/issues-to-consider-in-use-of-stored-data-or-tissues/index.html.

26. U.S. Department of Health and Human Services; Office for Human Research Protections. Guidance on Research Involving Coded Private Information or Biological Specimens. Oct 16, 2008. http://www .hhs.gov/ohrp/policy/cdebiol.html.

27. 45 CFR Part 46, Subpart A.

28. Epstein M. International Society for Pharmacoepidemiology. Guidelines for Good Pharmacoepidemiology Practices (GPP) Pharmacoepidemiol Drug Safety. 2005 August;14(8):589–95. See, for example on the essential elements of a protocol. [PubMed]

29. See 45 CFR 160.103.

30. 45 CFR 160102, Applicability, and 160103, definitions of covered entity, health care provider, health plan, health care clearinghouse, and transaction.

31. 45 CFR 160.103, definition of business associate.

32. 45 CFR 160.103 defines both "disclosure" and "use" for the purposes of the HIPAA Rules.

33. 45 CFR 160.203.

34. Maryland Health General Statute § 4–303(b)(4).

35. Modifications to the HIPAA Privacy, Security, Enforcement, and Breach Notification Rules Under the Health Information Technology for Economic and Clinical Health Act and the Genetic Information Nondiscrimination Act; Other Modifications to the HIPAA Rules; Final Rule, 78 Fed. Reg. 5566 (January 25, 2013) (codified at 45 CFR pts 160 and 164).

36. HITECH Act §13404(a); 45 CFR 164.104(b); 78 Fed. Reg. at 5591.

37. 15 USC 45(a)(1); See Federal Trade Commission. Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers. March 2012. https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

38. Family Educational Rights and Privacy Act (FERPA), 20 USC 1232g, 34 CFR Part 99.

39. See U.S. Department of Health and Human Services; Office for Civil Rights. Guidance on HIPAA and Individual Authorization of Uses and Disclosures of Protected Health Information for Research. June 2019. https://www.hhs.gov/sites/default/files/hipaa-future-research-authorization-guidance-06122018%20v2.pdf.

40. 45 CFR 164.501.

41. 67 Fed Reg 53231, August 14, 2002.

42. 45 CFR 46.102(d).

43. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule. May, 2005. NIH Publication Number 05-5308. http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp. National Institutes of Health. Research Repositories, Databases, and the HIPAA Privacy Rule. Jan, 2004. See also. NIH Publication Number 04-5489. http://privacyruleandresearch.nih.gov/research_repositories.asp.

44. 45 CFR 164. 512(b).

45. Centers for Disease Control and Prevention. HIPAA Privacy Rule and Public Health: Guidance from CDC and the U.S. Department of Health and Human Services. MMWR. 2003;52 [PubMed]

46. 45 CFR 164. 512(a).

47. 45 CFR 164.508(a).

48. 45 CFR 164.514(e)

49. 45 CFR 164.514(a)–(c).

50. 45 CFR 164.528.

51. 45 CFR 164.512(i)(1)(i).

52. 21 CFR 50.3(c).

53. The FDA has issued guidance pursuant to the 21st Century Cures Act indicating that it does "not intend to object to an IRB approving a consent procedure that does not include, or that alters, some or all of the elements of informed consent [under FDA regulations] or waiving the requirements to obtain informed consent" provided the IRB finds that certain criteria – comparable to those under the Common Rule – are satisfied. In the guidance, the FDA indicates that it intends to amend its regulations regarding informed consent to incorporate this waiver or alteration pathway, after which the FDA will withdraw the guidance. U.S. Food & Drug Admin., IRB Waiver or Alteration of Informed Consent for Clinical Investigations Involving No More Than Minimal Risk to Human Subjects at 4 (July 2017), available at https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM566948.pdf.

54. 21 CFR Parts 50, 56.

55. 21 CFR 56.109(c)-(d).

56. Part 11 applies to electronic records that are created, modified, maintained, archived, retrieved, or transmitted under any FDA regulatory requirements, or that are submitted to the FDA under the statutory requirements of the FDCA or Public Health Service Act, as well as to electronic signatures. 21 CFR 11.1.

57. Note that the FDA currently exercises enforcement discretion with respect to certain Part 11 requirements and has issued draft guidance to update recommendations in its prior 2003 guidance document regarding the scope and application of Part 11 requirements. U.S. Food & Drug Admin., Part 11, Electronic Records; Electronic Signatures – Scope and Application (Aug. 2003), available at https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm125125.pdf; U.S. Food & Drug Admin., Use of Electronic Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 – Questions and Answers (Draft Guidance, June 2017), available at https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm563785.pdf.

58. See 21 CFR 50.3(g) (definition of "human subject").

59. 21 CFR 50.3(g).

60. 21 CFR 50.23, 50.24 (permitting exception from informed consent requirements only in certain life-threatening situations or for emergency research).

61. Pub. L. 114-255 (2016).

62. U.S. Food & Drug Admin., Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices (Aug. 31, 2017), available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

63. U.S. Food & Drug Admin., Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices at 11 (Aug. 31, 2017), available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

64. U.S. Food & Drug Admin., Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices at 11 (Aug. 31, 2017), available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

65. U.S. Food & Drug Admin., Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices at 12-17 (Aug. 31, 2017), available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

66. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research. [March 4, 2014]. Revised October 4, 1999, http://www.cdc.gov/od/science/integrity/docs/defining-public-health-research-non-research-1999.pdf.
Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley and Los Angeles, CA: University of California Press; 2000. pp. 126–127. New York: The Milbank Memorial Fund.
Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies.1991. [March 4, 2014]. See also. Note 6, Introduction, noting that epidemiological practice and research may overlap. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm.

67. Bellin E, Dubler NN. The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001;91(9):1512–1517.[PMC free article] [PubMed]
Lindenauer PK, Benjami EM, et al. The role of the institutional review board in quality improvement: a survey of quality officers, institutional review board chairs, and journal editors. Am J Med. 2002;113(7):575–9.[PubMed]
Lo B, Groman M. Oversight of quality improvement: focusing on benefits and risks. Arch Intern Med. 2003;163(12):1481–6. [PubMed]

68. U.S. Department of Health and Human Services; Office for Human Research Protections. [March 4, 2014]. http://www.hhs.gov/ohrp.

69. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule. pp. 2–3. NIH Publication Number 05-5308.
45 CFR 164.501 for the definition of health care operations. See also.http://privacyruleandresearch.nih.gov/healthservicesprivacy.asp.

70. The use and disclosure restrictions under the Privacy Rule do not apply to data that has been de-identified in accordance with the Privacy Rule. Office for Civil Rights, Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, at 6 (Nov. 26, 2012), available at https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdfThe Common Rule defines "human subject" in part as an individual about whom an investigator obtains identifiable private information. 45 CFR 46.102(f). For guidance regarding the use of coded identifiable private information, see Office for Human Research Protections, Coded Private Information or Specimens Use in Research, Guidance (2008), available at https://www.hhs.gov/ohrp/regulations-and-policy/guidance/research-involving-coded-private-information/index.html.

71. 21 CFR 56.102(c); Secretary's Advisory Committee on Human Research Protections, Attachment E – Recommendations on FDA Draft Real-World Evidence Guidance (May 26, 2017), available at https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-e-august-2-2017/index.html (suggesting that identifiability is one criterion the FDA could adopt in clarifying which registries constitute clinical investigations).

72. The identifiability of genomic information and biospecimens was the subject of extensive debate during the notice and comment rulemaking process to amend the Common Rule. In the September 8, 2015 Notice of Proposed Rulemaking to amend the Common Rule, the Common Rule Agencies had proposed to consider any biospecimen to constitute a "human subject" even if the investigator does not have access to information that would enable him/her to identify the individual from whom the biospecimen was obtained. Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,936 (Proposed Rule, Sept. 8, 2015). This proposal was not finalized. See Section [3.3] for a more detailed discussion of considerations relating to genomic privacy.

73. A copy of the HHS version of the "Common Rule," 45 CFR Part 46, subpart A, and additional subparts B, C, and D regarding vulnerable populations, with which institutions must comply until January 21, 2019 (the general compliance date of the Common Rule Final Rule) may be obtained on the Web site of the Office for Human Research Protection (OHRP) in the U.S. Department of Health and Human Services. [March 4, 2014]. http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html. A copy of the Common Rule Final Rule, the fate of which is subject to some uncertainty due to the delays to its effective and compliance date following its publication, likewise may be obtained on the Web site of OHRP. https://www.gpo.gov/fdsys/pkg/FR-2017-01-19/pdf/2017-01058.pdf.

74. A copy of the "Privacy Rule," 45 CFR Parts 160 and 164, may be obtained on the Web site of the Office for Civil Rights (OCR) in the U.S. Department of Health and Human Services. [March 4, 2014]. http://www.hhs.gov/ocr/hipaa/finalreg.html.

75. 21 CFR Part 50, 56, 312, and 812 are key FDA regulations governing the scope of good clinical practice.

76. The Common Rule as adopted by the U.S. Department of Health and Human Services contains special protections for certain defined "vulnerable" populations, i.e., women, human fetuses, neonates, prisoners, and children.
45 CFR Part 46, Subparts B, C, D. See.

77. American College of Epidemiology: Policy Statement on Sharing Data from Epidemiologic Studies. May, 2002. [March 4, 2014]. http://www.acepidemiology .org/sites/default/files/DataSharing.pdf.

78. 45 CFR 46.102

79. National Cancer Institute. Surveillance Epidemiology and End Results.[August 27, 2012]. http://seer.cancer.gov.

80. 45 CFR 164.512(b).

81. Gostin LO, Lillienfeld DE, Stolley PD. Foundations of Epidemiology (revised). Oxford University Press; 1994. p. 104. Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley and Los Angeles, CA: University of California Press; 2000. See also. New York: The Milbank Memorial Fund. Note 43, at 114, Table 5.

82. 45 CFR 164.512(b)(1)(iii).

83. U.S. Food & Drug Admin., Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices, at 19 (Aug. 31, 2017), available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

84. Scott Gottlieb, M.D., FDA Budget Matters: A Cross-Cutting Data Enterprise for Real World Evidence (July 10, 2018), https://blogs.fda.gov/fdavoice/index.php/2018/07/fda-budget-matters-a-cross-cutting-data-enterprise-for-real-world-evidence/.

85. U.S. Food & Drug Admin., Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices, at 12 (Aug. 31, 2017), available at https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

86. 45 CFR 164.512(d).

87. 45 CFR 164.512(d)(1).

88. 45 CFR 46.102(d).

89. Whether the Common Rule applies by operation of law to human subjects research depends on whether the activity is funded or supported by a Common Rule Agency. See Section [2.2.1] for a discussion of how the Common Rule Final Rule amends the scope of the Common Rule and removes the ability of institutions to make a voluntary election in their FWAs to apply the Common Rule to all of their human subjects research, irrespective of the source of funding or support.

90. National Institutes of Health. Health Services Research and the HIPAA Privacy Rule. May, 2005. [March 13, 2014]. pp. 2–3. NIH Publication No. 05-5308. http://privacyruleandresearch .nih.gov/healthservicesprivacy.asp.

91. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research. Oct 4, 1999. [March 4, 2014]. pp. 126–7. http://www.cdc.gov/od /science/integrity/docs/defining-public-health-research-non-research-1999.pdf. Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley and Los Angeles, CA: University of California Press; 2000. New York: The Milbank Memorial Fund. Note 43.
Council for International Organizations of Medical Sciences. International Guidelines for Ethical Review of Epidemiological Studies.1991. [March 4, 2014]. See also. Note 6, Introduction, noting that epidemiological practice and research may overlap. http://www.cioms.ch/publications /guidelines/1991_texts_of_guidelines.htm. Bellin E, Dubler NN. The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001;91(9):1512–1517.And. Note 44.

92. Bellin E, Dubler NN. The quality improvement-research divide and the need for external oversight. Am J Public Health. 2001;91(9):1512–1517.Note 44. Lindenauer PK, Benjamin EM, et al. The role of the institutional review board in quality improvement: a survey of quality officers, institutional review board chairs, and journal editors. Am J Med. 2002;113(7):575–9. Lo B, Groman M. Oversight of quality improvement: focusing on benefits and risks. Arch Intern Med. 2003;163(12):1481–6.

93. See 45 CFR 160.103 for the definition of individually identifiable health information and 45 CFR 164.514(a)–(c) and (e) on the de-identification of health information and limited datasets, respectively.

94. But see Section [2.2.1] for changes to the scope of the Common Rule under the Common Rule Final Rule.

95. 45 CFR 46.102(f).

96. 82 Fed. Reg. 7149 (Jan. 19, 2017) (see § __.102(e)(7)).

97. 82 Fed. Reg. 7149 at 7169.

98. 45 CFR 46.111(a)(7).

99. 45 CFR 46.116(a)(5).

100. 45 CFR 46.111(a).

101. U.S. Department of Health & Human Services, Office for Civil Rights. Guidance on HIPAA and Individual Authorization of Uses and Disclosures of Protected Health Information for Research. https://www.hhs.gov/sites/default/files/hipaa-future-research-authorization-guidance-06122018%20v2.pdf.

102. U.S. Department of Health & Human Services, Office for Civil Rights. Guidance on HIPAA and Individual Authorization of Uses and Disclosures of Protected Health Information for Research. https://www.hhs.gov/sites/default/files/hipaa-future-research-authorization-guidance-06122018%20v2.pdf. As indicated in this guidance, OCR issued the guidance as interim guidance to allow for additional consideration of the issue.

103. See 45 CFR 164.514(a)–(c) and (e) on the deidentification of health information and limited datasets, respectively.

104. 45 CFR 164.514(e)(2).

105. 45 CFR 164.514(b)(2).

106. 45 CFR 164.514(e). The creation of a limited dataset also does not require the removal of "[a]ny other unique identifying number, characteristic, or code", nor a lack of "actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information", as required for de-identification at 45 CFR 164.514((b)(2)(i)(R) and (b)(2)(ii), respectively.

107. See, e.g., Centers for Medicare & Medicaid Services. CMS Cell Size Suppression Policy. May 8, 2017. [August 7, 2018]. https://www.resdac.org/resconnect/articles/26.

108. 45 CFR 164.514(b).

109. The 18 identifiers that must be removed from PHI under the Safe Harbor de-identification method are: (a) names; (b) all geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of the Census: (1) the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people, and (2) the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000; (c) all elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older; (d) telephone numbers; (e) fax numbers; (f) electronic mail addresses; (g) Social security numbers; (h) medical record numbers; (i) health plan beneficiary members; (j) account numbers; (k) certificate/license numbers; (l) vehicle identifiers and serial numbers, including license plate numbers; (m) device identifiers and serial numbers; (n) web universal resource locators (URLs), (o) Internet Protocol (IP) address numbers; (p) biometric identifiers, including finger and voice prints; (q) full face photographic images and any comparable images; and (r) any other unique identifying number, characteristic, or code (except codes assigned by the covered entity to re-identify the de-identified information).

110. 45 CFR 164.514(b)(1).

111. 67 FR 53182, 53233, August 14, 2002.

112. 21 CFR 50.3.

113. 45 CFR 164.514(e)(4).

114. 45 CFR 164.514(e)(4)(ii)(A).

115. 45 CFR 164.514(e)(4)(ii)(C)(2) and (5), respectively.

116. 67 Fed Reg 53182, 53236, August 14, 2002.

117. 45 CFR 164.504(e).

118. Secretary's Advisory Committee on Human Research Protections, Attachment C – Updated FAQs on Informed Consent for Use of Biospecimens and Data (April 11, 2018), available at https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-faqs-recommendations-and-glossary-informed-consent-and-research-use-of-biospecimens-and-associated-data/index.html.

119. 45 CFR 164.508.

120. 45 CFR 164.508(c).

121. 45 CFR 164.502(a)(1).

122. 45 CFR 46.116.

123. 21 CFR Part 50.

124. 45 CFR 164.508(b)(3).

125. 45 CFR 46.116; 21 CFR 50.25.

126. 45 CFR 164.512(i)(1)(i)(B).

127. 45 CFR 46.116(a)-(c).

128. Office for Human Research Protections. Attachment C – Recommendations for Broad Consent Guidance. July 26, 2017. [August 8, 2018]. https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html.

129. Office for Human Research Protections. Attachment C – Recommendations for Broad Consent Guidance. July 26, 2017. [August 8, 2018]. https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html.

130. 45 CFR 46.116(f)(1).

131. 21 CFR 50.25. U.S. Food & Drug Administration, Impact of Certain Provisions of the Revised Common Rule on FDA-Regulated Clinical Investigations (October 2018). https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM623211.pdf.

132. 67 Fed Reg 53182, 53226, August 14, 2002.

133. National Institutes of Health. Institutional Review Boards and the HIPAA Privacy Rule. Aug, 2003. [March 4, 2014]. pp. 15–16. NIH Publication Number 03-5428. http://privacyruleandresearch.nih.gov/irbandprivacyrule.asp.

134. 45 CFR 164.508(b)(3); 78 Fed. Reg. at 5612.

135. 45 CFR 164.508(c) and 164.508(c)(1)(iv). See also U.S. Department of Health & Human Services, Office for Civil Rights. Guidance on HIPAA and Individual Authorization of Uses and Disclosures of Protected Health Information for Research. https://www.hhs.gov/sites/default/files/hipaa-future-research-authorization-guidance-06122018%20v2.pdf.

136. 45 CFR 164.508(c)(2)(iii).

137. 42 USC 241(d)(4); National Institutes of Health; Office of Extramural Research. Grants & Funding. Frequently Asked Questions, Certificates of Confidentiality. Definitions – What does identifying characteristic mean? [March 4, 2014]. http://grants1.nih.gov/grants/policy/coc/faqs.htm#369.

138. 21st Century Cures Act, Pub. L. 114-255, Sec. 2012 (2016); Nat'l Insts. of Health, Notice of Changes to NIH Policy for Issuing Certificates of Confidentiality (NOT-OD-17-109), Sept. 7, 2017, available at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-109.html.

139. Nat'l Insts. of Health, Notice of Changes to NIH Policy for Issuing Certificates of Confidentiality (NOT-OD-17-109), Sept. 7, 2017, available at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-109.html.

140. Nat'l Insts. Of Health, Certificates of Confidentiality (CoC) Kiosk, Frequently Asked Questions, https://humansubjects.nih.gov/coc/faqs.

141. National Institutes of Health. NIH Announces Statement on Certificates of Confidentiality. Mar 15, 2002. [March 4, 2014]. Notice NOTOD-02-037. http://grants1.nih.gov/grants/guide/notice-files/NOT-OD-02-037.html.

142. Information about obtaining a certificate of confidentiality is available at the "Certificates of Confidentiality Kiosk" on the National Institutes of Health Web site. [March 4, 2014]. http://grants.nih.gov/grants/policy/coc/index.htm.

143. National Institutes of Health; Office of Extramural Research. Certificates of Confidentiality: Background Information. Febrary 14, 2006. [March 4, 2014]. http://grants.nih.gov/grants/policy/coc/background.htm.

144. 42 USCS 290dd-2 and 290ee-3; 42 CFR Part 2.

145. 42 CFR 2.52(a).

146. In re Philip Morris, 706 So. 2d 665 (La.App. 4 Cir. Jan. 28, 1998) holding that raw data from research on tobacco use is protected under Louisiana statutes that govern the confidentiality of public health data).

147. See, for example, Wis Stat. 146.38.

148. See supra note 55.

149. 45 CFR 164.512(i)(2)(ii).

150. 45 CFR 164.512(i)(2).

151. 45 CFR 164.512(i)(2)(iii) and (iv).

152. 45 CFR 46.116(f)(3).

153. Institutional Review Board Waiver or Alteration of Informed Consent for Minimal Risk Clinical Investigations, 83 Fed. Reg. 57,378 (Nov. 15, 2018); Institutional Review Board Waiver or Alteration of Informed Consent for Minimal Risk Clinical Investigations; Reopening of Comment Period, 84 Fed. Reg. 5968 (Feb. 25, 2019) (reopening comment period to allow for submission of comments by March 7, 2019).

154. U.S. Food & Drug Admin., IRB Waiver or Alteration of Informed Consent for Clinical Investigations Involving No More Than Minimal Risk to Human Subjects at 4 (July 2017), available at https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM566948.pdf.

155. U.S. Food & Drug Admin., IRB Waiver or Alteration of Informed Consent for Clinical Investigations Involving No More Than Minimal Risk to Human Subjects at 4 (July 2017), available at https://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM566948.pdf.

156. Jennifer Kulynych & Henry T. Greely, Clinical Genomics, Big Data, and Electronic Medical Records: Reconciling Patient Rights with Research When Privacy and Science Collide, J. Law & Biosciences, April 2017, at 127, available at https://www.n\cbi.nlm.nih.gov/pmc/articles/PMC5570692/pdf/lsw061.pdf.

157. The Common Rule and FDA regulations define “minimal risk” as where the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests. 45 CFR 46.102(i); 21 CFR 50.3(k).

158. 45 CFR 46.117(c)(1).

159. 45 CFR 46.117(c)(2); 21 CFR 56.109(c).

160. 45 CFR 46.117(c); 21 CFR 56.109(d).

161. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington, D.C.: National Academies Press; 1999.

162. 42 CFR § 3.402.

163. Nass SJ, Levit LA, Gostin LO, Committee on Health Research and the Privacy of Health Information, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. National Academies Press; The HIPAA Privacy Rule; Institute of Medicine; [March 4, 2014]. http://www.nap.edu/catalog/12458.html.

164. Nass SJ, Levit LA, Gostin LO, Committee on Health Research and the Privacy of Health Information, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research.National Academies Press; The HIPAA Privacy Rule; Institute of Medicine; [March 4, 2014]. p. 26. http://www.nap.edu/catalog/12458.html.

165. 45 CFR § 160.103, definition of health information; 78 Fed. Reg. at 5658-61.

166. 74 FR 42740 (August 24, 2009); 78 Fed. Reg. at 5638-58 (January 25, 2013); 45 CFR 164.400-164.414.

167. But see supra note 25.

168. Center for International Blood & Marrow Transplant Research. [March 4, 2014]. http://www.cibmtr.org/pages/index.aspx.

169. 45 CFR 164.528(b)(4); Jennifer Kulynych & Henry T. Greely, Clinical Genomics, Big Data, and Electronic Medical Records: Reconciling Patient Rights with Research When Privacy and Science Collide, J. Law & Biosciences, April 2017, at 33, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570692/pdf/lsw061.pdf. The Privacy Rule creates a legal right for patients to receive, upon request, an accounting of certain disclosures of their PHI that are made by health care providers, insurance plans, and their business associates. 45 CFR 164.528. The accounting must include disclosures that occur with a waiver of authorization approved by a privacy board or IRB. The Privacy Rule specifies the information that an accounting should contain and requires it to cover a six-year period or any requested shorter period of time. 45 CFR 164.518(b)(1). The content of an accounting may be more abbreviated if, during the period covered by an accounting, the entity made disclosures for a particular research purpose for 50 or more individuals. 45 CFR 164.518(b)(4).

170. 45 CFR 46.116(d)(4).

171. Compare Ga. Code Ann. § 31-33-3 (2008) and Ind. Code Ann. § 16-39-5-3 (2016) with N.H. Rev. Stat. Ann. § 151:21 (2015)

172. Joyce C, Patry W, Leaffer M, et al. Copyright Law. 3rd ed. New York and San Francisco: Matthew Bender & Co., Inc; 1994. Introduction: the landscape of copyright; pp. 1–41. (reprinted 1997):chapter 1.

173. 17 U.S.C. §§ 103; Feist Publications, Inc. v. Rural Telephone Service, Co., Inc., 499 U.S. 340, 347-349 (1991).

174. 45 CFR 164.524 - .528; 78 FR 5566 at 5606 (January 25, 2013).

175. 45 CFR 164.501.

Chapter 8. Informed Consent

1. Introduction

This chapter identifies the best practices for obtaining informed consent and permission for registry participation. It builds on some of the general ethical and legal principles discussed in Chapter 7, focusing specifically on the application of the regulations governing human subjects research and the requirements of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

The purpose of this chapter is to provide an ethical framework for obtaining informed consent and permission for registry participation and to distinguish registries from clinical research protocols with respect to these issues. It is not designed to provide specific legal guidance, nor can it substitute for Institutional Review Board (IRB) review. Moreover, legal discussion is limited to U.S. law, and more specifically, to federal as opposed to state statutes and regulations. Some states have guidelines governing the conduct of research involving human subjects or statutes addressing privacy, and an exploration of either area is beyond the scope of this chapter. Likewise, analysis of the relevant international standards and laws is left to others.

Case Examples 16, 17, and 18 provide descriptions of how different registries have approached informed consent.

2. Distinguishing Research and Other Activities

The purpose of this handbook is to provide guidance for registries used to evaluate patient outcomes. The focus of this chapter is on informed consent and authorization issues that arise in registries used for research. It is important to note that registries are increasingly being used for research purposes even when initially developed for clinical or other purposes, and thus it is suggested that in all cases, consideration should be given to the informed consent issues, as well as HIPAA privacy requirements, discussed in this chapter. The HIPAA Privacy Rule governs the use and disclosure of most individually identifiable health information (called protected health information or “PHI”) held by covered entities (health plans, health care clearinghouses, and most healthcare providers).

The key regulatory framework is found in the federal research regulations referred to as the “Common Rule.” As discussed in Chapter 7, the Common Rule applies to research involving human subjects conducted or supported by the 20 Federal departments and agencies, including the U.S. Department of Health and Human Services (HHS), that intend to follow the revised Common Rule. Significant amendments were made to the Common Rule in 2017. As of January 21, 2019, institutions are expected to be compliant with the revised Common Rule. There are also relevant regulations promulgated by the U.S. Food and Drug Administration (FDA). The FDA has authority over “all clinical investigations” regulated by the FDA— defined as “any experiment that involves a test article and one or more human subjects.”[1] This oversight may overlap with HHS, but will also extend to privately funded research. The discussion below focuses on the revised Common Rule, specifically referring to the HHS regulations and to a lesser extent FDA regulations; prior versions of this handbook address the old regulations.

Investigators whose research is sponsored by another federal agency should consult the guidelines issued by that agency. While the discussion below is likely to be applicable, there may be additional requirements in specific circumstances.

The HHS regulations apply to federally funded "research involving human subjects,"[2] where "research" is defined as a "systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge" and "human subject" is defined as a living person about whom the investigator obtains either data or biospecimens through intervention or interaction, or "obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens."[3] Identifiable information or biospecimens are those "for which the identity of the subject is or may readily be ascertained by the investigator or associated with the [information or] biospecimen."[4] The regulations require periodic (at least every 4 years) assessments of analytic technologies/techniques to determine what should be considered "identifiable." A list of the technologies will be published in the Federal Register for notice and comment and then maintained on a publicly available website.

Investigations that involve nonliving individuals, or that do not collect data through intervention/interaction with the individual, or do not collect "identifiable" information are not considered research subject to the regulations. Despite the scope of the HHS regulatory language, institutions may choose to apply the frameworks more broadly (sometimes under an "assurance," i.e., an agreement with HHS that the institution will apply the regulations to all research at the institution regardless of funding source). In other contexts FDA regulations will apply since the data from the research is going to be used in support of applications for marketing an FDA regulated product. The FDA regulations regarding informed consent in large part mirror the HHS regulations,[5] but the scope of research oversight of the two agencies is different.

Even when the activity in question meets the definition of research involving human subjects under the HHS regulations, a series of exemptions may apply.[6] There are a few exemptions particularly relevant to registry research, although they may not be used for research covered by the Common Rule Subpart C, which applies to prisoners (unless the research only incidentally includes prisoners).[7] First, consent is not required for "secondary research uses of identifiable private information or identifiable biospecimens, if at least one of the following criteria is met:"

- the identifiable information/biospecimens are publicly available,
- the information is recorded in way that identity is not readily ascertained, the investigator does not contact the subjects, and the investigator will not re-identify the subjects
- the research is regulated under HIPAA (e.g., is considered "healthcare operations" or "public health activities")
- the research is conducted by or on behalf of a federal agency using government generated /collected information.[8]

Second, "storage or maintenance of identifiable private information or identifiable biospecimens for potential secondary research use [is considered exempt] if an IRB conducts a limited IRB review and makes the determinations required by §46.111(a)(8)."[9] And finally, "research

involving the use of identifiable private information or identifiable biospecimens for secondary research use," is exempt from the normal requirements if the following conditions are met:

- Broad consent for the storage, maintenance, and secondary research use of the identifiable private information or identifiable biospecimens was obtained in accordance with §46.116(a)(1) through (4), (a)(6), and (d);
- Documentation of informed consent or waiver of documentation of consent was obtained in accordance with §46.117;
- An IRB conducts a limited IRB review and makes the determination required by §46.111(a)(7) and makes the determination that the research to be conducted is within the scope of the broad consent referenced in paragraph (d)(8)(i) of this section; and (iv) The investigator does not include returning individual research results to subjects as part of the study plan. This provision does not prevent an investigator from abiding by any legal requirements to return individual research results.[10]

Broad consent is discussed in more detail below in section 4.2.

2.1 Registry Research Versus Clinical Research

Differences in who is considered a participant in registry research and clinical research are worth noting, especially since the differences may impact how we think about informed consent. In particular, the use of a control group in a registry setting is often substantively different from the concept of a control group in a clinical research setting. Registry controls may be pulled from a general population—in some cases a population that may not have interacted with health professionals or a health institution. Unlike clinical controls, who may be exposed to placebos (and thus need to consent) or assigned to a standard treatment (and thus will already be involved in the treatment system), registry controls may be identified from an unaffected population. This raises ethical questions about the initial contact with an individual who may have no link to the registry topic area and who may view the contact as an unwelcome intrusion or perhaps even an incorrect indication of problematic health status.[11] Furthermore, since a clinical research trial may involve double-blind procedures, "controls" may agree to participate because of the potential for direct therapeutic benefits or even the indirect therapeutic benefits that come from better attendant care. In other situations, controls may participate because they hope to help others suffering from the same ailments (altruistic reasoning) or perhaps because they seek monetary compensation. In contrast, controls in a registry trial have no similar potential therapeutic (direct or indirect) or monetary benefits. While altruism may play a role in this context, its effects may be problematic. There is a great concern about the potential for selection bias in the creation of a control group for registry trials—as there is also significant concern about the effect of bias in clinical trials. Those who may agree to participate in a registry may be qualitatively different from those who do not agree, and this disparity can threaten the external validity of research findings. Concerns about selection bias will be heightened for diseases with a low prevalence in the general population, since there will be a greater possibility that the bias will affect the data. Developing consent requirements in such a way as to avoid selection bias will be extremely important in this setting.

Questions about adapting the regulatory requirements to research that does not fit the typical clinical model are not unusual. There are two other areas that have raised questions about how the Federal regulations apply and that are particularly relevant to registry evaluations: public health activities and quality improvement/quality assurance (QI/QA).

2.2 Public Health Activities

The HIPAA Privacy Rule expressly permits the disclosure of Protected Health Information (PHI) to a public health authority that is authorized by law to collect or receive such information for the purpose of preventing or controlling disease, injury, or disability, including for activities related to disease, injury, or vital event reporting. Thus, a covered entity may disclose PHI, without individual authorization, for a registry maintained by a public health authority (or by an entity acting under a grant of authority from or contract with such public agency) for authorized public health purposes; such as, for example, immunization registries, state cancer registries, birth and death registries, and general disease reporting (although this last type of reporting is often anonymous). The HIPAA Privacy Rule also allows the disclosure of PHI to a person subject to the jurisdiction of FDA for FDA-regulated–product reporting.

In addition to harmonizing the research regulations with the HIPAA-permitted disclosures, the revised Common Rule states explicitly that certain public health surveillance activities are deemed not within the definition of "research involving human subjects" and therefore not subject to the regulations.[12] This is designed "to allow the public health authority to monitor, assess, or investigate public health signals, onsets of disease outbreaks, or conditions of public health importance" without having to comply with the Common Rule. Even when a public health activity falls within the definition of research under the Common Rule, it may be covered by various exemptions.

The biggest challenge in this context is differentiating between public health activities and public health research; there have been a number of efforts in the past to clarify the two categories.[13] In many cases an activity might involve both traditional public health surveillance and research involving human subjects. According to the Belmont Report, the document on which the original Federal research regulations were based, if any aspect of an activity constitutes "research," then the entire activity should undergo regulatory review. By contrast, the Centers for Disease Control and Prevention (CDC) only consider a public health activity to be research if the primary intent of the activity is to contribute to or generate generalizable knowledge.[14] Local IRB policies in this area have varied; some focus on determining whether the primary intent of a specific activity is to gain generalizable knowledge, other IRBs categorically exclude all traditional public health department activities.

In November 2018 OHRP issued a guidance document addressing public health surveillance under the revised Common Rule, stating clearly that the elements of an activity will be considered separately with respect to the application of the Common Rule. It notes that "if an activity is composed of multiple components, some of which are not public health surveillance, OHRP's view is that only those components that serve to enable a public health authority to carry out one or more public health surveillance objectives should be considered potentially

eligible for the public health surveillance activity exclusion."[15] For research begun before the effective date of the revised regulations, the guidance states that an institution must—

- determine that the activity (or a set of activities) under consideration will transition to be conducted in accordance with the revised Common Rule.
- document (or have the IRB document) the transition decision (including the date of the decision) in accordance with §46.101(l)(4), and retain such documentation for at least 3 years in accordance with §46.115(b).
- apply the carve-outs from the definition of research as described in the revised common rule.

There is no indication of who determines the application of the exclusion in the Common Rule—it might be the public health authority, the organization or individual(s) undertaking the activity, or it could be an IRB. But the OHRP guidance does note that the decisions about applicability should be "thoughtful and deliberate." The determination of whether the activity is exempt rests upon the following three criteria:

- The activity must be a public health surveillance activity (45 CFR 46.102(*l*)(2));
- The activity must be conducted, supported, requested, ordered, required, or authorized by a public health authority (45 CFR 46.102(k) and 46.102(*l*)(2)); and
- The activity must be limited to that necessary to allow a public health authority to identify, monitor, assess, or investigate potential public health signals, onsets of disease outbreaks, or conditions of public health importance (including trends, signals, risk factors, patterns in diseases, or increases in injuries from using consumer products) (45 CFR 46.102(*l*)(2)).[15]

Two key factors in determining whether an activity is part of the public health surveillance carve-out under the revised regulations include evaluating whether the "collection, management, analysis, and interpretation of surveillance information or biospecimens is designed to inform a public health authority, and generally is followed by public health action or by the dissemination of information to public health programs and others to stimulate public health action."[15] The November guidance provides descriptions of these terms as used in the Common Rule, as well as examples of public health surveillance activities.

2.3 Quality Improvement/Quality Assurance Activities

As with certain public health activities, the HIPAA Privacy Rule provides an explicit permission to use or disclose PHI for "healthcare operations," which are defined as certain activities of a covered entity, including "conducting quality assessment and improvement activities…, provided that the obtaining of generalizable knowledge is not the primary purpose of any studies resulting from such activities."[16] Individual authorization for use or disclosure of PHI in this context is not required, but a covered entity can choose to obtain such individual authorization.

The new federal research regulations do not use the terms QI or QA, but specifically exempt from review research conducted under the HIPAA regulations for the purposes of "healthcare operations."[17] Thus many QI/QA activities will: (a) not meet the regulatory definitions of

"research," (b) not involve "human subjects," (c) fall under this or another delineated exemption, or (d) not be supported by HHS, involve an FDA regulated product, or otherwise be covered by an assurance of compliance. Some local IRBs appear to consider all QI/QA activities outside the scope of the regulations, as they would public health activities.[18]

In considering whether an activity fits into the "healthcare operations" exemption, it is important to note that the application of the research regulations does not rest on whether or not a procedure is considered "standard" or part of the "standard of care." Nor is lack of intent to publish the results a key factor. Instead, investigators should evaluate whether the activity in question meets the definition of research involving human subjects such that it falls within the regulatory framework, and then consider whether an exemption applies. This latter determination is usually made by an IRB or within the IRB office, not by the investigator.

Registries developed within an institution to implement a practice to improve the quality of patient care or to collect data regarding the implementation of such a practice would be exempt, as would registries designed to collect provider performance data for clinical, practical, or administrative purposes. Registries that involve existing data which is not individually identifiable also is not considered research involving human subjects as defined by the Common Rule. However, a QI/QA project that involves a clinical intervention (whether or not part of the standard of care) for purposes of gathering scientific evidence of efficacy (i.e., a systematic investigation designed to contribute to generalizable knowledge) may well be governed by the regulations, although another specific exemption may apply.[19] Even if the regulations apply, waivers or alterations to the consent process may be approved as noted below.

3. Growth of Database Research

3.1 Electronic Health Records and other Health Information Databases

The development of large-scale data registries raises a variety of regulatory questions, and this is nowhere more evident than in the discussions about electronic health records (EHRs). Not only are large amounts of information being collected, there is also a large increase in available biobanks, which are defined as facilities that store biological material (e.g., serum, genomic material, pathology specimens) from humans. These issues are explored in detail in Chapter 7. While the revised Common Rule directly addresses the questions of regulatory oversight for databank and biobank research (and informed consent in this context), the development of the final rule was subject to considerable debate.

Part of the problem is that there are few, if any, efforts to obtain individual consent for the creation of an EHR (or, for that matter, the creation of any health record) since consent is not required under the Common Rule for these activities; these are clinical documents which do not constitute research. Yet, these databases have enormous research potential. For example, Kaiser Permanente, a leader in the use of health information technology, created and maintains one of largest private-sector EHR systems, collecting health information from more than 11.8 million Kaiser members nationwide. As another example, there are efforts to develop (sometimes via state legislation) multi-payer claims registries to support comparative effectiveness research (CER). One primary concern that led to the recent changes in the Common Rule was that the

application of traditional consent models for secondary use of these databanks for research would prove inefficient (perhaps even preventing crucial research from occurring), or result in selection bias, impacting the usefulness of downstream analyses.

As the development of EHRs, claims registries, health information exchanges, and linkages between innumerable health databases moves forward, keeping records private becomes more difficult to manage. Personal health information may be accessed and shared in ways patients never imagined. Although studies consistently indicate that Americans are generally supportive of EHRs and even secondary uses of the data, they want to be informed about how and to what extent their information will be shared and disclosed to others.[20-23] At the same time, it is not clear from the empirical evidence that patients want full consent protections in this context. In fact, there are studies showing that patients are willing to forego full consent when it would undermine research. The extent of comfort with research absent full consent may vary. One study, for example, found that patients were more likely to be comfortable with the research uses of their EHR information when they were asked about the use by a specific entity (e.g., universities, hospitals, or disease foundations) rather than when asked in the abstract, and that they fully supported public health uses of their data.[24] Public education about the scope of research uses may alleviate remaining patient concerns about the use of data without full consent.[25] Similarly, addressing underlying fears about unauthorized access to identifiable data or discriminatory uses of the information can also be helpful in increasing support for this type of research.

3.2 Changing Ethics in the Context of Database Research

Perhaps the most challenging part of the shift to large database research and the current regulatory structure is the potential reframing of the underlying ethical issues. The pre-2018 Common Rule was based on the Belmont Report, which focused on the traditional clinical research context. The continued expansion of electronic data repository research and the potential for large-scale observational studies to replace some clinical trial data require consideration of whether the approaches used thus far should be altered. As the preamble to the final version of the revised Common Rule states:

> Since the Common Rule was promulgated, the volume and landscape of research involving human subjects have changed considerably. Research with human subjects has grown in scale and become more diverse. Examples of developments include: an expansion in the number and types of clinical trials, as well as observational studies and cohort studies; a diversification of the types of social and behavioral research being used in human subjects research; increased use of sophisticated analytic techniques to study human biospecimens; and the growing use of electronic health data and other digital records to enable very large datasets to be rapidly analyzed and combined in novel ways. Yet these developments have not been accompanied by major change in the human subjects research oversight system, which has remained largely unaltered over the past two decades.

While the revised Common Rule addresses some of these landscape changes, there are additional issues which will need to be considered in the future. Professor Barbara Evans identified three "novel challenges in applying familiar ethical frameworks" to databank research.[26] The first is

the possibility that with the shorter time period between research results and clinical application, the historical bifurcation between research and treatment (or even public health practice and research) may be incorrect. Perhaps IRBs will need to consider both the potential direct medical benefits of an observational study, and potential participant health risks such as negative insurance coverage determinations or changes in physician prescribing patterns. Or perhaps it is time to develop a robust concept of group benefit (and group risk) in contrast to the regulatory narrow focus on individuals. Second, the creation of these massive data banks that span numerous states (and sometimes countries) raises issues about whether the "local context review" that forms the basis for the IRB system continues to be relevant. It is worth noting that the revised Common Rule allows a single IRB of record for multi-site studies, rather than requiring multiple reviews, and NIH now requires this. How this approach will change the concept and importance of local review, which formed the basis for the initial federal regulations, remains to be seen. Finally, these new types of research raise questions about the meaning of vulnerability and susceptibility to harm, and who should be identified as a "vulnerable" population in need of additional protections. It may be that the groups traditionally considered vulnerable in the clinical research context are not especially vulnerable in this context. Conversely, there may be groups particularly vulnerable to re-identification, or for whom re-identification poses unique risks of psychosocial or economic harms, but which would not usually be considered vulnerable in clinical research. In fact, the need to understand potential group harms highlights the limitations of the traditional ethical framework that assumes the focus should be on the individual. More work is needed to consider the ethical framework that should guide large-scale database and observational studies, but such exploration is beyond the scope of this chapter.

4. Regulatory Consent Requirements

While a number of issues remain unanswered in this area, there is some clear guidance for registries that fall under the federal research regulations. As noted previously, there are two primary sets of federal regulations governing the conduct of research involving human subjects. HHS regulates research supported by federal money or covered under an institutional "assurance of compliance" (see Chapter 7). The FDA regulates research that will be used to support an FDA-regulated product. FDA has not (yet) updated its consent requirement to match the revised Common Rule so there are some differences. Both sets of regulations largely have the same consent requirements; relevant differences are indicated below. The HIPAA Privacy Rule also contains individual authorization requirements for uses and disclosures of individually identifiable health information for research purposes. Each of these federal regulatory areas will be discussed in turn.

4.1 Consent Requirements

Participants in research studies, or their legal representatives, must provide legally effective informed consent.[27] Information that a reasonable person would want should be provided in understandable language,[28] an opportunity for discussion offered, and the resulting consent must not be coerced or the product of undue influence.[29] Moreover, except for broad consent (discussed in 4.2 below), "the informed consent must begin with a concise and focused presentation of the key information that is most likely to assist… in understand the reasons why

one might or might not want to participate."[30] Finally, no consent may include exculpatory language which waives or appears to waive rights.[31] The following nine basic elements of information must be provided to research participants:[32]

1. A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental.
2. A description of any reasonably foreseeable risks or discomforts to the subject.
3. A description of any benefits to the subject or to others which may reasonably be expected from the research.
4. A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.
5. A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained.
6. For research involving more than minimal risk, an explanation as to whether there is any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained.
7. An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject.
8. A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.
9. One of the following statements about any research that involves the collection of identifiable private information or identifiable biospecimens: (i) A statement that identifiers might be removed from the identifiable private information or identifiable biospecimens and that, after such removal, the information or biospecimens could be used for future research studies or distributed to another investigator for future research studies without additional informed consent from the subject or the legally authorized representative, if this might be a possibility; or (ii) A statement that the subject's information or biospecimens collected as part of the research, even if identifiers are removed, will not be used or distributed for future research studies.

In addition, the FDA requires that informed consent forms for applicable clinical trials include a statement that the trial information will be entered into the National Institutes of Health (NIH) clinical trial registry.[33]

The Common Rule also requires, where appropriate, additional elements of informed consent, including the following:[34]

1. A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) which are currently unforeseeable.
2. Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent.
3. Any additional costs to the subject that may result from participation in the research.
4. The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject.
5. A statement that significant new findings developed during the course of research which may related to the subject's willingness to continue participation will be provided to the subject.
6. The approximate number of subjects involved in the study.
7. A statement that the subject's biospecimens (even if identifiers are removed) may be used for commercial profit and whether the subject will or will not share in this commercial profit;
8. A statement regarding whether clinically relevant research results, including individual research results, will be disclosed to subjects, and if so, under what conditions; and
9. For research involving biospecimens, whether the research will (if known) or might include whole genome sequencing (i.e., sequencing of a human germline or somatic specimen with the intent to generate the genome or exome sequence of that specimen).

HHS regulations also set forth the elements of a broad consent for the storage, maintenance, and secondary research use of identifiable private information or identifiable biospecimens. The details of these requirements are discussed below under a separate heading. FDA has not adopted this new language.

HHS regulations allow an IRB to approve a waiver or alteration of the consent requirements for minimal risk research where: the waiver or alteration will not affect the rights of the subjects, the research cannot practicably be carried out without the waiver, for identifiable private information or biospecimens that the research could not be practicably carried out without them remaining identifiable, and, when appropriate, subjects will be provided information after participation.[35] Waivers of informed consent are also allowed when investigators are merely gathering information to screen, recruit or determine eligibility of prospective subjects if the information is gathered either through oral/written communication or by accessing records or stored sample.[36] HHS also allows waivers of the documentation requirement for informed consent (see more below under 4.3), although FDA regulations do not. In all of cases of waiver, the participant may still be provided a written summary of the "key information essential to decision making" even if the documentation requirement is waived.

The FDA regulations explicitly allow for waivers in life-threatening situations and allow Presidential waivers for some military research.[37] Both HHS and FDA regulations allow for waiver of consent requirements for research conducted in specific types of emergency situations.[38,39]

4.2 Consent for Databank and Biobank Research

The revised Common Rule directly addresses a number of the issues regarding biobank and databank research, although it leaves some questions unanswered. One category of such research is considered exempt—"secondary research uses of identifiable private information or identifiable biospecimens" —if at least one of the following criteria is met:

- the identifiable information/biospecimens are publicly available,
- the information is recorded in way that identity is not readily ascertained, the investigator does not contact the subjects, and the investigator will not re-identify the subjects
- the research is regulated under HIPAA (e.g., "healthcare operations" or "public health activities")
- the research is conducted by or on behalf of a federal agency using government generated /collected information.

Another category of research on identifiable private information/biospecimens is allowed with "limited" IRB review—"storage or maintenance of identifiable private information or identifiable biospecimens for potential secondary research." Finally, "broad consent" (rather than traditional specific consent), may be used to provide authorization to store, maintain and conduct any secondary research on identifiable private information or identifiable biospecimens which were collected either as part of another research project, or collected for non-research purposes, without having to get specific subsequent informed consent for each future research endeavor.

The Secretary's Advisory Committee on Human Research Protection (SACHRP) advises the Secretary of HHS on issues related to the protection of human subjects. SACHRP created a useful template for the new broad consent pathway, allowing future uses of identifiable information and identifiable biospecimens, which complies with the revised Common Rule, incorporating both the general consent requirements and the specific requirements for the newly allowed "broad consent."[21] According to the template, investigators wishing to use the new broad consent pathway should create a broad consent document that includes:[40]

- A description of any reasonably foreseeable risks or discomforts to the subject;
- A description of any benefits to the subject or to others that may reasonably be expected from the research;
- A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained;
- A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled;
- When appropriate, a statement that the subject's biospecimens (even if identifiers are removed) may be used for commercial profit and whether the subject will or will not share in this commercial profit;

- When appropriate, for research involving biospecimens, whether the research will (if known) or might include whole genome sequencing (i.e., sequencing of a human germline or somatic specimen with the intent to generate the genome or exome sequence of that specimen);
- A general description of the types of research that may be conducted with the identifiable private information or identifiable biospecimens. This description must include sufficient information such that a reasonable person would expect that the Broad Consent would permit the types of research conducted;
- A description of the identifiable private information or identifiable biospecimens that might be used in research, whether sharing of identifiable private information or identifiable biospecimens might occur, and the types of institutions or researchers that might conduct research with the identifiable private information or identifiable biospecimens;
- A description of the period of time that the identifiable private information or identifiable biospecimens may be stored and maintained (which period of time could be indefinite), and a description of the period of time that the identifiable private information or identifiable biospecimens may be used for research purposes (which period of time could be indefinite);
- Unless the subject or legally authorized representative will be provided details about specific research studies, a statement that they will not be informed of the details of any specific research studies that might be conducted using the subject's identifiable private information or identifiable biospecimens, including the purposes of the research, and that they might have chosen not to consent to some of those specific research studies;
- Unless it is known that clinically relevant research results, including individual research results, will be disclosed to the subject in all circumstances, a statement that such results may not be disclosed to the subject; and
- An explanation of whom to contact for answers to questions about the subject's rights and about storage and use of the subject's identifiable private information or identifiable biospecimens, and whom to contact in the event of a research-related harm.[41]

If a broad consent is used, required elements of the broad consent cannot be waived by the IRB. Moreover, if a broad approach consent was used and the individual refused consent, the IRB may not waive elements of specific consent at a subsequent point. The regulations do not indicate what to do when an individual is asked for a broad consent and fails to respond. SACHRP suggests that only an "unambiguous written" refusal should be viewed as a refusal in this context and that the broad consent form clearly state the implications of a failure to respond to a broad consent request.[41] Part 5 of this chapter addresses some of these opt-in and opt-out issues.

SACHRP notes that in cases where a potential participant is asked to provide a broad consent and refuses, the entity gathering the data should (a) remove the data/sample from the database, or (b) fully de-identify the particular data/sample and continue to include it in the database, or (c) clearly mark the data/sample as not to be used for research and use it only for those activities which do not fall within the regulatory framework (e.g., public health or QI/QA activities). Data or samples which fall into this latter category could also, in theory, be marked and then the

individuals approached for specific consent for a particular research study at a later date (unless the individual had stated that s/he does not want to participate in any research, nor be approached later).

Even when a broad consent approach was not used to create the database, in certain circumstances an IRB can waive specific consent requirements for secondary research on identifiable information or specimens as long as the individual did not specifically refuse such a use. These waivers, however, will not be granted unless the IRB finds that it would be impracticable to use de-identified data. Finally, secondary research may be conducted on databases or biobanks without consent if the information or specimens are de-identified, which would take them out of the scope of research subject to the Common Rule. Note, however, that de-identification and aggregate reporting alone may not completely conceal identity.[42,43] For example, there is a considerable push to make de-identified, aggregate-level data from Genome Wide Association Studies (GWAS) publicly available in large repositories so that the data can be combined with other studies for more powerful analysis. An individual potentially could be re-identified by assessing the probability that an individual or relative participated in a GWAS through composite statistics across cohorts (such as allele frequency or genotype counts). As noted previously, the revised Common Rule requires periodic assessments of what constitutes identifiability.

While permitting broad consent is a significant boon for registry research, it is important to stress that ensuring compliance with the Common Rule will necessitate electronic tracking mechanisms for individual consent (linked with the identifiable data or biospecimen) to ensure that the information is only used in accord with the individual's preferences. Without tracking, the impact of broad consent will be considerably lessened. In cases where tracking is not possible, traditional consent may still be used, but with all the prior caveats and concerns about inefficiencies and skewed participation.

4.3 Format and Documentation of Consent

The revised Common Rule requires that all consent forms now begin with a concise statement of the key information needed to help the prospective participant (or their legally authorized representative) understand the reasons why s/he should or should not participate. The information has to be "presented in a way that facilitates comprehension." Broad consents are not required to meet this requirement.

Unlike treatment consents, research consents are usually written and the consent form functions both as documentation of the consent process, and in some cases as an aspect of the consent itself (since in long form the document contains all of the necessary consent disclosures and participants may be given the form to read as part of the consent process). The regulations allow an IRB to waive the written documentation requirement in whole or in part when (1) the only record linking the subject and the research would be the consent document and the principal risk of the study would be potential harm resulting from a breach of confidentiality, (2) the research involves no more than minimal risk of harm and involves no procedures for which written consent is not normally obtained in a clinical context, or (3) when the research involves no more than minimal risk of harm and the subjects or legally authorized representatives are members of

a distinct cultural group or community in which signing forms is not the norm, and there is an appropriate alternative mechanism for documentation.[44] Waivers of documentation are allowed for both specific and broad consents. One situation in which a waiver of written documentation (i.e. the participant's signature) for a broad consent might be used would be when the information in question is gathered through a phone survey. There may also be broad consent documentation waivers requested when involving a distinct cultural group in which signing forms is not the norm. In all of cases of waiver, the participant may still be provided a written summary of the "key information essential to decision making" even if the documentation requirement is waived. Note that the FDA regulations do not allow a waiver of documentation requirements.

Under certain conditions, an IRB may approve the use of a short-form written consent.[45] In these cases, oral presentation of informed consent information is accompanied by a short-form written consent document (stating that the elements of consent have been presented orally) and a written summary of what has been presented orally. A witness to the oral presentation is required, and the participant (or representative) must be given copies of the short form document and the summary. The participant must sign the short form document and the witness must sign both the short form and the summary.

E-consents may be considered written documentation under either set of regulations and are within the scope of the IRBs' power to authorize as an alternate way of fulfilling written documentation requirements. Both HHS and FDA specifically allow electronic signatures on research consent documents, provided they are legally valid in the specific jurisdiction.[46,47]

Finally, the new regulations require that for federally funded or conducted research a copy of an IRB-approved consent form must be published on a publicly available federal website after the study is closed to recruitment and no later than 60 days after the last study visit by any subject, as required by the protocol.[48] Even if a form has been modified during the course of the trial (see 4.4 below), or variations allowed for different kinds of participants (e.g., adults and children) or for different sites of a multisite trial, only one version is required to be posted. There is a mechanism to allow redaction of information that the sponsoring agency feels should not be posted on a public website. At this time, the consent form posting requirement can be satisfied by posting the form on ClinicalTrials.gov or in a docket folder on Regulations.gov.[49]

4.4 Informed Consent Form Revisions and Re-consent

Re-consent for a registry should be rare. It may be needed if a broad consent was not used initially and there are significant changes to the protocol, risks, or other aspects of information which should be shared with participants and consent for those changes is necessary. Such protocol revisions must be reviewed and approved by an IRB prior to the revised consent being utilized, except when necessary to eliminate apparent immediate hazards to subjects. Re-consent may also be necessary if the participants were below the age of consent when initially enrolled but reach the age of majority when the registry is still active.[50] The decision to change the informed consent form and subsequently obtain the re-consent of participants needs to be carefully considered due to possible challenges in obtaining the re-consent. Challenges may be particularly evident for registries that have been in place for several years. These difficulties are

part of what prompted the interest in broad general initial consents. In situations where the initial consent does not cover the change, registries may seek IRB waiver of re-consent requirements.

For studies in which re-consent is sought, registry developers should consider the potential effects of selection bias and the implications for external validity. Re-consented participants may be systematically different from non-re-consented participants. For example, participants that have not re-consented may have died or been lost to followup for health-related reasons, leaving an overall healthier group of participants. Additionally, even among those who can be contacted, individuals who agree to continue participation may be different from those who refuse to provide consent. As a result, one important requirement for studies that undertake re-consent may be to evaluate characteristics of the original study population, as compared with the subset of patients that do re-consent, and consider the implications for research outcomes. The evaluation of whether re-consents are more common for particular populations should be done for any analyses that have comparative arms.

Minor changes to a consent document do not necessitate re-consent. Re-consent may be appropriate, however, where the terms of the study or the background preconditions have changed. In some long-term studies, re-informing participants, but not re-consent, may be the best option. Even where re-consent is indicated, IRBs may waive requirements. Alternatively, data collection, sharing, and reporting for participants who cannot be re-consented could be maintained in accordance with the terms of the original consent. In those situations in which a re-consent process is implemented, participants should be told the reasons for the recontact and given a summary of consent form changes. Additionally, as with the original consent, documentation of the re-consent should be maintained as required by the registry, the IRB, and any relevant regulations. Going forward, the use of broad consent should address many of these problems.

4.5 Remaining Challenges for Registries

Even with the revised Common Rule regulations, which explicitly address database and biobank research, some of the regulatory requirements appear better suited to traditional clinical research trials than to registries. For example, of the nine basic elements listed earlier, requirements 4 (alternatives) and 6 (compensation/injury) are crafted to address issues raised in traditional clinical trials, rather than registries. Other elements have aspects that clearly encompass registry research (such as basic elements 1, 2, 7 and specifically 9), but other parts seem less applicable, since registries will not involve “experimental procedures” that must be identified, entail no physical “discomforts to the subject,” and do not pose a risk of (physical) “research-related injury.”

Other requirements may pose unique challenges for registries, such as basic element 8, which requires that subjects be informed about a right to withdraw. While registry participants may refuse to provide additional information about their medical status or care, withdrawing from a registry may undermine the data collection. In situations where the data have been anonymized, withdrawal will likely prove impossible. In many such cases, registry informed consents may contain language notifying subjects that in the event of withdrawal, data that was collected prior to the withdrawal may continue to be used and disclosed according to the consent in order to

preserve the scientific integrity of the registry. However, even where data have not been anonymized, some argue that the registry must retain all records to be maintain validity. The FDA explicitly requires the retention of identifiable data even after a subject withdraws from a study. HHS permits the retention of such data, but also permits the investigator to omit or destroy the data if retention is not required by FDA regulations for study integrity. OHRP suggests that IRBs provide guidance on documentation of participant withdrawal. Moreover, the OHRP guidance dated September 21, 2010, on this issue clarifies that once a subject withdraws, the investigator must stop interacting with the subject to obtain data, and stop collecting identifiable private information from other sources unless the subject specifically provides consent to the continued data collection.[51]

The best approach is to make sure that the informed consent form addresses what will happen after withdrawal (in additional to notifying participating of their right to withdraw). Withdrawal of consent after a broad consent might only affect future studies, but not current or past ones, or it might mean that the data/specimen will still be used but stripped of identifiers. This is a less than ideal solution for both sides—the patient will not be able to continue to control the use of the data and without identifiers research uses may be more limited. Potential participants will also need to be told that when a broad consent is used it may be difficult to track down all distributions of their data and that will affect the ability to "withdraw" their data from future studies.

4.5 HIPAA

The HIPAA Privacy Rule may apply to uses or disclosures of health information into or out of a registry, or the use of such information to create a registry, or both. In addition, because the HIPAA Privacy Rule protects individually identifiable health information held by covered entities, the Privacy Rule requirements may apply even if the human subjects research regulations do not. Moreover, the FDA requires that all of the qualified entities with which it contracts to provide analyses of drug safety data follow the minimal requirements of the Privacy Rule, regardless of whether they are a HIPAA-covered entity.[52] Chapter 7 describes the general Privacy Rule framework in this context and the specifics of coverage. The Privacy Rule requires a covered entity to obtain written authorization for the use or disclosure of an individual's PHI for research purposes unless the use or disclosure is permitted by another provision of the Rule (e.g., where an IRB waiver of authorization applies). It is worth emphasizing that a subject's informed consent to participate in research can be combined with a HIPAA authorization in one document, thus eliminating unnecessary pages in what are often extremely long consent documents and potentially reducing the confusion of participants. There are six core elements and three required statements for a valid HIPAA authorization:[53]

Core Elements

- A description of the PHI to be used or disclosed, identifying the information in a specific and meaningful manner.
- The names or other specific identification of the person or persons (or class of persons) authorized to make the requested use or disclosure.

- The names or other specific identification of the person or persons (or class of persons) to whom the covered entity may make the requested use or disclosure.
- A description of each purpose of the requested use or disclosure.
- Authorization expiration date or expiration event that relates to the individual or to the purpose of the use or disclosure. "End of the research study" or "none" are permissible for research, including for the creation and maintenance of a research database or repository.
- Signature of the individual and date. If the individual's legally authorized representative signs the Authorization, a description of the representative's authority to act for the individual must also be provided.

Required Statements

- A statement of the individual's right to revoke the Authorization in writing, and either: (1) a description of how to do so, and the exceptions to the right to revoke authorization, or (2) reference to the corresponding section of the covered entity's notice of privacy practices.
- Whether treatment, payment, enrollment, or eligibility for benefits can be conditioned on the individual's signing the Authorization, including research-related treatment, and consequences of refusing to sign the Authorization, if applicable.
- A statement of the potential for the PHI to be redisclosed by the recipient and no longer protected by the Privacy Rule. This may be a general statement that the Privacy Rule may no longer protect health information disclosed to the recipient.

Authorization is not needed for activities that are "preparatory to research," which may include scanning a patient database to determine feasibility for creating a registry. Before allowing an investigator access to PHI for such purposes, however, the covered entity must obtain from the researcher representations that: (1) the use or disclosure of PHI is sought solely for purposes preparatory to research, (2) no PHI will be removed from the covered entity during the review, and (3) access to the PHI is necessary for the research purposes.[54] These preparatory activities may aid investigators in the identification of potential research participants. Subsequent contact of potential research participants for purposes of obtaining authorization for the use or disclosure of the individual's PHI may be permitted under the Privacy Rule in a variety of ways depending on the relationship between the investigator and the covered entity. An investigator who is a workforce member of the covered entity is permitted to contact potential participants directly or through another person at the covered entity, such as a treating provider, to obtain authorization. Alternatively, a covered entity is permitted to hire a business associate—who may be an investigator—to contact patients to obtain authorization on behalf of the covered entity. Finally, a covered entity is permitted to provide contact information of potential research subjects to an investigator who is not part of the covered entity or a business associate, if the covered entity obtains documentation that an IRB or privacy board has waived the authorization requirement for the disclosure.

Additionally, uses or disclosures of a decedent's PHI to a research registry or from a registry for research purposes do not require an authorization (as long as certain representations are provided

to the covered entity that is providing the information).[55] This is consistent with the definition of "human subjects" under the research regulations which include only living individuals. Similarly, authorizations are also not required for uses or disclosures of de-identified datasets, provided the information has been de-identified in accordance with the Privacy Rule.[56] Nor are authorizations required for uses or disclosures of "limited datasets," as defined by the Rule (so long as a data use agreement is in place with the recipient of the limited dataset).[57] See Chapter 7.

In addition, an IRB or privacy board may waive or alter aspects of the HIPAA authorization requirements. Like the requirements for a waiver or alteration under the human subjects research regulations described above, these are limited to situations in which the research (including access to the PHI) could not practicably be carried out without the waiver or alteration, and the use or disclosure information involves no more than minimal risk to privacy because there is: (a) an adequate plan to protect the identifiers from improper use or disclosure; (b) an adequate plan to destroy identifiers if possible; and (c) adequate written assurances that the PHI will not be reused or disclosed except as required by law, as needed for research oversight, or for other research in a way permitted by the Privacy Rule.[58]

Finally, if a subject was enrolled in a research protocol prior to the compliance date of the Privacy Rule (for most covered entities, April 14, 2003) and pursuant to a valid informed consent, an authorization may not be required unless after the compliance date another informed consent is sought from the subject.[59] This provision may be especially relevant to registries that were created prior to the application of the Privacy Rule.

The HIPAA Privacy Rule also speaks to the issue of withdrawal from a registry. The Privacy Rule explicitly gives individuals the right to revoke their authorization for the use and disclosure of protected health information (the revocation must be in writing), except to the extent that a covered entity has already relied on the authorization. HHS guidance on the application of the Privacy Rule to research makes it clear that a covered entity that has disclosed PHI for research in reliance on an authorization is not required to retrieve information it disclosed prior to receiving the revocation, and may also continue to use and disclose PHI already obtained to the extent necessary to preserve the integrity of the study (e.g., as necessary to account for the subject's withdrawal). As noted above, FDA states that the data gathered as part of research under their regulatory authority is necessary and must be retained; but even for those registries outside the scope of FDA oversight, HIPAA permits the continued use of data as necessary to protect the integrity of the research.

In the past, concerns had been raised that the HIPAA authorization requirements were not sufficiently harmonized with the human subjects research informed consent requirements.[60,61] This was due to the fact that, while a HIPAA authorization could be combined with a research informed consent document (and elements already present in the research consent were not required to be repeated in the authorization), there were some situations in which an additional separate authorization may have been necessary for a separate research activity or future research activity. In January 2013, HHS published changes to the HIPAA Rules that addressed these issues. In particular, the Privacy Rule now permits the use of a compound authorization to authorize the use or disclosure of an individual's PHI for a conditioned research activity (e.g., a clinical trial where treatment is conditioned on the individual's authorization) as

well as an unconditioned research activity (e.g., the use or disclosure of PHI to create or contribute to a separate research database or repository), provided that the compound authorization clearly differentiates between the conditioned and unconditioned research components and provides the opportunity for the individual to opt in to the unconditioned research activities. The revised Common Rule continues to harmonize the research regulations with the Privacy Rule; it specifically modifies HHS' interpretation of the Privacy Rule's research authorization requirements to permit broad authorization for future research uses and disclosures, as long as the future research purposes are adequately described such that it would be reasonable for the individual to expect that her or his PHI could be used or disclosed for such future research.

4.6 Special Consent Issues: Incapacitated Adults and Children

In addition to the general requirements discussed above, there are also additional requirements for certain specific research populations. The revised Common Rule indicates that the focus of the IRB should be on populations that are vulnerable to coercion or undue influence and list four categories-- children, prisoners, economically or educationally disadvantaged individuals, and individuals with impaired decision making capacity. FDA has regulations that apply to children (which, for the most part, match the HHS regulations). Both also allow research to be conducted with adults lacking decisional capacity, although consent must be obtained by a "legally authorized representative," (LAR) who may be a guardian, proxy, or surrogate decision-maker (the terms are defined by state law). The revised Common Rule defines the LAR as:

> an individual or judicial or other body authorized under applicable law to consent on behalf of a prospective subject to the subject's participation in the procedure(s) involved in the research. If there is no applicable law addressing this issue, *legally authorized representative* means an individual recognized by institutional policy as acceptable for providing consent in the nonresearch context on behalf of the prospective subject to the subject's participation in the procedure(s) involved in the research.[62]

Likewise, HIPAA also allows for authorizations from "personal representatives" (again, generally defined by state law). Broad consents are allowed for research involving incapacitated adults and research involving children, provided the regulatory requirements for broad consent are met.

Of particular interest to registries are the research regulations pertaining to children. Research involving children must fit into one of four categories: minimal risk,[63] greater than minimal risk/prospect of direct therapeutic benefit,[64] minor increase over minimal risk/likely to yield generalizable knowledge about subject's disorder or condition,[65] and research not otherwise approvable but authorized by the Secretary of HHS or Commissioner of FDA in consultation with an expert panel.[66] Most registry research is likely to fall into the minimal risk category. For these studies, permission must be obtained from at least one parent/guardian and assent obtained from the child, if the child is capable of assenting.

Waivers of both permission and assent are possible. Under HHS regulations, a waiver of parental permission is allowed under the same conditions that allow for a waiver of informed consent in adult populations;[67] or when parental permission is not a reasonable requirement to protect the

subjects.[68] FDA regulations do not allow for waivers of parental permission. Both HHS and FDA regulations allow a waiver of assent when the research involves an intervention holding the potential for direct therapeutic benefit and is not available except through participation;[69] or when parental permission is waived in accord with section 46.116 of Subpart A. Moreover, when some or all of the children involved are not capable of providing assent, an IRB can determine that assent in not necessary (for the child or children in question, or for all children if appropriate). Both sets of regulations allow an IRB to determine that permission is only required from one parent, even when required from both under §§46.406 or 407 or §§50.53 or 54, in limited circumstances.[70] Where authorization must be obtained for the use or disclosure of PHI, the HIPAA Privacy Rule requires authorization from only one personal representative of the individual, such as one parent of a minor child, and does not require assent of the child.

OHRP has indicated that when the research in question involves a treatment for which the child would have legal authority to consent, the child's consent may suffice and parental permission may be unnecessary. The HIPAA Privacy Rule also generally provides that when a minor has legal authority to consent to a particular healthcare service without the involvement of a parent, the minor and not the parent has authority to act as the individual with respect to the PHI pertaining to that healthcare service. State statutes granting decision-making authority to minors vary. Many address issues such as treatment for sexually transmitted infections (STIs), access to contraception, and some even allow consent for mental health or substance abuse treatments. Registries involving these areas may be able to rely on the minor's consent, rather than the parental permission/assent framework. However, more specific legal guidance on the particulars of state statutory interpretation may be warranted in these situations.

Another important consideration is what to do when a minor who is involved in a registry reaches the age of majority. OHRP interprets the continuing consent standard to require that legal consent be sought from the participant upon reaching the age of majority. An authorization under the Privacy Rule, including one signed by a parent as the personal representative of a minor, remains valid until it expires or is revoked, even if such time extends beyond the child's age of majority. If the authorization expires on the date the minor reaches the age of majority, a covered entity would be required to obtain a new authorization signed by the individual in order to further use or disclose PHI covered by the expired authorization (or ensure that the use or disclosure falls within another regulatory permission in the Rule). Registries that involve children that will retain identifiable information past the child's age of majority will need to take steps to gain the appropriate consent and, if necessary, authorization for continued use. Less clear is whether investigators should seek a child's assent to continued participation when the initial consent was provided by parents at a time when the child lacked the capacity to play any role in decision making. If we follow the logic of reconsent in this context, continued assent is as important as the later consent to participation.

5. Approaches to Registry Consents

5.1 Current Practices and Problems

There are three current approaches to consent: opt-in, opt-out, and non-consent. An opt-in approach assumes that an individual will not be part of the registry until they have specifically

consented to participation. An opt-out approach assumes that all individuals will be part of a registry, unless there is a specific refusal to participate. Note that in order for an IRB to approve an opt-out approach for nonexempt, HHS-supported human subjects research, the researchers must document that the waiver of informed consent is appropriate for the research. Finally, a non-consent model does not seek or require individual consent or refusal, but includes all relevant individuals in a registry. This latter option will be problematic if the research falls within the federal regulations. The labeling of the approaches may vary in the literature, but the general concepts remain consistent. Additionally, some registries involve a mix of one or more approaches or a combined consent mechanism, where an opt-in approach is used for one aspect (access to a particular treatment) and non-consent for the other (listing in the treatment registry). This may also be referred to as "conditional access."

5.1.1 The Opt-In Approach

An opt-in procedure may involve a consent process similar to that used for clinical research protocols. It may be used separately for a registry, or it may be appended to a consent document used for a particular treatment (for example, individuals who consent to the use of a particular device may also be asked to participate in a registry for that device). While an opt-in approach has the benefit of assuring compliance with the Federal regulations, a number of the regulatory requirements are difficult to apply to registries (as discussed above). This led many to suggest a modified opt-in approach—using elements of the clinical research framework but adjusting to fit the registry model. But, even with a modified model, there are concerns that the strict informed consent requirements of the clinical research consent will have negative effects on subject selection, resulting in biases that will undermine the validity and thus affect the usefulness of the registry. An analysis of the Canadian Stroke Network estimated that dealing with consent issues cost $500,000 over the first 2–3 years of the registry, and the requirement to obtain written informed consent introduced significant selection biases undermining the usefulness of the registry.[71] The newly approved broad consent category should address many of these concerns and should be considered (along with an electronic consent tracking mechanism) for these types of registry studies.

5.1.2 The Opt-Out Approach

An opt-out procedure shifts the presumption from one in which each individual must consent to participate, to one in which each individual must refuse to participate. There is a great deal of discussion about the usefulness of an opt-out model, particularly for registries (e.g., organ donation registries). To be a valid opt-out model, individuals must be fully informed about the existence of the registry and their rights to opt-out of participation. In many cases, the information requirements are the same as the information requirements for an opt-in procedure—the only difference is that instead of explicitly agreeing to participate, the person must take steps explicitly to refuse to participate. While the information requirements may not change, the psychological shift may be significant. If the expectation is that everyone will participate, people may be more inclined to acquiesce. There is evidence in other areas of decision making that setting the default to participation results in greater inclusion than setting the default to non-participation, even when individuals are given an easy way to opt-in or opt-out.[72] While the Federal research regulations appear to assume an opt-in approach, in some circumstances an IRB

could approve a modification that allowed a shift to an opt-out. In order for an IRB to approve an opt-out approach for non-exempt, HHS-supported human subjects research, the researchers must document that the waiver of informed consent is appropriate for the research. An opt-out approach may be especially useful for some registries. Nonetheless, Privacy Rule requirements will preclude this approach unless the situation fits within one of the delineated permissible uses without an individual authorization (e.g., with a waiver of authorization for research or for public health activities).

IRBs could consider the opt-out approach for research that meets the five criteria for a waiver or alteration of consent under the HHS guidelines:

- The research involves no more than minimal risk to the subjects;
- The research could not practicably be carried out without the requested waiver or alteration;
- If the research involves using identifiable private information or identifiable biospecimens, the research could not practicably be carried out without using such information or biospecimens in an identifiable format;
- The waiver or alteration will not adversely affect the rights and welfare of the subjects; and
- Whenever appropriate, the subjects or legally authorized representatives will be provided with additional pertinent information after participation.[73]

For example, the Vermont Diabetes Information System (VDIS) was a quality-improvement, registry-based decision support and reminder system targeted to primary care physicians and their patients with diabetes. With IRB approval, VDIS incorporated an opt-out consent process.[74] Patients are notified by mail of their eligibility and inclusion in the registry and given a mechanism to opt-out by calling a toll-free number.

5.1.3 The Non-Consent Approach

Non-consent is not really a consent mechanism and thus will not be addressed here in detail. Nonetheless, this approach may, and probably should, entail providing participants with information about the registry. That is to say, even when consent is waived, there are ethical reasons to provide some notification and information. The format and process of disclosure may vary. In some cases, general public notifications (perhaps listing on a website, or posting prominently in a place likely to be seen) will be sufficient. In other cases, individual notification may be appropriate. A non-consent approach is used currently for registries that fall outside the federal research regulations (and thus not subject to the Common Rule or FDA regulations) such as state-mandated public health reporting or quality improvement activities. One primary methodological advantage of the non-consent approach in no-risk and minimal risk studies is that it can function to reduce concerns about biases introduced by the consent process, such as those that occur when individuals who consent to participate in the registry systematically differ from those who do not or cannot consent. Besides debates about when the use of a non-consent approach is acceptable (based on the level of permissible risk), most of the focus in this area should be on the type and extent of required notifications.

5.2 Scope of Consent

Consents may be broad or narrow. One term that is sometimes used is a so-called "blanket consent" approach, which asks for consent for all uses, unless one is specifically excluded. Blanket consent should be distinguished from "broad consent," described earlier. A broad consent refers to use in a wide array of biomedical research studies, it is not consent for all uses or even uses outside the biomedical research consent (such as for forensic use or for use by immigration authorities). A blanket consent model is sometimes seen in non-research contexts. For example, patients entering a health institution are sometimes asked to sign what might be described as a blanket consent (the validity of which may be questionable). But even in that context the consent is for all needed clinical care, and would not cover research. Another example may be a consent form for a specific clinical procedure which has a notation at the bottom of the form allowing the use of leftover tissue in any way deemed appropriate by the institution. Blanket consents are not generally viewed as valid exercises of autonomy and thus may not truly be considered to be "informed consent." At best, blanket consent may be viewed as a type of notification procedure, alerting individuals to the possible uses of their tissue or information. Neither the revised Common Rule nor the Privacy Rule permit blanket consents. Some registries may have been created using a type of blanket consent before the compliance date of the Privacy Rule, and may currently fall under an exemption to the human subjects research regulations; in these circumstances the previously obtained blanket consent may be deemed sufficient.

The real question related to the scope of consent is to what extent consent can and should authorize future unspecified uses. In other words, how broad a consent is permissible? The exercise of autonomy should include the ability to consent both to specific and to nonspecific research participation. An individual who would like to give expansive permission for the use of their data in any future registry (or for use in a particular registry, but include permission that the information may be shared with investigators for any future research query) is exercising a form of autonomy. In addition, part of the issue is in determining whether an expansive consent was truly informed. In the absence of specific details about the future uses, decision making is necessarily less informed than if every future use is spelled out clearly. However, the ethical doctrine of informed consent does not require this level of detail. Moreover, requiring multiple consent dialogues may respect autonomy less than permitting expansive consent if the individual does, in fact, wish to give permission and does not want continued recontact. The revised Common Rule now allows broad consent for biobank and databank research. By contrast, in other areas such as consent to participate in a clinical trial, expansive consents (e.g., "I give consent to participate in any clinical trial") are insufficient on both ethical and regulatory grounds.

5.3 Oversight and Community Consultation

Consent is only one aspect of the protections in place for human research participants. The second major part involves IRB review and oversight. Other chapters discuss the oversight roles for IRBs and registry governance boards.

Community consultation, a third concept that usually appears in the context of discussions on human subjects research consent, is not actually part of this system of protections. In fact, there is no simple community analog to individual consent. Consent requirements for research arise from the principle of autonomy, and there is no corresponding principle at the community level. Thus, concepts such as "community consent" or "community authorization" can be incoherent, in part because there is no unitary concept of a community. Communities may be defined on social, biological, religious, racial, cultural, or geographic grounds. Most people belong to multiple, sometimes overlapping, communities. Some of these communities may have a designated spokesperson, but this individual may not represent the interests of all members of the community. (Consider, for example, the complex relationship between the Pope and Catholics in the United States.) Other communities have no clearly identified spokesperson. It is inappropriate to consider community consultation as a replacement for individual consent. Rather than view community involvement as an aspect of consent, it should be considered as part of oversight (and an analog of IRB review).[75] Nevertheless, community involvement in the design and oversight of a registry may be particularly important when the registry involves socially identifiable groups that have been subject to historic discrimination or when it involves sensitive genetic information. In addition, in some cases, community involvement can enhance participant understanding of consent and thereby increase individual participation.

6. Consent Guidance

Given the nature of registry research, some elements of informed consent should be given special consideration, including: the scope of the use of registry data, potential for recontact, withdrawal, and information regarding the electronic data security and management to be employed.

6.1 Scope of Use of Registry Data

Registries constitute a valuable resource, since investigators often draw upon them to address questions extending far beyond those envisioned when the registries were first created. Therefore, informed consent for registry research that allows broad data sharing is optimal for promoting science. There may be instances, however, such as with respect to research on specific diseases (e.g., HIV/AIDS research), where more specific consent may be appropriate. In the meantime, registry developers should not only provide clear parameters regarding the scope of use of registry data when first creating the registry, but should also develop a mechanism to consider how future, possibly unanticipated, requests for data access will be evaluated. The registry governance board can play an important role in this situation.

6.2 Recontact

Individuals should be informed how their data/samples will be used at the point of entry into the registry. Whether and how participants will be recontacted should be established at the outset and included in the consent form. Exceptions should be considered specifically where data/samples were made irretrievably anonymous, since recontact would then be impossible. It is important to inform registry participants that the anonymization of their data will make withdrawal from the registry impossible.

6.3 Withdrawal

Many issues governing withdrawal from a registry have been discussed in this chapter. Consensus needs to be developed regarding whether withdrawal should be presented as an option to participants in the initial consent, and, if it is an option, how withdrawal will be managed. While withdrawal from a traditional research study is a basic subject right, withdrawal of collected data, even from clinical trials, may be restricted. It is extremely important that registry creators develop initial rules and procedures for withdrawal and fully inform participants of them.

6.4 Electronic Data Security

Given the public concerns about electronic data security, participants in the registry should be clearly informed as to the physical security of their data and/or biospecimens, including methods of coding and removal of identifiers, encryption techniques, potential for cloud computing, and quality assurance policies. As well, participants should be informed about the process of releasing and transferring data to future investigators as it relates to maintaining confidentiality. In some cases, this information will reassure participants, potentially increasing consent rates.

6.5 Proposed Consent Form Elements

The following is an outline of potential elements to consider when developing consent forms and engaging in consent dialogs for registry research. These elements were generated from the applicable HHS, FDA, and Privacy Rule requirements and include consent aspects particularly relevant to registry research. These are all issues that should be considered; there may be additional legal requirements (e.g., for a HIPAA authorization). The outline below should not be viewed as comprehensive or even applicable to all registries. Some registries will modify this outline to meet specific needs, while others will follow a consent procedure similar to one used for traditional clinical trials. However, this outline provides a starting place for understanding the scope of informed consent for registry participation. It is important to note that the responsibility for obtaining and assuring appropriate informed consent rests on multiple parties, including sponsors, investigators, Protocol Review Committees (PRCs), and IRBs. Moreover, despite the multitude of elements listed below, every effort should be made to keep consent forms as short as possible and at approximately a fifth- to eighth-grade reading level.

The following elements should be considered:

1. A statement that the individual is being asked to take part in a registry (or a research study, if applicable), including—
    a. The name of the specific registry for which consent is being obtained.
    b. An explanation of the purposes of the registry (why it was created, who will be included).
    c. The expected duration of participation.
    d. A description of the procedures entailed.
    e. The approximate size of the registry (if applicable and can be determined).

2. A description of any foreseeable risks or inconveniences (specifically risks related to any potential breach of confidentiality related to the data being collected).
    a. When human genetic research is anticipated, information should include possible consequences of genetic testing (e.g., insurance risks, paternity determinations, potential risks to family and community) and other related confidentiality risks, if such risks are determined to be real and relevant.
3. A description of the types of research that the repository will support, and any benefits to the subject or to others which may reasonably be expected, including—
    a. A statement about whether and how findings will be communicated to participants.
4. A statement describing the extent to which confidentiality of data/biospecimens identifying the subject will be maintained (including a description of the operations of the repository—how data/specimens will be stored and managed), including—
    a. If applicable, a statement about whether registry results will be published.
    b. A statement about the impact of participation on the subject's access to his/her medical records (e.g., that access may be limited until all work on the registry is completed).
5. The conditions and requirements under which data and/or specimens will be shared with recipient investigators, including—
    a. If applicable, a description that the data/specimens will be broadly shared and may be used for future research that is not yet identified.
    b. The fact that the data/specimens may be transferred to other institutions and explanation of a data transfer security plan.
6. A description of when recontact might be necessary, and how recontact will be handled.
7. A statement of whether there are any costs to participation and/or any payment for participation.
8. A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.
    a. The consequences of a subject's decision to withdraw from the research, including the possibility that the previously collected data will continue to be used, and procedures for orderly termination of participation by the subject.
9. Details on who to contact for answers to pertinent questions about the research and research subjects' rights.
10. As appropriate, any state-specific addenda.

References for Chapter 8

1. 21 CFR 50.3(c).
2. 45 CFR 46.101(a).
3. 45 CFR 46.102(e)(1) and (l).
4. 45 CFR 46.102 (e)(5).
5. The *21st Century Cures Act* requires harmonization ("to the extent practicable") between HHS and FDA research regulations by the end of 2019. 42 USCA Sec. 289.
6. 45 CFR 46.104.
7. 45 CFR 46.104(b)(2).
8. 45 CFR 46.104 (d)(4).
9. 45 CFR 46.104 (d)(7).
10. 45 CFR 46.104 (d)(8).
11. Casarett D, Karlawish J, Andrews E, et al. Bioethical Issues in Pharmacoepidemiologic Research. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. Sussex, England: John Wiley & Sons; 2005. p. 593.
12. 45 CFR 46.102 (l)(2).
13. *See, e.g.*, Office of the Assoc. Dir. for Sci., Ctrs. for Disease Control and Prevention, CDC-SA-2010-02, Distinguishing Public Health Research and Public Health Nonresearch. Document dated July 29, 2010. [August 10, 2018.] https://www.cdc.gov/od/science/integrity/docs/cdc-policy-distinguishing-public-health-research-nonresearch.pdf.
14. Centers for Disease Control and Prevention. Guidelines for Defining Public Health Research and Public Health Non-Research. Oct 4, 1999. [July 16, 2018]. http://www.cdc.gov/od/science/integrity/docs/defining-public-health-research-non-research-1999.pdf.
15. U.S. Department of Health and Human Services; Office for Human Research Protections. Activities Deemed Not to Be Research: Public Health Surveillance 2018 Requirements [November 7, 2018]. https://www.hhs.gov/ohrp/regulations-and-policy/requests-for-comments/draft-guidance-activities-deemed-not-be-research-public-health-surveillance/index.html.
16. 45 CFR 164.501.
17. 45 CFR 46.104(d)(4)(iii)
18. Johnson N, Vermeulen L, Smith KM. A survey of academic medical centers to distinguish between quality improvement and research activities. Qual Manag Health Care. 2006 Oct-Dec;15(4):215–20.
19. U.S. Department of Health and Human Services; Office for Human Research Protections. Quality Improvement Activities – FAQs. [July 16, 2018]. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/quality-improvement-activities/index.html.
20. National Committee on Vital and Health Statistics. Recommendations on Privacy and Confidentiality, 2006-2008. Section I: Privacy and Confidentiality in the Nationwide Health Information Network. May, 2009. [July 16, 2018]. https://ncvhs.hhs.gov/wp-content/uploads/2014/05/privacyreport0608.pdf.
21. Westin AF. How the Public Views Privacy and Health Research. March, 2008. [July 16, 2018]. https://www.ftc.gov/sites/default/files/documents/public_comments/health-care-delivery-534908-00001/534908-00001.pdf.

22. Privacy, Sec., and Confidentiality Subcomm., Nat'l Comm. on Vital and Health Statistics, Health Information Privacy Beyond HIPAA: A 2018 Environmental Scan of Major Trends and Challenges. Documented dated December 13, 2017. [August 10, 2018.] https://ncvhs.hhs.gov/wp-content/uploads/2018/05/NCVHS-Beyond-HIPAA_Report-Final-02-08-18.pdf.

23. Schneider S, Kerwin J, Robins C, et al. Consumer engagement in developing electronic health information systems: Final report. Rockville, MD: Agency for Healthcare Research and Quality; Jul, 2009. [July 16, 2018]. (Prepared by Westat, Rockville, Maryland, under Contract No. 233-02-0087). AHRQ Publication Number 09-0081-EF. http://healthit.ahrq.gov/sites/default/files/docs/citation/09-0081-EF.pdf.

24. Secretary's Advisory Comm. on Human Research Protections, *Attachment D - Recommendations for a Broad Consent Template*. [August 10, 2018.] https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-d-august-2-2017/.

25. Miller FG. Research on medical records without informed consent. J Law Med Ethics. 2008 Fall;36(3):560–6.

26. Barbara J. Evans. Appropriate Human-Subject Protections for Research use of Sentinel System Data. FDA Sentinel System Meeting Series Issue Brief: Legal Issues in Active Medical Product Surveillance. Mar, 2010. [July 16, 2018]. https://www.brookings.edu/wp-content/uploads/2012/04/Panel-3-Issue-Brief.pdf.

27. 45 CFR 46.116(a)(1).

28. 45 CFR 46.116(a)(3).

29. 45 CFR 46.116(a)(2).

30. 45 CFR 46.116(a)(5)(i).

31. 45 CFR 46.116(a)(6).

32. 45 CFR 46.116(b); 21 CFR 50.25(b).

33. Informed Consent Elements; Final rule. 76 Federal Register 2 (4 January 2011), pp. 256-270. [July 16, 2018]. https://www.gpo.gov/fdsys/pkg/FR-2011-01-04/pdf/2010-33193.pdf. ; 21CFR 50.25(c).

34. 45 CFR 46.116(c); 21 CFR 50.25(b).

35. 45 CFR 46.116(f).

36. 45 CFR 46.116(g)

37. 21 CFR 50.23.

38. U.S. Department of Health & Human Services. Informed Consent Requirements in Emergency Research. [July 16, 2018]. http://www.hhs.gov/ohrp/policy/hsdc97-01.html.

39. 21 CFR 50.24.

40. 45 CFR 46.116 (d)

41. Secretary's Advisory Comm. on Human Research Protections, *Attachment C - Recommendations for Broad Consent Guidance*. [August 10, 2018.] https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html.

42. Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet. 2008 Aug;4(8):e1000167.

43. Jacobs KB, Yeager M, Wacholder S, et al. A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies. Nat Genet. 2009 Nov;41(11):1253–7.

44. 45 CFR 46.117(c).

45. 45 CFR 46.117(b)(2); 21 CFR 50.27(b)(2).

46. U.S. Department of Health and Human Services; Office for Human Research Protections. FAQs, Can an electronic signature be used to document consent or parental permission? [July 16, 2018]. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/informed-consent/index.html.

47. Office for Human Research Protections, US Department of Health and Human Services, et al. The Use of Electronic Informed Consent. Documented dated December 2016. [August 10, 2018.] https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm436811.pdf.

48. 45 CFR 46.116(h)

49. Office for Human Research Protections. Department of Health and Human Services. Clinical Trial Informed Consent Form Posting (45 CFR 46.116(h)). [February 28, 2019.] https://www.hhs.gov/ohrp/regulations-and-policy/informed-consent-posting/index.html.

50. Office for Human Research Protections, Department of Health and Human Services, *Research with Children FAQs*. [August 10, 2018.] https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/children-research/index.html.

51. U.S. Department of Health & Human Services. Guidance on Withdrawal of Subjects from: Data Retention and Other Related Issues. [July 16, 2018]. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/guidance-on-withdrawal-of-subject/index.html.

52. Pub. L. No. 110-85, 121 Stat. 823 (2007) (Codified in 21 U.S.C.).

53. 45 CFR 164.508. [August 10, 2018]. For additional information on authorizations for research, see http://privacyruleandresearch.nih.gov/authorization.asp.

54. 45 CFR 164.512(i)(1)(ii).

55. 45 CFR 164.512(i)(1)(iii).

56. 45 CFR 164.514 (a)-(c).

57. 45 CFR 164.514(e).

58. 45 CFR 164.512(i)(2).

59. 45 CFR 164.532.

60. Nass SJ, Levit LA, Gostin LO, Committee on Health Research and the Privacy of Health Information, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. National Academies Press; The HIPAA Privacy Rule; Institute of Medicine; [June 24, 2020]. http://www.nap.edu/catalog/12458.html

61. Mark Barnes & David Forster, SACHRP, Subcommittee on Harmonization (SOH) Update: July 20, 2011. 2011. [July 16, 2018]. https://wayback.archive-it.org/3919/20150824191836/http://www.hhs.gov/ohrp/sachrp/mtgings/mtg07-11/soh.pdf.

62. 45 CFR 46.102(i).

63. 45 CFR 46.404; 21 CFR 50.51.

64. 45 CFR 46.405; 21 CFR 50.52.

65. 45 CFR 46.406; 21 CFR 50.53.

66. 45 CFR 46.407; 21 CFR 50.54.

67. 45 CFR 46.408(a), (b); 45 CFR 46.116.

68. 45 CFR 46.408(c).

69. 45 CFR 46.408(a); 21 CFR 50.55.

70. 45 CFR 46.408(b); 21 CFR 50.55(e).

71. Tu JV, Willison DJ, Silver FL, et al. Impracticability of informed consent in the Registry of the Canadian Stroke Network. N Engl J Med. 2004 Apr 1;350(14):1414–21.

72. Thaler RH, Sunstein CR. Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale: Yale University Press; 2008.

73. 45 CFR 46.116(f)(3).

74. Littenberg B, MacLean CD. Passive consent for clinical research in the age of HIPAA. J Gen Intern Med. 2006 Mar;21(3):207–11.

75. Marshall PA, Berg JW. Protecting communities in biomedical research. Am J Bioeth. 2006 May-Jun;6(3):28–30. discussion W46-8.

Case Examples for Chapter 8

Case Example 16. Issues With Obtaining Informed Consent

Description	The Ontario Stroke Registry is a registry of stroke patients in Canada. The registry is a non–consent-based registry that collects detailed clinical data on the acute stroke event from the onset of symptoms, including emergency medical service transport, emergency department care, and hospital discharge status. The purposes of the registry are to monitor and report on the quality of stroke care in Ontario and to provide a rich clinical database for research.
Sponsor	CorHealth Ontario, with original funding from the Canadian Stroke Network, Networks of Centres of Excellence, and Ontario Ministry of Health and Long Term Care
Year Started	2001
Year Ended	2014
No. of Sites	154
No. of Patients	100,000

Challenge

The registry began in 2001 with Phase 1, in which data were gathered from 21 hospitals in Canada. All patients admitted to the hospital or seen in the emergency department with symptoms of acute stroke within 14 days of onset or transient ischemic attack (TIA), as well as those with acute in-hospital stroke, were included in this phase. Research nurse coordinators identified eligible patients through daily reviews of emergency and admission patient lists and approached these patients for consent. Informed patient consent was required for full data collection, linkages to administrative data, and 6-month followup interviews.

Informed consent was required for full data collection. Unfortunately, consent was obtained for only 39 percent of eligible patients. Subsequent analyses showed that patients who consented to participate were not representative of the overall stroke population, as they were less likely to have severe or fatal stroke, and also less likely to have minor stroke or TIA.

Phase 2 of the registry began in 2002, with 21 hospitals and 4 Ontario Telestroke sites. In this phase, all patients admitted to the hospital or seen in the emergency department with symptoms of acute stroke within 14 days of onset or TIA were included. Patients with in-hospital stroke were no longer recruited. In order to standardize workload across the country, a random sample of eligible patients was selected to be approached for consent for full data collection. Consent was obtained from 50 percent of eligible patients.

After obtaining consent of only 39 percent and 50 percent of patients in Phases 1 and 2, the team realized that obtaining written patient consent for participation in the registry on a representative sample of stroke patients was impractical and costly. Patient enrollment threatened the viability

and generalizability of the stroke registry. The registry team published these findings in the New England Journal of Medicine in April 2004.

Proposed Solution

The registry team approached the Ontario Information and Privacy Commissioner to discuss a non-consent-based registry for Phase 3. As a result of these discussions, the registry was "prescribed" by the Privacy Commissioner under the Personal Health Information Protection Act, 2004. This decision allowed the registry to collect data legally on stroke patients without written consent.

Results

Phase 3 of the registry included all patients presenting to emergency departments of the 11 "Stroke Centres" in Ontario and 1 center in Nova Scotia with a diagnosis of acute stroke or TIA within 14 days of onset. Nurse coordinators identified eligible patients through daily reviews of emergency and admission patient lists. Patients were identified prospectively, with retrospective chart review, without consent. No followup interviews were done. Because informed consent was not required, the data collected provided a representative sample of stroke patients seen at tertiary care centers in Canada, making the data more viable for use in research and in developing initiatives to improve quality of care. Phase 4 of the registry expanded to include a population-based, province-wide audit of stroke care delivery on a random sample of patients from every acute care institution in Ontario. Using unique encoded identifiers, data from Phases 3 and 4 of the registry are linked to population-based administrative databases housed at the Institute for Clinical Evaluative Sciences to provide information on mortality, readmissions, physicians' services, and medication prescriptions.

Key Point

The impact of obtaining informed consent should be considered in developing a registry. Requiring that registries obtain the consent of patients with acute medical conditions such as stroke may result in limited selective participation, as it is not possible to obtain consent from all patients. For example, patients who die in the emergency department and patients who have brief hospital visits may be missed. Mechanisms such as obtaining a waiver of informed consent or using the approach outlined in this case may be alternatives.

For More Information

- https://www.ices.on.ca/Research/Research-programs/Cardiovascular/Ontario-Stroke-Registry
- Tu JV, Willison DJ, Silver FL, et al. The impracticability of obtaining informed consent in the Registry of the Canadian Stroke Network. N Engl J Med. 2004;350:1414–21. PMID: 15070791. DOI: 10.1056/NEJMsa031697.

Case Example 17. Operationalizing Informed Consent for Children

Description	TARGetKids! (The Applied Research Group for Kids) is a prospective registry enrolling healthy children aged 0-5 years. The aim of the registry is to link early nutritional exposures to later health outcomes including obesity, micronutrient deficiency, and developmental outcomes.
Sponsor	University of Toronto
Year Started	2008
Year Ended	Ongoing
No. of Sites	15 primary care practices in Toronto, Canada
No. of Patients	>10,000

Challenge

Research involving children faces unique challenges, including special requirements related to the informed consent process. In this pediatric patient registry, patients are recruited at their annual well-child visits and followed up for ten years during subsequent annual well-child visits. Participation involves completion of age-specific questionnaires related to the child (asking for nutritional, behavioral, and developmental information), collection of anthropometric measurements, and collection of the child's blood sample (4-7 mL) by a trained pediatric phlebotomist.

Consent for the registry is provided by one or both parents. By signing consent, parents also authorize the collection of their child's health card number to allow researchers to access the child's health records. Registry staff anticipated challenges in obtaining informed consent for these activities, particularly given the infrequency of contact with patients and the fact that blood sample collection is not part of normal clinical care during these annual visits. After reviewing the registry protocol, a research ethics board recommended that steps be taken to minimize coercion when recruiting and obtaining consent from patients.

Proposed Solution

Two weeks before a scheduled well-child visit for an eligible patient, the physician's office mails a short informational letter to the child's home. The purpose of the letter is to provide information about the registry and prepare parents for their contact with registry staff during the visit. By providing this information in advance, the letter minimizes the possibility that parents will feel coerced to consent to registry participation.

On the day of the visit, the child's parents are approached by a registry research assistant to provide consent for participation of their child in the registry. The research assistant explains what participation entails (i.e., completing questionnaires, collection of anthropometric measurements, and collection of a blood sample). If the parent spontaneously expresses that they wish to participate in the registry but don't wish to participate in one of these activities, they are

given the option to opt out of one portion of the registry (e.g., blood collection) while still consenting for others (e.g., questionnaires and anthropometric measurements).

Results

The registry is now following over 10,000 children aged 0 to 5 years from 15 primary care practices in Toronto and Montreal, Canada. The participation rate for the registry is 49 percent of all eligible children in the targeted practices (defined as children aged 0-5 years with a well-child appointment scheduled). The primary reasons parents decline to participate in the registry include lack of time to answer the questionnaire, no legal guardian accompanying the child, the need to discuss participation with spouse, and their child not feeling well that day.

Of consenting parents, about 50 percent consent to the registry blood sample. One possible reason parents choose to consent to the blood sample is that they see value in the test results (i.e., for iron or vitamin D deficiency) which are not standard of care for children in Canada. Parents may perceive this as an added benefit, although, in an attempt to minimize coercion, the informed consent form and registry staff do not emphasize this fact.

Key Point

Providing patients (and parents of pediatric patients) with information about the registry in advance can give them time to prepare questions and thoughtfully consider whether they wish to consent to participation. A flexible consent structure that allows patients to opt out of activities of a sensitive nature can reduce barriers to consent and participation.

For More Information

- https://www.targetkids.ca/
- Carsley S, Borkhoff CM, Maguire JL, et al. Cohort Profile: The Applied Research Group for Kids (TARGet Kids!). Int J Epidemiol. 2015;44(3):776-88. PMID: 24982016. DOI: 10.1093/ije/dyu123
- Morinis J, Maguire J, Khovratovich M, et al. Paediatric obesity research in early childhood and the primary care setting: The TARGet Kids! Research Network. Int J Environ Res Public Health. 2012 Apr;9(4):1343–54. Epub 2012 Apr 16. PMID: 22690197. DOI: 10.3390/ijerph9041343.
- Birken CS, Maguire J, Mekky M, et al. Parental factors associated with screen time in preschool children in primary-care practice: a TARGet Kids! study. Public Health Nutr. 2011 Apr 5:1–5. PMID: 21466741. DOI: 10.1017/S1368980011000516.
- Maguire JL, Birken CS, Jacobson S, et al. Office-based intervention to reduce bottle use among toddlers: TARGet Kids! Pragmatic, randomized trial. Pediatrics. 2010 Aug;126(2):e343–50. PMID: 20624802. DOI: 10.1542/peds.2009-3583.

Case Example 18. Obtaining a Waiver of Informed Consent

Description	The STS/ACC TVT Registry™ tracks patient safety and real-world outcomes related to transcatheter valve replacement and repair procedures. The registry collects data on patient demographics, procedure details, and facility and physician information to support analyses of patient outcomes and clinical practice patterns. The Centers for Medicare & Medicaid Services (CMS) has approved the registry as meeting the requirements outlined in the national coverage decisions for transcatheter aortic valve replacement (TAVR) and transcatheter mitral valve repair (TMVR). The participant agreement for the registry permits the registry sponsors to conduct cardiovascular research using a limited dataset. The registry sponsors' policy requires that all research be conducted consistent with the Common Rule.
Sponsor	The Society of Thoracic Surgeons (STS) and the American College of Cardiology (ACC)
Year Started	2012
Year Ended	Ongoing
No. of Sites	617 hospitals
No. of Patients	67,353

Challenge

Aortic stenosis is the most common valvular abnormality in the United States, and the prevalence of this condition is expected to increase as the U.S. population ages. Until recently, surgical aortic valve replacement has been the only effective treatment for adults with severe symptoms. Transcatheter heart valve procedures provide is new treatment options for patients who are not eligible for conventional heart valve replacement or repair surgery, but evidence is lacking on long-term patient outcomes.

In 2012, CMS issued a Medicare National Coverage Decision for TAVR. Under the decision, CMS permits Medicare coverage for TAVR only when (1) the procedure is performed with a device approved by the U.S. Food and Drug Administration (FDA) consistent with labeled indications and any other FDA requirement; (2) the procedure is performed in facilities meeting certain requirements; and (3) when all patients undergoing the TAVR procedure are included in a national TAVR registry or participate in an approved clinical study. The national TAVR registry must consecutively enroll TAVR patients, accept all manufactured devices, follow patients for at least one year, and comply with all relevant regulations related to the protection of human research subjects. The National Coverage Decision specifically requires that the registry collect data on the following outcomes: stroke, all-cause mortality, transient ischemic attacks, major vascular events, acute kidney injury, repeat aortic valve procedures, and quality of life.

The development of a national registry to meet these requirements was challenging, particularly due to the need to collect at least one year of followup and quality of life data. The registry was

expected to enroll hundreds of sites and thousands of patients, making it time consuming, administratively cumbersome, and expensive to obtain local institutional review board (IRB) approval for each site and informed consent for each patient.

Proposed Solution

The registry developers determined that the national TAVR registry was most likely to be successful if it collected data that was already routinely documented as part of the standard of care and was able to obtain a waiver of informed consent from a central IRB.

When designing the data collection instruments for the registry, the developers worked closely with surgeons and interventional cardiologists to understand which data are already collected. The developers were able to limit the registry data collection effort to data already collected routinely, thereby allowing registry data to be abstracted from the medical record with no data collected solely for the purposes of the registry. In particular, the registry developers carefully considered how to collect followup data and quality of life data without requiring the collection of information solely for the purposes of the registry.

Based on discussions with surgeons and interventional cardiologists, the developers determined that patients are seen for followup care routinely at 30 days and 1 year following the procedure. Published guidelines have established the use of the Kansas City Cardiomyopathy Questionnaire (KCCQ) to assess quality of life as a standard of care for TAVR patients at these followup visits. The registry was designed to use the data collected at these followup visits, including the KCCQ, to meet the requirements for collecting long-term outcomes and quality of life information.

Results

The registry began collecting data in 2012 and has been approved by CMS as meeting the requirements of the Medicare National Coverage Decision. The ACC and STS, the institutions operating the registry, are considered the only entities engaged in research, while the participating sites are considered to be providing data only. The registry was approved only by the single IRB designated under the ACC/STS's Federalwide Assurance. Based on the registry protocol, the IRB determined that the ACC/STS are engaged in research on human subjects and granted a waiver of informed consent. The waiver of informed consent was awarded primarily because the participating sites are collecting and submitting information that is already documented in the medical record as part of the standard of care. As the registry operators, the ACC and STS submit data, including patient identifiers, to CMS as required by the National Coverage Decision. However, the ACC and STS only conduct research on a limited dataset, and any research studies not covered by the protocol are submitted to the IRB for review.

Because the ACC and STS have IRB approval and a waiver of informed consent, and because the data are already collected as part of the standard of care, the individual sites participating in the registry do not necessarily need to go through an IRB prior to enrolling in the registry. Some individual sites elect to submit the registry to their local IRB, in many cases because they intend to use the data they collect for the registry in additional research studies and are required by policy to seek IRB review. The local IRBs generally have reached the same conclusion as the central IRB. However, a local IRB may reach a different conclusion, perhaps due to differences in the data collection process at an individual site. For example, the data collection process at an individual site may provide an opportunity to consent the patient, in which case the IRB may not

grant a waiver of informed consent. In these cases, the individual site will follow the advice of the local IRB.

Since its launch in 2012, the registry has collected data on over 70,000 patients, resulting in a clinically rich data repository that has produced dozens of publications and presentations. The registry also meets multiple needs for different stakeholders; for example, CMS receives data to inform policy decisions on requirements for TAVR (such as the proposed decision memo released in early 2019), the FDA receives data to monitor device safety, and industry can leverage the existing registry infrastructure to meet post-approval FDA requirements in an efficient manner and to generate data to support expansion of approved indications.

Key Point

Protecting the subjects whose data will be used is of the utmost importance when developing a registry. When developing a registry, sponsors should consider the planned data collection effort in the context of requirements for IRB review and informed consent. This approach may help the registry identify a strategy that protects patients' privacy without overburdening the participating sites.

For More Information

- https://www.ncdr.com/TVT/Home/Default.aspx.
- Carroll JD, Shuren J, Jensen TS, et al. Transcatheter Valve Therapy Registry Is A Model For Medical Device Innovation And Surveillance. Health Affairs. 2015;34(2):328-34. PMID: 25646114. DOI: 10.1377/hlthaff.2014.1010.

Chapter 9. Registry Governance

1. Introduction

Registries function in a dynamic environment and are often shaped by the complex relationships among individual health, public health policy, economics, geography, and culture. Complexity within registries stems from the topics being studied, stakeholders with different agendas, and the legal and political climates for such research, among other factors. For example, studies of newly emerging infectious diseases or studies of addiction may be conducted in a highly politicized environment,[1] while geographical and cultural distances among stakeholders may introduce challenges in multinational registries. Even registries that exist within a single institution may need to address competing stakeholder needs and challenges such as changes in funding.

Governance is an important tool to help registries manage complexities such as these across the registry lifecycle, from the initial planning phase through the dissemination of results. Registry governance refers to a formalized, often written, structure or plan, for managing the registry and guiding decision making related to registry funding, operations, and dissemination of information.[2] Registry governance can take many forms depending on the scope of the registry, the number of stakeholders, and the purpose of the registry. Ultimately, the goal of governance is to provide a mechanism for individuals to work together to achieve the goals of the registry.

This chapter presents considerations for good registry governance and addresses how patients can contribute.

2. Governance Roles

Many types of stakeholders may be involved in registry governance depending on the scope and purpose of the registry. Frequently, stakeholders involved in governance include the principal investigator (PI), clinical experts, biostatisticians and/or epidemiologists, and representatives of the funding agency/sponsor. Some registries may include additional stakeholders, such as representatives from payers, industry, government agencies, and patient advocacy organizations.

Inclusion of different stakeholder perspectives in the registry governance framework is important to ensure that the registry meets the needs of these stakeholders, but decision-makers from different sectors have conflicting priorities. Engaging each stakeholder group with the common goal of improving healthcare quality and patient outcomes through sharing of data and other resources is vital to the success of the registry. Such collaborations have occurred successfully in several industries where no single entity had the resources or expertise to drive an entire field.[3,4] Successful collaborations satisfy the needs of multiple stakeholders and provide immediate value as well as long-term returns, while driving productivity and, sometimes, the development of best practices.

Successful collaborations require clear goals and adept leadership to unite all stakeholders involved with a registry, as discussed below. Building and maintaining trust and collaboration with important decision-makers and relevant, reputable contributors is paramount.[5-7] In addition,

it is critical to set appropriate expectations at the start of the project by establishing agreement among participating individuals on the time and resources to be committed.

3. Governance Structure and Functions

3.1 Governing Strategy

Because effective leadership is vital to robust governance, it is important to have overarching leadership to steer registry business and develop consensus among stakeholders. Leadership may be in the form of a single person, often the PI, or it may derive from a committee of individuals made up of, for example, co-PIs and representatives from various partners in the registry (e.g., funding sources, reporting entities, or subcontractors that handle operational aspects of the registry).

A framework to divide governance duties and to address potential areas of conflict should be developed at the registry's inception by involved parties. In general, topics central to the governance of registries will include the research agenda, funding, data, ethics, science, and dissemination. The registry governance plan should describe how the registry will function and who will be responsible for each task or functional area. Asking the following (or similar) questions can help determine duties and solidify further governance structure:

- *Research agenda:* Who sets priorities and establishes the goals of the registry? What are the respective roles of the stakeholders in setting the research agenda?
- *Funding:* Who takes responsibility for securing the necessary funds for operating the registry, including start-up and maintenance costs? Who will manage the acquired funds, and how will they be allocated to perform the various tasks? What role in decision making, if any, does the financial sponsor(s) have?
- *Data:* What data are needed and how will they be collected? What technologies will be used for data collection (e.g., electronic data capture, digital devices, electronic acquisition of patient reported outcomes)? What quality assurance procedures will be used? Who owns the information collected by the registry, and, for a tissue repository, who owns the materials in it? Will the registry adapt as new data needs are identified, and, if so, how?
- *Ethics*: How will the registry protect patient privacy? Will informed consent be required? What ethical oversight is required (e.g., institutional review board approval)?
- *Science*: How will the registry ensure that enrollment and followup procedures are followed? How will recruitment and patient retention be monitored over time? How will the registry address changes in the clinical area of interest (e.g., the introduction of a new treatment, the emergence of a new research question)?
- *Dissemination:* Who has access to the data for research purposes? Can third parties request access for specific research purposes? Who reviews publications resulting from analyses of registry data?

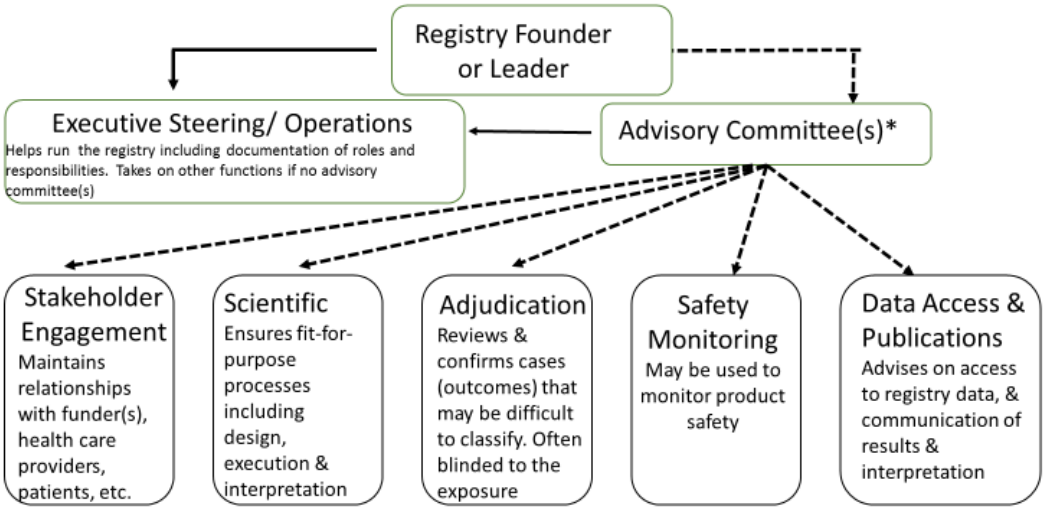
In all governance models, oversight policies and procedures for decision making, operations, and reporting of results should be transparent to stakeholders. Many registries essentially operate as

small businesses with complex operational components including regulatory requirements, finances, informed consent, data entry software, progress reports, periodic meetings, and scientific analyses and publications. These day-to-day operations should be conducted as a structured effort, with well-defined lines of authority and responsibility. The structure should also have the flexibility to adapt to evolving science and changing regulatory requirements, such as the periodic updates to the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and the introduction of the EU General Data Protection Regulation (GDPR) in Europe.

3.2 Governance Committees

Many registries use committee systems to divide and assign oversight duties. The registry may use one or more advisory boards or other leadership committees comprised of representatives of all or some stakeholders in the registry depending on the purpose. Examples of various types of committees and functions are illustrated in Figure 9-1.

Figure 9-1. Committee functions comprising governance



An Executive Steering Committee generally assumes responsibility for the major financial, administrative, legal/ethical, and scientific decisions that determine the direction of the registry. These decisions are made with appropriate input from legal, scientific, and administrative experts. Depending on their capabilities and the size and resources of the registry, this group may also assume other advisory functions. The funding source (or sources) generally have a seat on such steering committees, though whether they have a formal voting role should be considered in the context of specific areas of decision making, described in subsequent sections. An operations committee generally is used for small or less complex registries and usually takes on all the functions shown in Figure 9-1. The Principal Investigator is generally included on this steering committee.

Advisory committees may be organized to review the registry data, develop consensus statements, provide recommendations on modifications or enhancements to the registry, and/or assist in the dissemination of information and the formulation of strategies to encourage enrollment. Advisory committee actions generally are limited to making recommendations to the ultimate decision maker, whether an executive committee, financial sponsor, or the registry developer/leader. Some advisory committees make decisions by consensus rather than by a formal vote, whereas majority voting is used by others. Consensus-driven decisions can be particularly challenging when partners are not aligned, or a single member takes an intransigent position. In situations where disputes are difficult to resolve, it is helpful to be able to rely on pre-established committee voting rules to develop a path forward. Documentation (i.e., meeting minutes) should be created, distributed, and approved to memorialize decisions and actions taken for all types of advisory committees.

The mix of stakeholders (people or entities who have a stake in the findings of the registry) and experts (e.g., clinicians, epidemiologists, statisticians) involved in advisory committees relates largely to the purpose of the registry. Broad stakeholder involvement in governance committees is most desirable when the outcome or findings can impact many stakeholders. Depending on the size of the registry, governance may be assumed by various oversight committees made up of interested individuals who are part of the design team (internal governance) or who remain external to the day-to-day operations of the registry (external governance). An external advisory board can assist registry operations in two important ways: (1) providing guidance to the registry developer for the clinical, scientific and/or technical aspects of the registry operations and (2) establishing the scientific independence of the registry. The latter function is especially valuable when controversies arise, such as those relating to patient safety and treatment.

Differences in the registry purpose and scope, the research questions of interest, the relevant data sources, and many other factors will influence the degree of involvement and role of oversight groups. In other words, committee functions are described below to highlight the roles that are typically assumed by the registry governance structure; these should be modified, as appropriate, to meet the needs of a specific registry. Advisory committees may be differentiated as follows:

- *Stakeholder engagement:* In some registries, an advisory committee may be established to maintain relationships with key stakeholders, such as the funding source(s), healthcare providers, collaborating researchers, and patients. Some registries include representatives of these stakeholders on advisory committees, while others use a committee to coordinate stakeholder outreach activities. Examples of outreach activities include interactive workshops or exchange forums between public and private sectors, dissemination of publications and news releases, and updates at professional meetings.[8] Plans for stakeholder engagement should be considered at the outset of the registry and documented.
- *Scientific:* A scientific advisory committee often includes experts in areas such as database content, clinical research or clinical care, epidemiology, and biostatistics. This group may determine the overall direction of the research program, including identifying updates that may be needed to address changes in clinical practice or newly approved treatments. This committee also may recommend or lead specific data analyses and reporting of results. In some registries, the scientific advisory committee also is responsible for data access and publications (see below).

- *Adjudication:* Adjudication may be used to review and confirm cases (diagnosis) and/or endpoints (outcomes) that may be difficult to classify accurately and consistently. Individuals performing this function are generally blinded to the exposure (product or process) under study so that the confirmation of outcomes is made without knowledge of exposure.
- *Safety:* Most registries will not require a formal safety monitoring committee. In studies of experimental drugs or devices, a formal Data Safety Monitoring Board commonly is used to detect safety problems that would indicate that the experimental product is unsafe and clinical research should not proceed. In the context of registries, situations may arise in which the registry is responsible for the primary accumulation of safety data on a specific intervention; in such situations, a safety monitoring board or individual charged with this responsibility may be useful for conducting periodic reviews (e.g., annually) to look for unusual patterns indicating a serious safety problem.
- *Data Access and Publications/Dissemination of Results:* The plan for disseminating registry findings should be discussed early in the planning phase of the project, since preparing results for wide dissemination requires considerable time and effort. This function, often covered by the Data Access and Publications Committee (DAPC), should address the process by which registry investigators (and possible external parties) access and perform analyses of registry data for the purpose of submitting abstracts to scientific meetings and developing manuscripts for peer-reviewed journal submission. The role of all stakeholders in the publication process should be specified, as well as any plans to allow external parties to access the data.

Many external parties may request access to registry data. For example, a government regulatory agency may request registry data to fulfill safety reporting requirements or other obligations. Pharmaceutical companies or device manufacturers may request data from patients who receive their products or may request registry data to use as controls for a clinical trial. An external investigator(s) may request permission to access registrywide aggregate data for the purposes of evaluating a new scientific question. Lastly, data requests from the media or the public are possible, although rare. Any information released to the public by the registry (via newsletters, public website, or other methods) should be reviewed prior to release to ensure that data confidentiality is not compromised.

A process for reviewing and responding to such requests from other investigators or entities should be considered in registries that may generate broad external interest, if the registry stakeholders and participants are agreeable to such use. The DAPC is often used to develop and implement policies and manage requests for data access in a transparent, consistent way that is agreeable to all stakeholders; for small registries, this function may be handled by the Operations Committee.

Depending on available resources, a registry may conduct analyses in-house to support external requests; contract with an outside agency to conduct analyses; or leave this task to the data requestors. If an outside agency or data requestor will be conducting analyses of registry data, it is essential for the Operations Committee or DAPC to retain oversight of these activities, especially those that are intended for use in manuscripts submitted to peer-reviewed publications. It is also important to ensure that data access complies with all applicable ethical, legal, and

regulatory requirements and that appropriate care is taken to protect patient privacy and maintain data security and integrity.

Authorship (including that of registry sponsors) in scientific publications should satisfy the conditions of the Uniform Requirements for Manuscripts Submitted to Biomedical Journals.[9] For manuscripts that are prepared by external investigators, many registries ask to review manuscripts prior to publication to ensure that the data and operations of the registry are represented correctly.

3.3 Committee Operations

Governance committees generally should be small to facilitate decision making. An odd number of members is preferable to avoid ties in voting. In practical experience, Executive Steering Committees (or entire Advisory Boards) are generally composed of fewer than ten members. It is important to recognize that these members may solicit input and advice from stakeholders or experts who are not part of the Steering Committee. Some governance committees may create formal subcommittees to develop recommendations on specific issues. In those situations, the Steering Committee member who leads those subcommittees will represent the subcommittee members at the Steering Committee meetings.

Transparency regarding any perceived or actual conflicts of interest is important for effective governance. A plan for identifying and managing actual and perceived financial or intellectual conflicts of interest (COI) should be developed at the registry outset. The plan should clearly spell out the timeline and process for obtaining completed COI and financial disclosure forms from participating members and for reviewing and addressing any potential conflicts. It may be helpful to define what constitutes a problematic COI, and this definition should be reviewed on a periodic basis and revised if needed. Committee members should promptly notify the registry developer/leader should they identify an emerging potential COI (minor or major) during their service.

3.4 Governance in Change Management

Anticipating and planning for change is good practice for all patient registries. Governance can play a critical role in identifying the need for change and providing oversight to ensure that the transition goes as smoothly as possible, with minimal disruption to registry operations.

As an example, an ongoing registry may transition to new organizational leadership or to a new source of funding. In these situations, membership of the key advisory board(s) should be reviewed to ensure appropriate representation of stakeholders. During the transition, advisory board(s) can act as advocates of change by publicly supporting the transition and helping to engage and motivate clinicians at the participating centers. External stakeholders, such as patient advocacy groups and regulatory agencies/health authorities, may also be informed of the transition and, depending on the goals of the transition, potentially enlisted as additional public advocates for the registry transition.

Of note, registry transitions may require revisiting the registry's data access policies and procedures. If a DAPC is already in place, the committee should be charged with (1) determining

how changes in the registry will affect the policies and procedures for accessing data, and (2) reviewing the operational plan for executing analysis plans with respect to the registry transition. Furthermore, if the transition involves a change in registry stakeholders, the procedures for conducting analyses and developing publications should be re-examined. New stakeholders may need to be involved in the prioritization of analysis plans, conduct of analyses, and/or the review of scientific abstracts and manuscripts.

4. Special Applications

4.1 Patient Engagement[i]

As with many types of outcomes research, there is a growing trend to involve patients and patient advocacy groups in the planning and operation of patient registries. In some cases, patient organizations are the funding source for the registry and are actively involved in all aspects. (See Case Example 20.) In other cases, patients work as research partners with the registry, offering valuable insights on user burden (e.g., when the registry is collecting patient-reported outcomes), registry feasibility, approaches to recruitment, and patient support needs. They may also serve on advisory committees to facilitate engagement with the patient community. Working with patients brings certain unique considerations, and varying levels of patient involvement may be appropriate and/or feasible for different registries. These considerations are discussed in depth in the supplemental eBook, *21st Century Registries*,[10] and summarized here in Table 9-1.

Table 9-1. Key points to consider when engaging patients in governance

Domain	Key Points
Roles and responsibilities	• Be sensitive to time constraints. Before engagement, clearly detail expectations (e.g., anticipated commitment of time and types of activities). • Be clear about the patient partner roles – do not expect community members to perform academic duties. • Include patient partners in the development and design of any research materials that may be used to recruit registry participants. • Include patient partners in discussions about the exposures and outcomes of interest in the study. • Consider events that patient partners can lead to increase engagement.

[i] *Adapted from Gliklich RE, Dreyer NA, Leavy MB, Christian JB (eds). 21st Century Patient Registries. EBook addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. (Prepared by L&M Policy Research, LLC under Contract No. 290-2014- 00004-C.) AHRQ Publication No. 17(18)-EHC013-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018.*

Domain	Key Points
Trust and transparency	• Engage patient partners early in the process. • Create a true patient-researcher partnership community by bringing the registry team, patients, caregivers, patient advocates, and other stakeholders together for discussions. • Acknowledge patient and/or caregiver contributions frequently. • Maintain ongoing relationships to build trust and credibility. • Conduct frequent reviews of how well patient engagement is working and if changes need to be made to improve patient partner input and involvement. • Present ongoing and final study results to patient partners and other stakeholders.
Training and education	• Present to researchers and patients on the importance of patient-centered research to help increase awareness. • Educate researchers on how to be more effective listeners so that they hear and act on recommendations from patient partners. • Educate patients in how to be more effective communicators and bring their voice to the table (e.g., European Patients Academy on Therapeutic Innovation (EUPATI) and the Patient-Centered Outcomes Research Institute (PCORI)). • Be sensitive to technology limitations and how to best mitigate them (e.g., patient partners may have physical limitations that make it difficult to use small keyboards or read small print). • Recognize that patients and clinicians often speak two different languages (e.g., medical/research versus lay language) and may need facilitators who speak both languages to make sure that patient input is considered. It may be useful to provide patient partners with education in medical and research terminology relating to the registry.

It is also important to prepare patients for team meetings and to conduct them regularly. Patients should have the opportunity to ask questions before and during meetings (particularly for those who are unfamiliar with research). Providing pre-meeting information materials will promote comfort with the topic and enable informed discussion. Team building exercises such as premeeting "icebreakers" also may be useful, especially when engaging stakeholders with differing experiences/perspectives. It is also important to create an environment where patients feel comfortable speaking up at registry team meetings.

The Arthritis Patient Partnership with Comparative Effectiveness Researchers Registry (AR-PoWER PPRN) is an example of a registry that includes patient partners in registry governance. The registry established a governance structure in collaboration with investigators and patient community members that includes an Executive Board, a Patient Governor Group, and a Research Advisory Board[11] as follows:

- *The Executive Board* includes the PI, one Co-PI, and the Executive Director and a Board Member of the Global Healthy Living Foundation, a nonprofit disease advocacy organization for arthritis patients that is a founding partner organization in the AR-PoWER Registry.

- *The Patient Governor Group* is a patient steering committee composed of 11 patients with conditions relevant to the research mission of the registry from a variety of professional, geographic, and demographic backgrounds. This group reviews and selects research topics to advance for further consideration based on priority of the evidence need.
- *The Research Advisory Board* is composed of 10 rheumatology researchers from the University of Alabama at Birmingham. It reviews topics for feasibility and funding.

The Patient-Governor Group and the Research Advisory Board Chairs present recommendations to the Executive Board for approval of projects to be conducted in AR-PoWER. The registry is part of PCORI's Patient-Centered Clinical Research Network (PCORnet), and stakeholder engagements follow the PCORI Engagement Rubric, focusing on the six principles: reciprocal relationships, co-learning, partnerships, transparency, honesty, and trust.[12]

4.2 Public-Private Partnerships[ii]

"Public-private partnership" (PPP) is a broad term that refers to any partnership in which at least one entity is a public agency (e.g., a government entity) and at least one other entity is a private organization. As both government and private groups have shown increased interest in patient registries, PPPs have become more prevalent as a means to develop and support patient registries and data linkage projects. The scope can range from partnerships at the local level, including local and regional health agencies, to national and international health agencies and other private institutions or organizations (e.g., professional associations, patient advocacy groups). In a research context, a partnership implies some joint collaboration to achieve a common scientific goal. Partners may contribute intellectual capital, funding, data, or other services.

PPPs may take many forms. Some possible models include: partnerships among federal agencies to examine safety and effectiveness; partnerships among health agencies from several countries on an international level to describe the clinical course of a disease and understand whether there are any effective treatments; partnerships with state agencies for quality improvement; and partnerships for evidence development for coverage decisions. Regardless of the structure and purpose of a PPP, effective and transparent governance is necessary to enable the registry to achieve its objectives, as described in this chapter.

More information is provided in Chapter 24 of the third edition of the User's Guide.[13] Case Example 19 describes a PPP to develop a registry.

5. Key Principles for Successful Governance

In summary, some principles to guide successful governance strategies are as follows:

[ii] *Adapted from Dreyer NA, Eloff BC, Kirklin JK, et al. 'Public-Private Partnerships.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

- All aspects of governance should be codified in a written format that can be reviewed, shared, and refined over time.
- The value of strong leadership is paramount. The registry leadership will need to have sufficient time and skills to manage the operational and scientific aspects of developing and operating a registry and the equally challenging task of leading an interdisciplinary team with diverse interests towards a common goal.
- Mutual respect among all partners is necessary for a strong working relationship.
- Transparency regarding any perceived or actual conflicts of interest is important for effective governance. A plan for identifying and managing actual and perceived financial or intellectual COIs should be developed at the registry outset.
- The expectations of each research partner (e.g., stakeholders, clinicians, etc.,) should be explicitly delineated, pragmatic, and transparent. The registry should aim to return value to all partners, and such value will be different for each partner. Meetings with research partners early in the planning phase are particularly useful to set appropriate expectations. At this time, roles and responsibilities can be clearly defined and mutually agreed upon so that all stakeholders benefit equally, and so that the respective roles are understood before the first governance committee meeting.[14,15]
- Registry partnership agreements and legal contracts should clearly specify who will have ownership/access rights to the registry data, and if and how external access to data will be considered and provided. In many registries, the registry sponsor owns the registry data. However, registries may have multiple sponsors and multiple stakeholders. Establishing clarity here will help guide data use and external access policies.
- Policies and procedures should be developed to support stakeholder engagement and transparency. These may include published charters for any formal advisory boards, public websites with information about the registry, periodic public stakeholder meetings, newsletters, and email listservs.

References for Chapter 9

1. Dreyer N, Toovey S, Oner A, et al. Investigating outbreaks of novel infectious disease: an international case study. Journal of Clinical Studies. 2013;5:52-3.

2. Cross-Border Patient Registries Initiative. Methodological guidelines and recommendations for efficient and rational governance of patient registries – version 1.0. https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20150512_co013_en.pdf. Accessed June 10, 2019.

3. Reich MR, ed. Public-Private Partnerships for Public Health. Cambridge, MA: Harvard Center for Population and Development Studies; 2002. https://cdn1.sph.harvard.edu/wp-content/uploads/sites/480/2012/09/Partnerships_book.pdf. Accessed June 10, 2019.

4. Nikolic IA, Maikisch H, Health Nutrition. An Overview with Case Studies from Recent European Experience. Washington, D.C.: The International Bank for Reconstruction and Development/The World Bank; 2006. Population (HNP) Discussion Paper Public-Private Partnerships and Collaboration in the Health Sector.

5. Wagner JA, Prince M, Wright EC, et al. The Biomarkers Consortium: practice and pitfalls of open-source precompetitive collaboration. Clin Pharmacol Ther. 2010;87(5):539-42. PMID: 20407460. DOI: 10.1038/clpt.2009.227.

6. Goodman M, Almon L, Bayakly R, et al. Cancer outcomes research in a rural area: a multi-institution partnership model. J Community Health. 2009;34(1):23-32. PMID: 18850070. DOI: 10.1007/s10900-008-9123-7.

7. Omobowale EB, Kuziw M, Naylor MT, et al. Addressing conflicts of interest in Public Private Partnerships. BMC Int Health Hum Rights. 2010;10:19. PMID: 20615242. DOI: 10.1186/1472-698X-10-19.

8. HIV-related Public-Private Partnerships and Health Systems Strengthening. Joint United Nations Programme on HIV/AIDS (UNAIDS); Geneva, Switzerland: 2009.

9. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication. http://www.icmje.org. Accessed June 10, 2019.

10. Gliklich RE, Dreyer NA, Leavy MB, Christian JB (eds). 21st Century Patient Registries. EBook addendum to Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition. (Prepared by L&M Policy Research, LLC under Contract No. 290-2014- 00004-C.) AHRQ Publication No. 17(18)-EHC013-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018. www.effectivehealthcare.ahrq.gov.

11. Nowell WB, Curtis JR, Crow-Hercher R. Patient Governance in a Patient-Powered Research Network for Adult Rheumatologic Conditions. Med Care. 2018;56 Suppl 10 Suppl 1:S16-S21. PMID: 30074946. DOI: 10.1097/MLR.0000000000000814.

12. Concannon TW, Meissner P, Grunbaum JA, et al. A new taxonomy for stakeholder engagement in patient-centered outcomes research. J Gen Intern Med. 2012;27(8):985-91. PMID: 22528615. DOI: 10.1007/s11606-012-2037-1.

13. Gliklich R DN, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014. http://www.effectivehealthcare.ahrq.gov.

14. Ciccone DK. Arguing for a centralized coordination solution to the public-private partnership explosion in global health. Glob Health Promot. 2010;17(2):48-51. PMID: 20587631. DOI: 10.1177/1757975910365224.

15. Centers for Medicare & Medicaid Services. Guidance for the Public, Industry, and CMS Staff. Coverage with Evidence Development. November 20, 2014. https://www.cms.gov/medicare-coverage-database/details/medicare-coverage-document-details.aspx?MCDId=27. Accessed June 4, 2019.

Case Examples for Chapter 9

Case Example 19. Developing a Public-Private Partnership for Comparative Effectiveness Research

Description	The Registry In Glaucoma Outcomes Research (RiGOR) was a prospective observational study designed to assess comparative effectiveness of medications, laser trabeculoplasty and incisional surgery in patients with open-angle glaucoma.
Sponsor	Agency for Healthcare Research and Quality (AHRQ)
Year Started	2011
Year Ended	2013
No. of Sites	45 community and academic ophthalmologic practices
No. of Patients	2,597

Challenge

In 2009, the Institute of Medicine disseminated a landmark report, "Initial National Priorities for Comparative Effectiveness Research," that listed research priorities for the newly enacted American Recovery and Reinvestment Act (ARRA). Among the 100 priority research topics identified was evaluating the different treatment strategies for primary open-angle glaucoma (OAG). Since the disease disproportionately affects African-Americans, understanding the effectiveness of treatment strategies in minority populations was also of special interest. With ARRA funding, AHRQ sought to develop high-quality scientific evidence to inform decisionmaking by clinicians and patients. An approach was needed to obtain continued and expanded input from the various stakeholders while addressing existing evidence gaps.

Proposed Solution

A diverse group of stakeholders was assembled to implement the registry, provide scientific guidance, develop dissemination plans, and further key research based on study findings. The principal investigator and co-principal investigators represented AHRQ, the American Academy of Ophthalmology (AAO), the University of California at Los Angeles (UCLA) Jules Stein Eye Institute, and the Outcome DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Center. AHRQ provided oversight and financial support to the project, with scientific leadership from the principal and co-principal investigators; the Outcome DEcIDE Center managed the operational aspects of the study; AAO and UCLA engaged sites and investigators and provided guidance on clinical issues. The stakeholder committee was comprised of individual clinical advisors and representatives from the Glaucoma Research Foundation, American Glaucoma Society, National Medical Association, and State-level healthcare organizations.

Developing the study protocol, initiating startup activities and decisions, and analyzing and reporting the findings required continued communications among all stakeholders. A communication plan was developed to outline project team roles and organizational structures for each stakeholder. Regular stakeholder committee meetings created a forum to discuss design

issues, share study status, solicit input on unexpected challenges, and discuss research plans. Site- and patient-recruitment efforts were designed to maximize geographic diversity and enrollment of minority populations. Quarterly study newsletters and investigator meetings coinciding with the AAO annual meeting were also implemented to maintain site interest.

Results

Launched in 2011, RiGOR was a prospective, observational, cohort study designed to compare the effectiveness of treatment strategies for primary OAG. Different treatment strategies studied in the registry included laser surgery, other procedures (such as incisional surgery or other glaucoma procedures), and medications. All treatment decisions were made at the discretion of the treating physician according to their usual practice. Data collection included patient demographics, medication, visual measures, glaucoma severity, surgical characteristics, adverse events, and patient-reported outcomes, and occurred at baseline, 3 months, 6 months, and 12 months.

The registry completed data collection in 2013. Registry data were used to examine the effectiveness of different treatment strategies and to describe practice patterns and treatment changes and quality of life and visual acuity outcomes in patients with OAG. The registry also met its objective of enrolling a high percentage of minority patients. Of the total enrolled patients, 21.5% were African-American and 8.3% were of Hispanic or Latino ethnicity. Study analyses provided valuable information regarding visual function and treatment patterns among different racial/ethnic populations.

Key Point

The public-private partnership model can be an effective approach to engaging multiple stakeholders in an effort to address a comparative effectiveness research question. When working with multiple stakeholders, it is critical to clearly identify roles and communicate regularly with all stakeholders to address any design, operational, or analytical issues, solicit input from all contributors, share study findings, and maintain stakeholder engagement.

For More Information

- Coleman AL, Lum FC, Velentgas P, et al. Impact of treatment strategies for open angle glaucoma on intraocular pressure: the RiGOR study. J Comp Eff Res. 2016;5(1):87-98. PMID: 26691269. DOI: 10.2217/cer.15.58.
- Coleman AL, Lum FC, Velentgas P, et al. Practice patterns and treatment changes for open-angle glaucoma: the RiGOR study. J Comp Eff Res. 2016;5(1):79-85. PMID: 26691017. DOI: 10.2217/cer.15.57.
- Coleman AL, Lum FC, Gliklich RE, et al. Quality of life and visual acuity outcomes in the Registry in Glaucoma Outcomes Research study. J Comp Eff Res. 2016;5(1):99-111. PMID: 26691427. DOI: 10.2217/cer.15.59.
- Coleman AL, Lum FC, Velentgas P, et al. RiGOR: a prospective observational study comparing the effectiveness of treatment strategies for open-angle glaucoma. J Comp Eff Res. 2016;5(1):65-78. PMID: 26690903. DOI: 10.2217/cer.15.56.

Case Example 20. Developing a Patient-Centered Governance Structure

Description	ArthritisPower®, the ARthritis Patient Partnership with Comparative Effectiveness Researchers (AR-PoWER) Patient-Powered Research Network (PPRN) is a patient-led, patient-centered registry designed to support comparative effectiveness research in rheumatic and musculoskeletal disease, or RMD. The registry is part of PCORnet, the National Patient-Centered Clinical Research Network, and is a collaborative project involving the Global Healthy Living Foundation, the CreakyJoints arthritis patient community, and rheumatology researchers from the University of Alabama at Birmingham.
Sponsor	Funded by the Patient-Centered Outcomes Research Institute (PCORI)
Year Started	2015
Year Ended	Ongoing
No. of Sites	N/A
No. of Patients	>18,000

Challenge

In 2014, PCORI invested in the development of a collaborative of PPRNs to support patient-centered comparative effectiveness research. By definition, PPRNs are ‘powered’ by patients and/or caregivers, which suggests that patient stakeholders should be active participants in the registry governance. However, at the time of the development of ArthritisPower, limited information was available about how best to engage patients in registry governance.

Proposed Solution

The co-PIs designed the governance structure in collaboration with patient community members. ArthritisPower established a governance structure made up of an Executive Board, a Patient Governor Group (PGG), and a Research Advisory Board (RAB). The Executive Board is composed of the Scientific Co-Principal Investigator, the patient organization executive director, and the research advisor; the Patient Co-Principal Investigator serves as the chair of the Executive Board. The PGG is the patient steering committee and is composed of 10-12 patient members of ArthritisPower who come from a variety of professional, geographic, and demographic backgrounds with conditions relevant to the research mission of the registry. The PGG collects feedback from the arthritis patient community using surveys and the “Send Feedback” function on the registry’s website and mobile application. The RAB is made up of researchers, clinicians, industry representatives and includes the PGG patient chair.

Results

Since its launch in 2015, ArthritisPower has built online tools and a mobile application to engage and educate patients and to collect information on diagnosis, symptoms, and medications. The program has captured data on over 18,000 patients and produced numerous publications and presentations.

Key Point

Patients can play an active role in registry governance, particularly in registries that are designed to enroll patients directly and support patient-centered research.

For More Information

- https://arthritispower.creakyjoints.org/.
- Nowell WB, Curtis D, Thai M, et al. Digital Interventions to Build a Patient Registry for Rheumatology Research. Rheumatic diseases clinics of North America. 2019;45(2):173-86. PMID: 30952391.
- Nowell WB, Curtis JR, Crow-Hercher R. Patient Governance in a Patient-Powered Research Network for Adult Rheumatologic Conditions. Med Care. 2018;56 Suppl 10 Suppl 1:S16-s21. PMID: 30074946 PMCID: PMC6143204.

Section III

Operating Registries

Chapter 10. Recruiting and Retaining Participants in the Registry

1. Introduction

Recruitment and retention of participants are essential elements in the design and operation of a registry. Registries are often intended to be representative of a certain population of patients and reflective of the practices of certain providers and geographic areas. The problems commonly associated with clinical studies—such as difficulties with patient enrollment, losses to followup, and certain sites contributing the majority of patients—can also have profound consequences on validity of registry data. When registry patients are not representative of the target population, the value of the results is diminished. For example, in regard to policy determinations, the enrolled sites or providers must be representative of the types of sites and providers to which the policy determination would apply in order for the results of the registry to be generalizable. Differences in how effectively sites enroll or follow patients can skew results and overly reflect the sites with the most data. This oversampling within a particular site or location must also be considered in sample size calculations. If the sample size of a key unit of analysis (patient, provider, or institution) is not sufficient to detect a clinically important difference, the validity of the entire registry is weakened. (See Chapter 3 and Chapter 13.)

Well-planned strategies for enrollment and retention are critical to avoiding these biases that may threaten registry validity. Because registries typically operate with limited resources and with voluntary rather than mandatory participation, it is particularly important to balance the burdens and rewards of participation in the registry. The term “voluntary” in this context is intended to mean that participation in the registry by either providers or patients is not mandated (e.g., by the U.S. Food and Drug Administration), nor is participation required as a necessary condition for a patient to gain access to a healthcare product or for a provider to be eligible for payment for a healthcare service. Registries that are not voluntary have different drivers for participation.

In general, the burden of participation should be kept as low as possible, while the relative rewards, particularly nonmonetary rewards, should be maximized. As described in Chapters 2 and 5, minimizing burden typically starts with focusing on the key goals of the registry.

Building participation incentives into a registry should also be included in the planning phase. A broad range of incentives—spanning a spectrum from participation in a community of researchers, to access to useful data or quality improvement benefits, to continuing medical education, to public recognition or certification, to payments or access to patients—have been used in registries. The ability to offer certain incentives (e.g., linking payment for a service to participation in a registry or access to patients) may be available only to certain registry developers (e.g., payers, licensing entities). Many registries incorporate multiple types of incentives, even when they pay for participation. Monetary incentives (e.g., from payers or sponsors) are very helpful in recruiting sites. However, because the payments should not exceed fair market value for work performed, registries cannot solely rely on these incentives. Many nonmandated registries have achieved success in recruitment and retention by providing a

combination of ethical incentives that are tailored to and aligned with the specific groups of sites, providers, and patients that are asked to participate. (See Case Examples 21 and 22.).

2. Recruitment

Depending on the purpose of a registry, recruitment may occur at any of three levels: facility (e.g., hospital, practice, and pharmacy), provider, or patient. While recruitment at these levels is frequently part of a design to accrue a sufficient number of patients for sample size purposes, such as for a safety registry, the individual levels may also constitute potential units of analysis (and as such, may further affect sample size, as discussed in Chapter 3). As an example, a registry focused on systems of care that is examining both hospital system processes and patient outcomes might need to consider characteristics of the individual patients, the providers, and/or the places where they practice (i.e., clusters). If the question is about the practices of orthopedic surgeons in the United States, the registry will be strengthened by describing the number and characteristics (e.g., age, gender, and geographic distribution) of U.S. orthopedic surgeons, perhaps by citing membership data from the American Academy of Orthopedic Surgeons. This will allow documentation of the similarities and differences in the characteristics of the surgeons participating in the registry compared with the target population. (See Chapter 3.)

2.1 Hospital Recruitment

A hospital or health system may choose to participate in a patient registry for many reasons, including the research interest of an individual investigator or champion, the ability of the hospital to achieve other goals through the registry (such as requirements for reimbursement, certification, or recognition), or the general interest of the institution in the disease area (e.g., specialty hospitals). Increasingly, external mandates to document compliance with practice standards provide an incentive for hospitals to participate in registries that collect and report mandatory hospital performance or quality-of-care data. For example, some registries enable providers to participate in the Centers for Medicare and Medicaid Services' Merit-Based Incentive Payment System (MIPS). In these cases, hospital administrators may be willing to supply the staff time to collect these data without additional financial incentives from the registry sponsor, provided that registry participation meets the criteria of these external programs. In other cases, participation in a quality monitoring or health system surveillance registry may be required by payers or governments for reimbursement, differential payments, or patient referrals under various programs (e.g., centers of excellence programs, pay-for-performance programs). One particular example, CMS's Coverage With Evidence Development programs,[1] which may require participation in a registry for the center or provider to qualify for payment for a procedure, can have a dramatic impact on registry participation. Registry participation requirements have existed for implantable cardioverter defibrillators for preventing sudden cardiac death in heart failure, for bariatric surgery, for positron emission tomography scan use in cancer, and for other procedures and devices. These requirements have rapidly resulted in high participation rates for registries meeting them.

The presence of quality assurance departments in U.S. hospitals provides an infrastructure for participation in many hospital-based registries, and these departments are therefore a natural target for recruiting. However, hospital size, service line (e.g., disease-specific centers), and competing activities may limit institutional interest. The American Hospital Association database

provides a valuable resource for identifying hospitals by key characteristics, including hospital ownership, number of beds, and the presence of an intensive care unit.

Table 10-1 summarizes the key factors for successful hospital recruitment and lists specific methods that might be used to recruit hospitals. While programs need not incorporate all of these characteristics or use all of these methods, successful programs typically incorporate several.

Table 10-1. Hospital recruitment

Keys to hospital recruitment	• The condition being studied satisfies one of the hospital's quality assurance mandates. Sufficient funds, data, or other benefits will be realized to justify the effort required to participate. • The confidentiality of the hospital's performance data is ensured, except to the extent that the hospital elects to report it. • Clinically relevant, credible, timely, and actionable self-assessment data—ideally, data that are risk adjusted and benchmarked—are provided back to the hospital to help it identify opportunities for enhancing patient care outcomes. • High-profile hospitals (regional or national) are participating in the registry. • Burden is minimized (e.g., data collection fits within the hospital workflow; existing data sources are used to the extent feasible). • Participation assists the hospital in meeting coverage and reimbursement mandates, gaining recognition as a center of excellence, or meeting requirements for pay-for-performance initiatives.
Methods of hospital recruitment	• Identify eligible hospitals from the American Hospital Association database. • Use stakeholder representatives to identify potentially interested hospitals. • Enroll hospitals through physicians who work there and are interested in the registry. • Use invitation letters or calls to directors of quality assurance or the chief of the clinical department responsible for the condition targeted by the registry. • Ask physician members of an advisory board (if applicable) to network with their colleagues in other hospitals. • Reach out to physicians or hospital administrators through relevant professional societies or hospital associations. • Leverage mandates by external stakeholders, including third-party payers, health plans, or government agencies.

2.2 Physician Recruitment

A physician practice may or may not choose to participate in a voluntary registry for many reasons. As with hospitals, these reasons can include the research interests of the physician and the ability of the practice to achieve other goals through the registry (such as reimbursement or recognition). When deciding to participate, physicians often focus on several concerns:

- *Relevance*: Does the registry have meaning for the practice and patients?
- *Trust*: Are the registry leaders credible? Are the goals clearly stated?
- *Risks*: Will confidentiality be maintained? Are patient records secure?
- *Effort*: Will the amount of effort expended be fairly compensated?

- *Disruption*: Will participation disrupt workflow of the staff?
- *Value*: What benefits will be derived from participation? Will it improve the care provided? Will it enhance the evidence base for future practice?

Physicians who manage only a few patients per year with the disease that is the subject of the registry are less likely to be interested in enrolling their patients than physicians who see many such patients—unless the disease is rare or extremely rare, in which case the registry may be of great interest.

Because most registries are voluntary and physicians in nonacademic practice settings may have less infrastructure and staff available to enroll their patients, recruitment of representative physicians is a major challenge for registries that aim to compare physician practices across a full spectrum of practice settings. In general, community-based physicians are less well equipped than hospital-based or academic physicians to collect data for research studies because they work in busy practices geared to routine clinical care rather than research. To increase recruitment of nonacademic physicians, it can be helpful to clearly explain the purpose and objectives of the registry; how registry data will be used; how much staff time is required; and, specifically, that individual results will not be shared (except at the direction of the physician) or published, and that registry outcomes data will be released only in large aggregates that protect the identities of individual hospitals, physicians, and patients. In addition, any incentives should be clearly articulated.

Table 10-2 describes the key factors for successful physician recruitment and lists several methods that might be used for recruiting physicians.

Table 10-2. Physician recruitment

Keys to physician recruitment	• The condition being studied is part of the physician's specialty. • The registry is a valuable scientific endeavor. • The registry is led by respected physician opinion leaders. • The registry is endorsed by leading medical, government, or patient advocacy organization(s). • The effort needed to recruit patients and collect and submit data is perceived as reasonable. • Useful practice pattern and/or outcome data are provided. • The registry meets other physician data needs, such as maintenance of certification requirements, credentialing requirements, or quality-based, differential, reimbursement payment programs (pay-for-performance).
Methods of physician recruitment	• Purchase mailing lists from physician specialty organizations. • Ask opinion leaders in the field to suggest interested colleagues. • Partner with local and national medical societies or large physician hospital organizations. • Use stakeholder representatives to identify interested physicians. • Recruit and raise awareness at conferences. • Advertise using email and the Web. • Leverage practice-based research networks.

2.3 Vetting Potential Participants

Once potential hospital or physician participants have been identified, it is important to vet them to ensure that the registry is gathering the appropriate mix of data. Issues to consider when vetting potential participants include—

- Representativeness
- Hospital characteristics (e.g., bed size, geographic location)
- Physician characteristics (e.g., specialty training)
- Practice setting (health maintenance organization [HMO], private practice)
- Ability to recruit patients
- Volume of target cases
- Internal resources
- Availability of a study coordinator
- Availability of Internet connectivity for studies with electronic data capture
- Prior performance, including reliability and accuracy of data entry

2.4 Patient Recruitment

Patients may be recruited based on the judgment of the physician who provides their care; the diagnosis of a disease; receipt of a procedure, operation, device, or pharmaceutical; membership in a health insurance plan; or membership in a group of individuals who have a particular exposure. Recruitment of patients by the physician who is providing their care is one of the most successful strategies. Since registries should not modify the usual care that physicians provide to their patients, there should be little or no conflict between their role of physician and that of participant in the registry. (See Chapter 7.) In addition, patients may see participation in the registry as an opportunity to increase their communication with their clinician. Another incentive for many patients is the feeling that they are contributing to the knowledge base of sometimes poorly understood and undertreated conditions.

While many registries recruit patients through a healthcare provider, some registries recruit patients directly. In this study design, patients are contacted via Internet, mobile applications, mailed surveys, telephone, face-to-face interview or other means and invited to participate. This method is most useful for registries in which patients can be reasonably good and complete reporters about the exposures and outcomes of interest and when recruitment through medical care providers is particularly challenging. For example, patients may be better reporters than medical care providers when the research questions concern personal habits, exposures, or quality of life. Patients may also feel comfortable reporting sensitive information (e.g., recreational drug use or sexual activity) for research purposes that they may be unwilling to share honestly with medical care providers.[2,3]

Recruitment of patients presents different challenges, depending on the nature of the condition being studied. In general, patient recruitment plans should address the following questions:

- Does the plan understand the needs and interests of potential participants?
- Does the plan address patient recruitment issues and procedural challenges, including informed consent and explanation of risks?
- What are the patient retention goals? What is a reasonable followup period? What is a reasonable followup rate? When does reduced retention compromise validity?
- What, if any, patient incentives are offered, including different types of incentives and the ethical, legal, or study validity issues to be considered with patient incentives?
- What are the costs of patient recruitment and retention?

Table 10-3 summarizes the key factors for patient recruitment and lists several specific methods that might be used for recruiting patients, grouped by the basic categories of patients at the time of recruitment.

Table 10-3. Patient recruitment

Keys to patient recruitment	• Communicate to the patient that registry participation may help to improve care for all future patients with the target condition. • Keep the survey forms short and simple. • Write all patient materials (brochures, consent forms) in a manner that is easily understandable by the lay public. • Provide incentives. These can be nonmonetary, such as functions relevant to the patient's care (reports) or community (newsletters, portals). In some cases, monetary incentives can be offered if approved by the institutional review board. • Actively plan how to include minorities or other populations of interest.
Methods of patient recruitment	• Noninstitutionalized residents of the general U.S. population: o Recruit via letter survey, telephone, or email. o Recruit during well-patient visits to outpatient clinics. o Recruit via patient advocacy and support groups, health information websites, mobile applications, social media, etc. • Outpatients attending the clinic of a physician who is participating in the registry: o Recruit through the patient's physician. o Recruit via brochures placed in physician's office. • Hospital inpatients who are hospitalized for treatment of a condition that is the subject of the registry: o Recruit through the patient's physician. o Recruit through hospitalists or consultant specialists. o Recruit through a hospital research coordinator. • Residents of nursing homes and similar long-term care facilities: o Establish a relationship with the nursing home and staff.

2.5 Partnerships To Facilitate Recruitment

Many agencies/organizations can assist in the recruitment of physicians and patients. These partners may have access to patients or their families and physicians who treat the condition, and they may lend credibility to the effort. These agencies/organizations include—

- Government agencies
- Physician professional associations or State medical associations
- Certifying boards (e.g., American Board of Neurological Surgeons)
- Patient advocacy groups (e.g., Muscular Dystrophy Association)
- Nonprofit foundations (e.g., Robert Wood Johnson Foundation)
- Industry (e.g., pharmaceutical companies)
- HMOs and other third-party insurance providers

2.6 Procedural Considerations Related To Recruitment

When developing a recruitment plan for a registry, consideration should be given to the procedural concerns that may be factored into potential participants' decisions. These concerns include the roles and responsibilities of each party, the need and process for obtaining institutional review board (IRB) approval, and the management of patient and provider confidentiality.

The contract between registry sites and the sponsor or coordinating center should clearly state the roles and responsibilities of the participants, the registry-coordinating center, and the sponsor. If remuneration is being offered, the data-entry requirements that need to be fulfilled before payments are made should be stated. It is often helpful to explain to sites the concept of fair market value. Where the Health Insurance Portability and Accountability Act (HIPAA) applies, consideration should also be given to new requirements in the HIPAA Privacy Rule that prohibit a covered entity from disclosing protected health information in exchange for remuneration from the recipient of that information without the individual's authorization, subject to a number of exceptions (see 45 CFR 164.502(a)(5)(ii)). There is no specific formula (such as whether to separate startup payments from per-patient payments), but total remuneration must reflect work effort for the specific registry. Some individual factors, ranging from location to specialty, may have a bearing on fair market value. It is also important to spell out which entity will have ownership of the data and how the data will be used.

The contract should clearly explain the registry policy regarding any necessary approvals. If review by an IRB is required, generic templates can be offered to participants to assist them in obtaining ethical and IRB approval. Because the costs of obtaining IRB approval are often substantial, it is essential that the contract with the participants clearly indicate which party is responsible for bearing this cost. If the registry developer believes that IRB or privacy board review or approval is not required or may be waived, then a clear rationale should be provided to the prospective participants. As discussed in Chapter 7, the research purpose of the registry, the type of entity that creates and maintains it, whether the Common Rule applies to the particular site, and the extent to which the data are individually identifiable largely determine which regulatory requirements apply. For example, for registries limited to certain purposes, such as quality improvement, institutions may not need IRB approval.[4]

Patient privacy and participant confidentiality should be addressed in the registry materials. Methods of ensuring patient privacy need to be clearly elucidated in all registry-related

documentation. Case report forms and patient logs must be designed to minimize patient identification (such as by transmitting limited datasets rather than more identifiable information, if such information is not required to meet a registry objective).

The intended management of the confidentiality of participating providers should be explained in the contract. Mechanisms for protecting provider confidentiality, including Certificates of Confidentiality and Patient Safety Organizations, are discussed in Chapter 7. If third-party or public reporting is an intended component of the registry, the specific data to be shared, the level of the disclosure (e.g., hospital and/or physician level), and the permitted receiving entities need to be articulated and the control mechanisms explained.

3. Retention

3.1 Providers

Once hospitals and physicians are recruited to participate in a registry, retaining them becomes a key to success. The factors identified as important for recruitment are important for retention as well. A critical factor in retention is delivery on promises made during recruitment (e.g., that the burden of participation is low). By carefully pilot testing all aspects of the registry prior to full recruitment, there is less likelihood that problems will arise that threaten the registry's reputation. Registries with an advisory board or steering committee can use this resource to help with retention (see Chapter 9). A visible and independent advisory board adds transparency and credibility, sets appropriate expectations among its peers on what to expect from a registry (e.g., compared with a clinical trial), ensures that the burden of the registry is minimized (or at least never outweighs its value to participants), and maintains the relevance and currency of the registry for the investigators. Ideally, advisory board members serve as ambassadors for the program. The level of credibility, engagement, practicality, and enthusiasm of the advisory board can significantly affect provider recruitment and retention. For example, an advisory board whose clinical members are not themselves participating in the registry will have greater difficulty than a board with participating members in addressing the concerns of participating practices that invariably arise over the course of the registry. Including patients or patient advocacy organization representatives on the advisory board also can support patient retention efforts. These representatives can provide feedback to the board on patient issues or concerns about the registry, and they can facilitate communication about the registry's purpose and value to their peers or members.

Throughout the registry's duration, communication from the data coordinating center and the advisors, as well as community building, are important for strong retention. Early and continued engagement of the site champions or principal investigators is very important. Some registries use periodic face-to-face meetings of principal investigators from participating sites. When this approach is not economically feasible, well-planned online meetings can serve the same purpose.

Visibility of the registry at relevant national meetings can help maintain clinician awareness and sense of community, and regular demonstration of its value through presentations and publications reinforces the credibility of the registry to its participants. As the dataset grows, so too does the value of the registry for all participants, and regular updates on the registry growth can be important. Finally, enhancing site value through nonfinancial rewards can be particularly

useful in retention, and the registry should continually seek to bring value to the participants in creative and useful ways.

Participation retention tools include:

- Websites
- Newsletters
- Telephone helplines
- Instruction manuals
- Training meetings
- Site audit/retraining visits
- Customer satisfaction/opinion surveys
- Regular data reports to stakeholders
- Presentations at conferences
- Regular reports to registry participants on registry growth and publications
- Ability of participating physicians to publish based on registry data (depending on the data access policy of the registry)

3.2 Patients

Retaining patients as active participants in registries with longitudinal followup is an ongoing challenge. Many factors need to be considered in developing a retention plan, including how long the patient is likely to return to the enrolling site. Patients enrolled in a primary care practice for a chronic illness can likely be followed in that practice for some time, although there should be a plan for how the registry will (or will not) address the issue of patients who transfer to unenrolled practices.

Different approaches are needed for patients who are unlikely to return to the medical care provider after an initial encounter or treatment, such as after a surgical procedure, medical device implantation or vaccination. Options include enlisting site staff to reach out to patients beyond their standard interactions,[5] following patients directly through a central patient management center,[6] and linking to other data sources (e.g., claims data) to obtain key long-term outcomes data on patients who are lost to followup. Retention plans, including contingencies, should be considered during registry planning, as they may require additional permissions (e.g., for direct contact) or data elements (e.g., for linkage).

Maintaining ethical incentives for patient participation (ranging from newsletters to payments) is also important for some registries (e.g., those that collect patient-reported outcomes data). Beyond planning for how to retain patients in a registry, it is important to track actual versus expected followup rates over time and to respond if rates are not meeting expectations. The resources available for patient retention efforts should also be clear. Followup rates can often be improved with more efforts, such as more attempts to contact the patient, but these efforts add costs and, at some level, will yield diminishing returns.

Additionally, some patient advocacy organizations may already have established networks or patient communities that contain members who are eager to participate in research programs and who are likely to complete any studies in which they enroll. One example would be the Cystic Fibrosis Foundation. Patient-Powered Research Networks (PPRNs), established by the Patient-Centered Outcomes Research Institute (PCORI) as authorized by the Affordable Care Act, offer a myriad of patient registries, including Improve Care Now: A Learning System for Children with Crohn's Disease and Ulcerative Colitis[7] and the Multiple Sclerosis Patient Powered Research Network[2] to name just two examples. Patients within these networks/communities share their experiences with each other and may be interested in participating in patient registries. A patient organization may even sponsor a registry among its community members, often referred to as patient-generated registries, which may be used for recruitment in studies organized by others. The various characteristics that motivate these patients to participate in a study may make them different from other patients with the disease or exposure of interest, and researchers must consider the generalizability of results. However, those motivating factors may not necessarily interfere with any biologic relationships under study and may enhance retention.

Patient advocacy groups or individual patient partners who participate in the registry design may also be able to help design recruitment and retention procedures that appeal to potential participants. These partners may be better positioned to understand potential barriers from the patient perspective and, as trusted members of their respective communities, can ensure that the recruitment procedures are sensitive to the needs of the patients. These partners can be "champions" for the registry and serve as a liaison between researchers and patients, helping to identify and recruit patients and explaining the study to potential participants in lay terms. Their efforts can also include preparing and/or reviewing patient outreach materials, identifying more effective methods to increase awareness of the study in the community of patients, and developing new methods to recruit difficult-to-reach patients.[8-12]

4. Pitfalls in Recruitment and Retention

Pitfalls abound in recruitment and retention. The most important of these pitfalls is the risk of selection bias. Targeting hospital-based or academic physicians to the exclusion of community-based physicians is tempting because the former are often more accessible and are frequently more open to involvement in, and more experienced in, research projects. Similarly, targeting high-volume practices or centers will improve efficiency of patient enrollment, but may not yield an adequately representative sample of care practices. If an advisory board or committee is used to help design the registry and aid in recruitment, there may be a tendency for advisors to recruit known colleagues or to target disease experts, when a wider range of participants may be necessary to provide the appropriate data to meet the research goals. Including representatives from the range of anticipated site types on the advisory board can be helpful.

Even with an appropriate mix of physician participants in a registry, biases in patient recruitment may still occur. For example, older and more seriously ill patients may be excluded because of challenges in enrollment and followup or poorer outcomes. From the outset, physicians involved in recruitment efforts need to be aware of the potential for bias, and they must understand the importance of adhering to well-delineated inclusion and exclusion criteria. They must also adhere to the registry's enrollment strategy, which is typically designed to reduce this bias (e.g., sequential enrollment). In addition, overly demanding data collection requirements can affect

retention. The schedule should be designed to obtain relevant data in a timely fashion without overtaxing the resources of patients and providers. It is also important to consider approaches that will distinguish patients who are lost to followup from those who have missing data for other reasons (such as a patient who missed a visit but is still in the registry).

Another major pitfall is confusing terminology. This can be a major problem when the registry is international. When designing training materials, instruction manuals, and questionnaires, it is critical that the language and terminology be clear and concise. Materials that are translated into other languages must undergo strict quality assurance measures to ensure that terms are translated properly (e.g., back translation).

References for Chapter 10

1. Centers for Medicare & Medicaid Services. Guidance for the Public, Industry, and CMS Staff. Coverage with Evidence Development. November 20, 2014. https://www.cms.gov/medicare-coverage-database/details/medicare-coverage-document-details.aspx?MCDId=27. Accessed June 4, 2019.

2. Multiple Sclerosis Patient-Powered Research Network. https://www.pcori.org/research-results/2015/multiple-sclerosis-patient-powered-research-network-iconquerms%E2%84%A2. Accessed June 10, 2019.

3. Fleurence RL, Curtis LH, Califf RM, et al. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc. 2014;21(4):578-82. PMID: 24821743. DOI: 10.1136/amiajnl-2014-002747.

4. Dokholyan RS, Muhlbaier LH, Falletta JM, et al. Regulatory and ethical considerations for linking clinical and administrative databases. Am Heart J. 2009;157(6):971-82. PMID: 19464406. DOI: 10.1016/j.ahj.2009.03.023.

5. Gheorghiade M, Abraham WT, Albert NM, et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA. 2006;296(18):2217-26. PMID: 17090768. DOI: 10.1001/jama.296.18.2217.

6. Spertus JA, Peterson E, Rumsfeld JS, et al. The Prospective Registry Evaluating Myocardial Infarction: Events and Recovery (PREMIER)--evaluating the impact of myocardial infarction on patient outcomes. Am Heart J. 2006;151(3):589-97. PMID: 16504619. DOI: 10.1016/j.ahj.2005.05.026.

7. ImproveCareNow: A Learning Health System for Children with Crohn's Disease and Ulcerative Colitis. https://www.pcori.org/research-results/2013/improvecarenow-learning-health-system-children-crohns-disease-and-ulcerative. Accessed June 10, 2019.

8. Rhodes P, Nocon A, Booth M, et al. A service users' research advisory group from the perspectives of both service users and researchers. Health Soc Care Community. 2002;10(5):402-9. PMID: 12390226.

9. Allen J, Mohatt GV, Rasmus SM, et al. The tools to understand: community as co-researcher on culture-specific protective factors for Alaska Natives. J Prev Interv Community. 2006;32(1-2):41-59. PMID: 17000601. DOI: 10.1300/J005v32n01_04.

10. Burrus BB, Liburd LC, Burroughs A. Maximizing participation by black Americans in population-based diabetes research: the Project DIRECT pilot experience. J Community Health. 1998;23(1):15-27. PMID: 9526723.

11. Minkler M, Fadem P, Perry M, et al. Ethical dilemmas in participatory action research: a case study from the disability community. Health Educ Behav. 2002;29(1):14-29. PMID: 11822550. DOI: 10.1177/109019810202900104.

12. Elliott E, Watson AJ, Harries U. Harnessing expertise: involving peer interviewers in qualitative research with hard-to-reach populations. Health Expect. 2002;5(2):172-8. PMID: 12031057. DOI: 10.1046/j.1369-6513.2002.00158.x.

Case Examples for Chapter 10

Case Example 21. Using Registry Tools To Recruit Sites

Description	The objective of the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure) registry was to improve quality of care and promote evidence-based therapies in heart failure. The registry provided a comprehensive process-of-care improvement program and gathered data that allowed hospitals to track their improvement over time.
	GlaxoSmithKline
	2003
	2005
	270
	More than 50,000

Challenge

The registry was designed to help hospitals improve care for patients hospitalized with heart failure. The objective was to accelerate the adoption of evidence-based guidelines and increase the use of the guideline-recommended therapies, thereby improving both short-term and long-term clinical outcomes for heart failure patients.

Proposed Solution

To increase compliance with guidelines, the registry team promoted the implementation of a process-of-care improvement component and the use of comprehensive patient education materials. They combined these materials into a hospital toolkit, which included evidence-based practice algorithms, critical pathways, standardized orders, discharge checklists, pocket cards, and chart stickers. The toolkit also included algorithms and dosing guides for the guideline-recommended therapies and a comprehensive set of patient education materials. The team engaged the steering committee in designing the toolkit to ensure that the materials reflected both the guideline-recommended interventions and the practical aspects of hospital processes.

In addition to the toolkit, the registry offered point-of-care tools, such as referral notes and patient letters, that could be customized for each patient based on data entered into the registry. The registry also included real-time performance reports that hospitals could use to assess their improvement on a set of standardized measures based on the guidelines.

Results

The hospital toolkit was a key component of the registry's marketing campaign. Hospitals could view the toolkit at recruitment meetings, but they did not receive their own copy until they joined the program. The toolkit gained credibility among hospitals because its creators included some of the most prominent members of the heart failure research and treatment community. Hospitals also actively used the reports to track their improvement over time and identify areas for

additional work. Overall, the registry recruited 270 hospitals and met its patient accrual goal six months ahead of schedule.

Key Point

Nonfinancial incentives, such as patient education materials, toolkits, and reports, can encourage sites to join a registry. Incentives that also add value for the site by improving their processes or providing materials that they use frequently can aid retention.

For More Information

- Fonarow GG, Abraham WT, Albert NM, et al. Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF): rationale and design. Am Heart J. 2004;148(1):43–51. PMID: 15215791. DOI: 10.1016/j.ahj.2004.03.004.
- Gheorghiade M, She L, Abraham WT, et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA. 2006;296:2217–26. PMID: 17090768. DOI: 10.1001/jama.296.18.2217
- Fonarow GC, Abraham WT, Albert NM, et al. Association between performance measures and clinical outcomes for patients hospitalized with heart failure. JAMA. 2007;297:61–70. PMID: 17200476. DOI: 10.1001/jama.297.1.61.

Case Example 22. Using a scientific advisory board to support investigator research projects

Description	The National LymphoCare Study is a large, prospective, disease-based registry designed to examine the disease presentation, treatment patters, and clinical outcomes of patients diagnosed with follicular lymphoma. There are a number of open clinical questions about follicular lymphoma treatment, including whether anthracyclines should be used early in the course of disease, and whether there is a group of patients for whom observation (as opposed to active treatment) is the best choice, given the indolent nature of the disease. The registry recruited consecutive patients diagnosed with follicular lymphoma between March 2004 and March 2007 at participating sites in the United States and follows patients for up to 10 years. Specific outcomes of interest include overall response rate, progression-free survival, time to subsequent therapy, and overall survival for common frontline and subsequent therapeutic strategies.
	Genentech, Inc., and Biogen Idec, Inc.
	2004
	2017
	250 community and academic sites
	2,740

Challenge

The National LymphoCare Study included a large number of community-based sites in addition to many academic sites. Many of the principal investigators at the community-based sites were interested in using the registry data to answer clinical questions, but they did not have sufficient research experience to design a research question, conduct data analysis, and share the results with the scientific community. One aim of the registry sponsors and scientific advisory board (SAB) was to facilitate research among the community investigators, both to increase interest in the registry and to increase the scope of research questions addressed using registry data.

Proposed Solution

The registry sponsors and the SAB developed a plan to allow investigators at enrolling sites to propose a question of interest; work with an SAB member, clinical scientists, epidemiologists, and biostatisticians to develop an analysis plan to answer the question; and present findings at scientific meetings. The plan was implemented in 2007, when the registry issued a call for research proposals to all participating investigators. The proposal outlined the types of data that were available at that point (e.g., descriptive data on demographics, initial treatments, etc.). Several community-based investigators sent in proposals, which the SAB then reviewed. The SAB selected the proposals that it felt were most appropriate for the available data and that answered the most valuable questions from a clinical standpoint.

The community investigator for each selected proposal was then paired with a member of the SAB to further develop the research question. This process included conference calls and emails to refine the question and the high-level analytic plan. Once the plan was ready, the investigator and the SAB member submitted the proposal and analytic plan to the registry sponsor. The sponsor provided support for analytic design and biostatistics. The investigator, in consultation with the SAB member, developed an abstract based on the results. Abstracts were reviewed by the full SAB before being submitted for presentation.

Results

The research program was been well received by community-based investigators, who have the opportunity to author their own research projects with mentoring from an experienced advisor. The SAB was also enthusiastic about working with community-based physicians on research methodology and adding to the scientific knowledge about this disease. The process resulted in several abstracts that were presented at scientific meetings.

Key Point

Community-based investigators who participate in a registry may be interested in pursuing research opportunities but may not have all of the necessary resources or expertise. By utilizing an engaged advisory board, a registry can provide investigators with research opportunities, resulting in more publications and presentations based on registry data, and potentially more engaged investigators.

For More Information

- Friedberg JW, Taylor MD, Cerhan JR, et al. Follicular lymphoma in the United States: first report of the National LymphoCare Study. J Clin Oncol. 2009;27:1202–8. PMID: 19204203. DOI: 10.1200/JCO.2008.18.1495.

- Friedberg JW, Wong EK, Taylor MD, et al. Characteristics of patients with stage I follicular lymphoma (FL) selected for watchful waiting (WW) in the US: report from the National LymphoCare Study (NLCS). American Society of Hematology. 2007. Abstract 3315.
- Link BK, Taylor MD, Brooks JM, et al. Correlates of treatment intensity for initial management of follicular lymphoma (FL) in the United States: report from the National LymphoCare Study (NLCS). American Society of Hematology. 2007. Abstract 2612.
- Matasar MJ, Saxena R, Wong EK, et al. Practice patterns in the diagnosis of follicular lymphoma (FL): report from the National LymphoCare Study (NLCS). American Society of Hematology. 2007. Abstract 2613.
- Nabhan C, Morawa E, Bitran JD, et al. Patterns of care in follicular lymphoma (FL): are minorities being treated differently? Report from the National LymphoCare Study (NLCS). American Society of Hematology. 2007. Abstract 367.

Chapter 11. Obtaining Data and Quality Assurance

1. Introduction

This chapter focuses on the procedures for obtaining registry data and associated quality assurance principles. Data management—the integrated system for obtaining, cleaning, storing, monitoring, reviewing, and reporting on registry data—determines the utility of the data for meeting the goals of the registry. Quality assurance, on the other hand, aims to assure that the data were, in fact, collected in accordance with these procedures and that the data stored in the registry database meet the requisite standards of quality, which are generally defined based on the intended purposes. In this chapter, the term registry coordinating activities refers to the centralized procedures performed for a registry, and the term registry coordinating center refers to the entity or entities performing these procedures and overseeing the registry activities at the site and patient levels.

Because the range of registry purposes can be broad, a similar range of data collection procedures may be acceptable, but only certain methodologies may be suitable for particular purposes. Furthermore, certain end users of the data may require that data collection or validation be performed in accordance with their own guidelines or standards. For example, a registry that collects data electronically and intends for those data to be used by the U.S. Food and Drug Administration (FDA) should meet the systems validation requirements of that end user of the data, such as Title 21 of the Code of Federal Regulations Part 11 (21 CFR Part 11). Such requirements may have a substantial effect on the registry procedures. Similarly, registries may be subject to specific processes depending on the type of data collected, the types of authorization obtained, and the applicable governmental regulations.

Requirements for data collection and for quality assurance should be defined during the registry inception and creation phases. Certain requirements may have significant cost implications, and these should be assessed on a cost-to-benefit basis in the context of the intended purposes of the registry. This chapter describes a wide range of centralized and distributed data collection and quality assurance activities currently in use or expected to become more commonly used in patient registries.

2. Obtaining Data

2.1 Database Requirements and Case Report Forms

Chapter 1 defined key characteristics of patient registries for evaluating patient outcomes. They include specific and consistent data definitions for collecting data elements in a uniform manner for every patient. As in randomized controlled trials, the case report form (CRF) is the paradigm for the data structure of the registry. A CRF is a formatted listing of data elements that can be presented in paper or electronic formats. Those data elements and data entry options in a CRF are represented in the database schema of the registry by patient-level variables. Defining the registry CRFs and corresponding database schema are the first steps in data collection for a registry. Chapter 5 escribes the selection of data elements for a registry. All data elements should be modeled within the CRF, even those that might be obtained from secondary sources (Chapter

6). For data that are obtained from secondary sources, the CRF should be structured so that it does not change the meaning of a secondary data element (for instance, populating a series of co-morbidities on a CRF with diagnoses obtained from the electronic health record (EHR).

Two related documents should also be considered part of the database specification: the data dictionary (including data definitions and parameters) and the data validation rules, also known as queries or edit checks. The data dictionary and definitions describe both the data elements and how those data elements are interpreted. The data dictionary contains a detailed description of each variable used by the registry, including the source of the variable, coding information if used, and normal ranges if relevant. For example, the term "current smoker" should be defined as to whether "smoker" refers to tobacco or other substances and whether "current" refers to active or within a recent time period (e.g., within the last year). Data validation rules refer to the logical checks on data entered into the database against predefined rules for either value ranges (e.g., systolic blood pressure less than 300 mmHg) or logical consistency with respect to other data fields for the same patient; these are described more fully below. While neither registry database structures nor database requirements are standardized, the Clinical Data Interchange Standards Consortium (CDISC)[1] is actively working on representative models of data interchange and portability using standardized concepts and formats. Chapter 5 further discusses these models, which are applicable to registries as well as clinical trials.

2.2 Procedures and Personnel

Data collection procedures need to be carefully considered in planning the operations of a registry. Successful registries depend on a sustainable workflow model that can be integrated into the day-to-day clinical practice of active physicians, nurses, pharmacists, and patients, with minimal disruption. Registry developers can benefit tremendously from preliminary input from the healthcare workers, study coordinators, or patients who are likely to be participants.

2.2.1 Pilot Testing

One method of gathering input from likely participants before the full launch of a registry is pilot testing. Whereas feasibility testing, which is discussed in Chapter 2, focuses on whether a registry should be implemented, pilot testing focuses on how it should be implemented. Piloting can range from testing a subset of the procedures, CRFs, or data capture systems, to a full launch of the registry at a limited subset of sites with a limited number of patients.

The key to effective pilot testing is to conduct it at a point where the results of the pilot can still be used to modify the registry implementation. Through pilot testing, one can assess comprehension, acceptance, feasibility, and other factors that influence how readily the patient registry processes will fit into patient lifestyles and the normal practices of the healthcare provider. Chapter 5 discusses pilot testing in more detail.

2.2.2 Documentation of Procedures

The data collection procedures for each registry should be clearly defined and described in a detailed manual. The term manual here refers to the reference information in any appropriate form, including hard copy, electronic, or via interactive Web or software-based systems.

Although the detail of this manual may vary from registry to registry depending on the intended purpose, the required information generally includes protocols, policies, and procedures; the data collection instrument(s); and a listing of all the data elements and their full definitions. If the registry has optional fields (i.e., fields that do not have to be completed on every patient), these should be clearly specified.

In addition to patient inclusion and exclusion criteria, the screening process should be specified, as should any documentation to be retained at the site level and any plans for monitoring or auditing of screening practices. If sampling is to be performed, the method or systems used should be explained, and tools should be provided to simplify this process for the sites. The manual should clearly explain how patient identification numbers are created or assigned and how duplicate records should be prevented. Any required training for data collectors should also be described.

If paper CRFs are used, the manual should describe specifically how they are used and which parts of the forms (e.g., two-part or three-part no-carbon-required forms) should be retained, copied, submitted, or archived. If electronic CRFs are used, clear user manuals and instructions should be available. These procedures are an important resource for all personnel involved in the registry as well as for external auditors who might be asked to assure the quality of the registry.

The importance of standardizing procedures to ensure that the registry uses uniform and systematic methods for collecting data cannot be overstated. At the same time, some level of customization of data entry methods may be required or permitted to enable the participation of particular sites or subgroups of patients within some practices. As discussed in Chapter 10, if the registry provides payments to sites for participation, then the specific requirements for site payments should be clearly documented, and this information should be provided with the registry documents.

2.2.3 Personnel

All personnel involved in data collection should be identified, and their job descriptions and respective roles in data collection and processing should be described. Examples of such “roles” include patient, physician, data entry personnel, site coordinator, help desk, data manager, data analyst, quality analyst, terminologist, and monitor. The necessary documentation or qualification required for any role should be specified in the registry documentation. As an example, some registries require personnel documentation such as a curriculum vitae, protocol signoff, attestation of intent to follow registry procedures, or confirmation of completion of specified training.

2.3 Data Sources

The sources of data for a registry may include new information collected from the patient, new or existing information reported by or derived from the clinician and the medical record, and ancillary stores of patient information, such as laboratory results. Since registries for evaluating patient outcomes should employ uniform and systematic methods of data collection, all data-related procedures—including the permitted sources of data; the data elements and their definitions; and the validity, reliability, or other quality requirements for the data collected from

each source—should be predetermined and defined for all collectors of data. As described in Section 3 below, data quality is dependent on the entire chain of data collection and processing. Therefore, the validity and quality of the registry data as a whole ultimately derive from the least, not the most, rigorous link.

In Chapter 6, data sources are classified as primary or secondary, based on the relationship of the data to the registry purpose and protocol. Primary data sources incorporate data collected for direct purposes of the registry (i.e., primarily for the registry). Secondary data sources consist of data originally collected for purposes other than the registry (e.g., standard medical care, insurance claims processing). A registry may contain one or both kinds of data sources. The sections below incorporate and expand on these definitions.

2.3.1 Primary Data Collection

The major sources of primary data in a registry are patients and clinicians. Patient-reported data are data specifically collected from the patient for the purposes of the registry rather than interpreted through a clinician or an indirect data source (e.g., laboratory value, pharmacy records). Such data may range from basic demographic information to validated scales of patient-reported outcomes (PROs). From an operational perspective, a wide range of issues should be considered in obtaining data directly from patients. These range from presentation (e.g., font size, language, reading level) to technologies (e.g., paper-and-pencil questionnaires, computer inputs, telephone or voice inputs, or hand-held patient diaries). Mistakes at this level can inadvertently bias patient selection, invalidate certain outcomes, or significantly affect cost. Limiting the access for patient reporting to particular languages or technologies may limit participation. Patients with specific diagnoses may have difficulties with specific technologies (e.g., small font size for visually impaired, paper and pencil for those with rheumatoid arthritis). Other choices, such as providing a PRO instrument in a format or method of delivery that differs from how it was validated (e.g., questionnaire rather than interview), may invalidate the results. For more information on patient-reported outcome development and use, see Chapters 4 and 5.

Clinician-reported or -derived data can also be divided into primary and secondary subcategories. As an example, specific clinician rating scales (e.g., the National Institutes of Health Stroke Scale)[2] may be required for the registry but not routinely captured in clinical encounters. Some variables might be collected directly by the clinician for the registry. Data elements that the clinician must collect directly (e.g., because of a particular definition or need to assess a specific comorbidity that may or may not be routinely present in the medical record) should be specified. These designations are important because they determine who can collect the data for a particular registry or what changes must be made in the procedures the clinician follows in recording a medical record for a patient in a registry. Furthermore, the types of error that arise in registries (discussed in Section 3) will differ by the degree of use of primary and secondary sources, as well as other factors. As an example, registries that use medical chart abstracters, as discussed below, may be subject to more interpretive errors.[3]

2.3.2 Secondary Sources

Data from secondary sources can be obtained in several different ways. These include manual abstraction, direct import, transformation, and computational derivation. Each of these methods

are described in the sections below, along with potential caveats related to data quality. Note that these quality issues are different than those that would be uncovered during traditional data cleaning procedures (Section 2.4). Those measures are concerned with data once they have been entered into the CRF. The issues described here occur upstream from that process, as data are transferred out of the secondary source.

2.3.2.1 Manual Abstraction

Manual abstraction is the process by which a data collector other than the clinician interacting with the patient extracts clinician-reported data. While physical examination findings, such as height and weight, or laboratory findings, such as white blood cell counts, are straightforward, abstraction usually involves varying degrees of judgment and interpretation.

Clarity of description and standardization of definitions are essential to the assurance of data quality and to the prevention of interpretive errors when using manual abstraction. Knowledgeable registry personnel should be designated as resources for the data collectors in the field, and processes should be put in place to allow the data collectors in the field continuous access to these designated registry personnel for questions on specific definitions and clinical situations. Registries that span long periods, such as those intended for surveillance, might be well served by a structure that permits the review of definitions on a periodic basis to ensure the timeliness and completeness of data elements and definitions, and to add new data elements and definitions. A new product or procedure introduced after the start of a registry is a common reason for such an update.

Abstracting data is often an arduous and tedious process, especially if free text is involved, and it usually requires a human reader. The reader, whose qualifications may range from a trained "medical record analyst" or other health professional to an untrained research assistant, may need to decipher illegible handwriting (paper or scanned documents), translate obscure abbreviations and acronyms, and understand the clinical content to sufficiently extract the desired information. Registry personnel should develop formal chart abstraction guidelines, documentation of processes and practical definitions of terms, and coding forms for the analysts and reviewers to use. Generally, the guidelines include instructions to search for specific types of data that will go into the registry (e.g., specific diagnoses or laboratory results). Often the analyst will be asked to code the data, using either standardized codes from a codebook (e.g., the ICD-10 [International Classification of Diseases, 10th Revision] code) corresponding to a text diagnosis in a chart, or codes that may be unique to the registry (e.g., a severity scale of 1 to 5).

All abstraction and coding instructions must be carefully documented and incorporated into a data dictionary for the registry. Because of lack of precision in natural language, the clinical data abstracted by different abstracters from the same documents may differ. This is a potential source of error in a registry. To reduce the potential for this source of error, registries should ensure proper training on the registry protocol and procedures, condition(s), data sources, data collection systems, and most importantly, data definitions and their interpretation. While training should be provided for all registry personnel, it is particularly important for non-clinician data abstracters. Training time depends on the nature of the source (charts or CRFs), complexity of the data, and number of data items. A variety of training methods, from live meetings to online meetings to interactive multimedia recordings, have all been used with success.[4] Training often

includes test abstractions using sample charts. For some purposes, it is best practice to train abstracters using standardized test charts. Such standardized tests can be further used both to obtain data on the inter-rater reliability of the CRFs, definitions, and coding instructions and to determine whether individual abstracters can perform up to a defined minimum standard for the registry. Registries that rely on medical chart abstraction should consider reporting on the performance characteristics associated with abstraction, such as inter-rater reliability.[5] Examining and reporting on intra-rater reliability may also be useful. Some key considerations in standardizing medical chart abstractions are:

- Standardized materials (e.g., definitions, instructions)
- Standardized training
- Testing with standardized charts
- Reporting of inter-rater reliability

2.3.2.2 Direct Import

When data in the secondary source are in electronic format, the simplest method of transmitting secondary data to a registry is through direct import. In this case, there is a 1:1 correlation between the fields in the secondary source and the registry CRF. The questions/fields in the secondary source have exactly the same meaning as the CRF, with the same data types and field formats, so no translation is necessary. It is rare for this scenario to occur for more than a handful of variables. Most fields will have differences in their value sets or formatting that require some sort of transformation or mapping, as described in the next section.

2.3.2.3 Transformation

The most common way that secondary data are imported into a registry is through a transformation process. Data are extracted from the secondary source, transformed to look like the target field, then loaded into the registry. This process is often call extract-transform-load, or ETL. For the purposes of this discussion, a distinction is made between transformation (described here) and derivation (described below). It is a somewhat artificial separation, but important in highlighting some of the issues that must be considered. Transformation refers to the process of translating data into a consistent format to support integration and analysis. In the context of transferring secondary data to a registry, transformation typically involves converting data elements from the secondary source to match the format required by the registry.

With numeric data, or data that are meant to represent numeric values (e.g. height and weight measurements), transformation rules should be established to specify whether to include or remove leading zeroes, decimals, dashes, etc. Even though these data are available in many source systems, they may not be represented in the same manner, so registry personnel should decide if they want to receive the data as they are and then transform them or have personnel at each of the participating sites complete the transformation locally. The former approach can reduce the chance for error, since only one group is developing the transformation rules.

For categorical variables, a simple type of transformation occurs when the same variable/concept is collected in different sources, but the value sets differ. This scenario can often be resolved by

creating a mapping between the different value sets, but those mappings should be validated to ensure the meaning of the data are not changed. More complicated situations occur when the secondary source may have two registry concepts captured in a single field or vice versa. This frequently occurred with data on Race and Ethnicity that were captured in EHRs prior to Meaningful Use. Hispanic, which is an Ethnicity value, was often listed as a category under Race, whereas many registries captured Race and Ethnicity as two separate fields. When using the EHR data to populate demographics, the registry personnel had to decide how they to handle transform these data (e.g., if "Hispanic" is entered as a value for Race, populate the Ethnicity field with Hispanic and leave the value for Race as "Unknown"), and what steps they want to take, if any, in dealing with the resulting missing data.

Standardization of concepts is also an important type of transformation. For many medical terms, different data sources may have different local terms to represent the same concept (e.g., acetaminophen vs. Tylenol). In these cases, mapping the local terms to a standard vocabulary (e.g., LOINC, RxNorm) is critical to ensure that data from different data sources are integrated and interpreted appropriately.

If obtaining data from a source that has been mapped to a CDM, it is important to understand the mapping between the original source and the CDM and then from the CDM to the registry. In some cases, a CDM may have a limited value set for a given field, whereas there is a great deal of granularity in the source (e.g., EHR encounter type). In these cases, the CDM value set may not be sufficient for the needs of the registry, and it may be necessary to use the raw source values instead. Many CDMs allow these raw source values to be stored as part of each record, but sites may not always populate them.

Documentation of transformations is an obvious but critical part of maintaining traceability to the original source. A formal documentation process for transformations is not only good practice but may be required for registries used for regulatory purposes.

2.3.2.4 Computational Derivation

Due to the ubiquity of secondary data sources, particularly the EHR, it can be tempting to try to use as much of the data as possible to reduce the primary data collection burden. Given that the data elements in the secondary source may not have exactly the same meaning, some type of computational derivation is required to make the secondary data "fit" the primary registry element. Depending on the quality of these derivations, additional data collection or validation may still be necessary.

Examples of some common types of derivations are the use of EHR data to assign conditions or co-morbidities, or "computable phenotypes." These can be relatively straightforward, such as declaring that someone has a condition based on the presence of specified diagnosis codes, or they can involve machine learning or other techniques that consider a large number of variables. These phenotype algorithms are often designed for specific purposes (e.g., recruitment for clinical trial, identify those with known disease, identify those at risk), so registry personnel ensure that they are using the algorithm that is best fit for their purpose. The Electronic MEdical Records and Genomics (eMERGE) Network has developed and published a number of

computable phenotype algorithms over the years.[6] Many of the network's algorithms, and the algorithms of other researchers, can be found in the PheKB repository.[7]

An additional example is the use of medication orders to determine exposure or history (ever/never exposure). In this scenario, the completeness of the secondary source will determine how much information can be derived. For instance, with EHR orders, the presence of an order may be indicative of exposure (the "ever" case), but the absence of an order does not necessarily mean that the patient was never exposed. If a registry were collecting medication history, additional followup would likely be necessary.

Another form of computational derivation occurs when trying to extract meaning from narrative text. While a great deal of information is captured in the EHR in structured fields, a substantial portion of information is recorded as text. Physician progress notes, consultations, and radiology reports, are all examples of narrative text that may be typed directly by the clinician or dictated and transcribed (many EHRs also include the ability to generate text based on responses to structured fields, which would not fall into this category, since those responses can be extracted as structured data). While manual abstraction of free text occurs frequently, computational methods can be used to extract information from free text that is stored electronically. This is referred to as *natural language processing* (NLP), which is another form of computational derivation. The goal of NLP is to parse free text into meaningful components based on a set of rules or mathematical probabilities that enable the program to recognize key words, understand grammatical constructions, and resolve word ambiguities. Information can be extracted and delivered to the registry along with structured data, and both can be stored as structured data in the registry database. In registries where some sites are using NLP to populate a field while other sites use different methods (e.g., abstraction, transformation or direct import), it is worth noting the source of the data, as each introduces different types of potential error.

An increasing number of NLP software packages are available (e.g., cTAKES[8],CLAMP,[9] MetaMap,[10] MedLEE,[11] and a number of commercial products). NLP software operates best when it is trained in specific clinical domains with structured documents (e.g., radiology, pathology) that are coupled with large training datasets. Despite significant investment and progress in recent years, it is still relatively difficult to deploy NLP at scale for all-purpose chart abstraction. Projects that have found success tend to operate centralized models with a single processing pipeline, which has a lower cost than trying to do NLP at every site. A centralized approach typically requires the transfer of protected health information (PHI), however, which can present legal and regulatory hurdles.

Computational derivations can add tremendously to the ability for registries to obtain efficient and accurate data from secondary sources. However, in order to use derived data from these technologies such as machine learning or NLP, registry evaluators will need to understand the accuracy and methods of validation for the derivations used to the greatest extent possible. Increasingly, standardized performance metrics will be reported with models used in derivations such as Area Under the Curve (AUC), positive predictive value, precision, recall and so forth.

2.4 Data Entry Systems

Once the primary and any secondary data sources for a registry have been identified, the registry team can determine how data will be entered into the registry database. Many techniques and technologies exist for entering or moving data into the registry database, including paper CRFs, direct data entry, facsimile or scanning systems, interactive voice response systems, and electronic CRFs. There are also different models for how quickly those data reach a central repository for cleaning, reviewing, monitoring, or reporting. Each approach has advantages and limitations, and each registry must balance flexibility (the number of options available) with data availability (when the central repository is populated), data validity (whether all methods are equally able to produce clean data), and cost. Appropriate decisions depend on many factors, including the number of data elements, number of sites, location (local preferences that vary by country, language differences, and availability of different technologies), registry duration, followup frequency, and available resources.

2.4.1 Paper CRFs

With paper CRFs, the clinician enters clinical data on the paper form at the time of the clinical encounter, or other data collectors abstract the data from medical records after the clinical encounter. CRFs may include a wide variety of clinical data on each patient gathered from different sources (e.g., medical chart, laboratory, pharmacy) and from multiple patient encounters. Before the data on formatted paper forms are entered into a computer, the forms should be reviewed for completeness, accuracy, and validity. Paper CRFs can be entered into the database by either direct data entry or computerized data entry via scanning systems.

With direct data entry, a computer keyboard is used to enter data into a database. Key entry has a variable error rate depending on personnel, so an assessment of error rate is usually desirable, particularly when a high volume of data entry is performed. Double data entry is a method of increasing the accuracy of manually entered data by quantifying error rates as discrepancies between two different data entry personnel; data accuracy is improved by having up to two individuals enter the data and a third person review and manage discrepancies. With upfront data validation checks on direct data entry, the likelihood of data entry errors significantly decreases. Therefore, the choice of single versus double data entry should be driven by the requirements of the registry for a particular maximal error rate and the ability of each method to achieve that rate in key measures in the particular circumstance. Double data entry, while a standard of practice for registrational trials, may add significant cost. Its use should be guided by the need to reduce an error rate in key measures and the likelihood of accomplishing that by double data entry as opposed to other approaches. In some situations, assessing the data entry error rates by re-entering a sample of the data is sufficient for reporting purposes.

With hard-copy structured forms, entering data using a scanner and special software to extract the data from the scanned image is possible. If data are recorded on a form as marks in checkboxes, the scanning software enables the user to map the location of each checkbox to the value of a variable represented by the text item associated with the checkbox, and to determine whether the box is marked. The presence of a mark in a box is converted by the software to its corresponding value, which can then be transmitted to a database for storage. If the form contains hand-printed or typed text or numbers, optical character recognition software is often

effective in extracting the printed data from the scanned image. However, the print font must be of high quality to avoid translation errors, and spurious marks on the page can cause errors. Error checking is based on automated parameters specified by the operator of the system for exception handling. The comments on assessing error rates in the section above are applicable for scanning systems as well.

2.4.2 Electronic CRFs

An electronic CRF (eCRF) is defined as an auditable electronic form designed to record information required by the clinical trial protocol to be reported to the sponsor on each trial subject.[12] An eCRF allows clinician-reported data to be entered directly into the electronic system by the data collector (the clinician or other data collector). Site personnel in many registries still commonly complete an intermediate hard-copy worksheet representing the CRF and subsequently enter the data into the eCRF. While this approach increases work effort and error rates, it is still in use because it is not yet practical for all electronic data entry to be performed at the bedside, during the clinical encounter, or in the midst of a busy clinical day.

An eCRF may originate on local systems (including those on an individual computer, a local area network server, or a hand-held device) or directly from a central database server via an Internet-based connection or a private network. For registries that exist beyond a single site, the data from the local system must subsequently communicate with a central data system. An eCRF may be presented visually (e.g., computer screen) or aurally (e.g., telephonic data entry, such as interactive voice response systems). Specific circumstances will favor different presentations. For example, in one clozapine patient registry, both pharmacists and physicians can obtain and enter data via a telephone-based interactive voice response system as well as a Web-based system. The option is successful in this scenario because telephone access is ubiquitous in pharmacies and the eCRF is very brief.

A common method of electronic data entry is to use Web-based data entry forms. Such forms may be used by patients, providers, and interviewers to enter data into a local repository. The forms reside on servers, which may be located at the site of the registry or co-located anywhere on the Internet. To access a data entry form, a user on a remote computer with an Internet connection opens a browser window and enters the address of the Web server. Typically, a login screen is displayed and the user enters a user identification and password, provided by personnel responsible for the website or repository. Once the server authenticates the user, the data entry form is displayed, and the user can begin entering data. As described in “Cleaning Data,” many electronic systems can perform data validation checks or edits at the time of data entry. When data entry is complete, the user submits the form, which is sent over the Internet to the Web server.

Smart phones or other mobile devices may also be used to submit data to a server to the extent such transmissions can be done with appropriate information security controls. Mobility has recently become an important attribute for clinical data collection. Software has been developed that enables wireless devices to collect data and transmit them over the Internet to database servers in fixed locations. As wireless technology continues to evolve and data transmission rates increase, these will become more essential data entry devices for patients and clinicians.

2.4.3 Electronic Upload

Aside from manual entry, the most common way for data to be transferred to a registry is through electronic transfer, or electronic upload. While this can occur with some primary data, it typically occurs with secondary data sources. The ease of extracting data from electronic systems for use in a registry depends on the design of the registry systems, and the ability of the source to make the requested data accessible. Registry systems may support a variety of input formats (e.g., flat files, web services, etc.), and organizations, including HL7,[13] the Office of the National Coordinator for Health Information Technology (ONC), the National Institute of Standards and Technology,[14] CDISC, and others, have worked to define a number of common formats.

When electronically transferring data from one system to the registry, additional steps for exception handling are necessary. This can include situations where a record in a secondary source is updated or deemed invalid after the data have already been transferred to the registry. The registry software must be able to receive that notification, flag the erroneous value as invalid, and insert the new, corrected value into its database. Additional logic may be necessary if a registry is receiving the same information from multiple sources that could potentially be in conflict (e.g., date of birth recorded in multiple systems). Finally, it is important to recognize that the use of an electronic-to-electronic interchange requires not only testing but also validation of the integrity and quality of the data transferred. For registries that intend to report data to FDA or to other sponsors or data recipients with similar requirements, including electronic signatures, audit trails, and rigorous system validation, the ways in which the registry interacts with these other systems must be carefully considered.

2.5 Cleaning Data

Data cleaning refers to the correction or amelioration of data problems, including missing values, incorrect or out-of-range values, responses that are logically inconsistent with other responses in the database, and duplicate patient records. While all registries strive for “clean data,” in reality, this is a relative term. How and to what level the data will be cleaned should be addressed upfront in a data management manual that identifies the data elements that are intended to be cleaned, describes the data validation rules or logical checks for out-of-range values, explains how missing values and values that are logically inconsistent will be handled, and discusses how duplicate patient records will be identified and managed.

2.5.1 Data Management Manual

Data managers should develop formal data review guidelines for the reviewers and data entry personnel to use. The guidelines should include information on how to handle missing data; invalid entries (e.g., multiple selections in a single-choice field, alphabetic data in a numeric field); erroneous entries (e.g., patients of the wrong gender answering gender-based questions); and inconsistent data (e.g., an answer to one question contradicting the answer to another one). The guidelines should also include procedures to attempt to remediate these data problems. For example, with a data error on an interview form, it may be necessary to query the interviewer or the patient, or to refer to other data sources that may be able to resolve the problem. Documentation of any data review activity and remediation efforts, including dates, times, and results of the query, should be maintained.

For secondary data sources, the data analyst group should define formal data transformation, cleaning, and monitoring rules and procedures. For example, when multiple updates transactions are uploaded from the data source, the rules and procedures should specify whether all the updates or only the last version of the record should be kept in the registry and whether every field in the record should be updated with the most recent value or only a subset of fields.

2.5.2 Automated Data Cleaning

Ideally, automated data checks are preprogrammed into the database for presentation at the time of data entry or data upload. These data checks are particularly useful for cleaning data at the site level while the patient or medical record is readily accessible. Even relatively simple edit checks, such as range values for laboratories, can have a significant effect on improving the quality of data. Many systems allow for the implementation of more complex data edit checks, and these checks can substantially reduce the amount of subsequent manual data cleaning. A variation of this method is to use data cleaning rules to deactivate certain data fields so that erroneous entries cannot even be made. A combination of these approaches can also be used. It should be noted that specifying that a data check is required (i.e., must be resolved before the user can proceed) can sometimes lower data completeness rates, as users abandon data collection entirely if they cannot resolve the failed required check. For paper-based entry methods, automated data checks are not available at the time the paper CRF is being completed but can be incorporated when the data are later entered into the database.

2.5.3 Manual Data Cleaning

Data managers perform manual data checks or queries to review data for unexpected discrepancies. This is the standard approach to cleaning data that are not entered into the database at the site (e.g., for paper CRFs entered via data entry or scanning). By carefully reviewing the data using both data extracts analyzed by algorithms and hand review, data managers identify discrepancies and generate “queries” to send to the sites to resolve. Even eCRF-based data entry with data validation rules may not be fully adequate to ensure data cleaning for certain purposes. Anticipating all potential data discrepancies at the time that the data management manual and edit checks are developed is very difficult. Therefore, even with the use of automated data validation parameters, some manual cleaning is often still performed. For fields populated with data from secondary data sources, remediation may not be possible (e.g., a field was not populated in the EHR during the patient’s visit) without resorting to primary data collection or including data from yet another secondary source. If multiple secondary sources are considered, data checks will be needed to handle situations where the information may be in conflict (e.g., discrepant values for date of death).

2.5.4 Query Reports

The registry coordinating center should generate, on a periodic basis, query reports that relate to the quality of the data received, based on the data management manual and, for some purposes, additional concurrent review by a data manager. The content of these reports will differ depending on what type of data cleaning is required for the registry purpose and how much automated data cleaning has already been performed. Query reports may include missing data,

"out-of-range" data, or data that appear to be inconsistent (e.g., positive pregnancy test for a male patient). They may also identify abnormal trends in data, such as sudden increases or decreases in laboratory tests compared with patient historical averages or clinically established normal ranges. Qualified registry personnel should be responsible for reviewing the abnormal trends with designated site personnel. The most effective approach is for sites to provide one contact representative for purposes of queries or concerns by registry personnel. Depending on the availability of the records and resources at the site to review and respond to queries, resolving all queries can sometimes be a challenge. Creating systematic approaches to maximizing site responsiveness is recommended.

2.5.5 Data Tracking

For most registry purposes, tracking of data received (paper CRFs), data entered, data cleaned, and other parameters is an important component of active registry management. By comparing indicators, such as expected to observed rates of patient enrollment, CRF completion, and query rates, the registry coordinating center can identify problems and potentially take corrective action— either at individual sites or across the registry as a whole. This is discussed further in the Quality Assurance section below.

2.5.6 Coding Data

As further described in earlier in this chapter and in Chapter 5, the use of standardized coding dictionaries is an increasingly important tool in the ability to aggregate registry data with other databases and reduce variation in information semantics. As the health information community adopts standards, registries should routinely apply them unless there are specific reasons not to use such standard codes. While such codes should be implemented in the data dictionaries during registry planning, including all codes in the interface is not always possible. Some free text may be entered as a result. When free text data are entered into a registry, recoding these data using standardized dictionaries (e.g., MedDRA, WHODRUG, SNOMED®) may be worthwhile. There is cost associated with recoding, and in general, it should be limited to data elements that will be used in analysis or that need to be combined or reconciled with other datasets, such as when a common safety database is maintained across multiple registries and studies.

2.5.7 Storing and Securing Data

When data on a form are entered into a computer for inclusion in a registry, the form itself, as well as a log of the data entered, should be maintained for the regulatory archival period. Data errors may be discovered long after the data have been stored in the registry. The error may have been made by the patient or interviewer on the original form or during the data entry process. Examination of the original form and the data entry log should reveal the source of the error. If the error is on the form, correcting it may require re-interviewing the patient. If the error occurred during data entry, the corrected data should be entered and the registry updated. By then, the erroneous registry data may have been used to generate reports or create cohorts for population studies. Therefore, instead of simply replacing erroneous data with corrected data, the registry system should have the ability to flag data as erroneous without deleting them and to insert the corrected data for subsequent use.

Once data are entered into the registry, the registry database should be backed up on a regular basis. There are two basic types of backup, and both types should be considered for use as best practice by the registry coordinating center. The first type is real-time disk backup, which is done by the storage hardware used by the registry server. The second is a regular (e.g., daily) backup of the registry to removable media. In the first case, as data are stored on disk in the registry server, they are automatically replicated to two or more physical hard drives. In the simplest example, called "mirroring," registry data are stored on a primary disk and an exact replica is stored on the mirrored disk. If either disk fails, data continue to be stored on the mirrored disk until the failed disk is replaced. This failure can be completely transparent to the user, who may continue entering and retrieving data from the registry database during the failure. More complex disk backup configurations exist, in which arrays of disks are used to provide protection from single disk failures.

The second type of periodic backup is needed for disaster recovery. Ideally, a daily backup copy of the registry database stored on removable media should be maintained off site. In case of failure of the registry server or disaster that closes the data center, the backup copy can be brought to a functioning server and the registry database restored, with the only potential loss of data being for the interval between the regularly scheduled backups. The lost data can usually be reloaded from local data repositories or re-entered from hard copy. Other advanced and widely available database solutions and disaster recovery techniques may support a "standby" database that can be located at a remote data center. In case of a failure at the primary data center, the standby database can be used, minimizing downtime and preventing data loss.

2.6 Managing Change

As with all other registry processes, the extent of change management will depend on the types of data being collected, the source(s) of the data, and the overall timeframe of the registry. There are two major drivers behind the need for change during the conduct of a registry: internally driven change to refine or improve the registry or the quality of data collected, and externally driven change that comes as a result of changes in the environment in which the registry is being conducted.

Internally driven change is generally focused on changes to data elements or data validation parameters that arise from site feedback, queries, and query trends that may point to a question, definition, or CRF field that was poorly designed or missing. If this is the case, the registry can use the information coming back from sites or data managers to add, delete, or modify the database requirements, CRFs, definitions, or data management manual as required. At times, more substantive changes, such as the addition of new forms or changes to the registry workflow, may be desirable to examine new conditions or outcomes. Externally driven change generally arises in multiyear registries as new information about the disease and/or product under study becomes available, or as new therapies or products are introduced into clinical practice. Secondary data sources may also change, resulting in registry changes. Change and turnover in registry personnel is another type of change, and one that can be highly disruptive if procedures are not standardized and documented.

A more extensive form of change may occur when a registry either significantly changes its CRFs or changes the underlying database. Longstanding registries address this issue from time to

time as information regarding the condition or procedure evolves and data collection forms and definitions require updating.

Proper management of change is crucial to the maintenance of the registry. A consistent approach to change management, including decision making, documentation, data mapping, and validation, is an important aspect of maintaining the quality of the registry and the validity of the data (see Chapter 2). While the specific change management processes might depend on the type and nature of the registry, change management in registries that are designed to evaluate patient outcomes requires, at the very least, the following structures and processes:

- *Detailed manual of procedures*: As described earlier, a detailed manual that is updated on a regular basis—containing all the registry policies, procedures, and protocols, as well as a complete data dictionary listing all the data elements and their definitions—is vital for the functioning of a registry. The manual is also a crucial component for managing and documenting change management in a registry.
- *Governing body*: As described in Chapters 2 and 8, registries require oversight and advisory bodies for a number of purposes. One of the most important is to manage change on a regular basis. Keeping the registry manual and data definitions up to date is one of the primary responsibilities of this governing body. Large prospective registries, such as the National Surgical Quality Improvement Program, have found it necessary to delegate the updating of data elements and definitions to a special definitions committee.
- *Infrastructure for ongoing training*: As mentioned above, change in personnel is a common issue for registries. Specific processes and an infrastructure for training should be available at all times to account for any unanticipated changes and turnover of registry personnel or providers who regularly enter data into the registry.
- *Method to communicate change*: Since registries frequently undergo change, there should be a standard approach and timeline for communicating to sites when changes will take place.

In addition to instituting these structures, registries should also plan for change from a budget perspective (Chapter 2) and from an analysis perspective (Chapter 13).

3. Quality Assurance

In determining the utility of a registry for decision making, it is critical to understand the quality of the procedures used to obtain the data and the quality of the data stored in the database. As patient registries that meet sufficient quality criteria (discussed in Chapters 1 and 14) are increasingly being seen as important means to generate evidence regarding effectiveness, safety, and quality of care, the quality of data within the registry must be understood in order to evaluate its suitability for use in decision making. Registry planners should consider how to ensure quality to a level sufficient for the intended purposes (as described below) and should also consider how to develop appropriate quality assurance plans for their registries. Those conducting the registry should assess and report on those quality assurance activities.

Methods of quality assurance will vary depending on the intended purpose of the registry. A registry intended to serve as key evidence for decision making[15] (e.g., coverage determinations,

product safety evaluations or other regulatory decision making, or performance-based payment) will require higher levels of quality assurance than a registry describing the natural history of a disease. Quality assurance activities generally fall under three main categories: (1) quality assurance of data, (2) quality assurance of registry procedures, and (3) quality assurance of computerized systems. Since many registries are large, the level of quality assurance that can be obtained may be limited by budgetary constraints.

To balance the need for sufficient quality assurance with reasonable resource expenditure for a particular purpose, a risk-based approach to quality assurance is highly recommended. A risk-based approach focuses on the most important sources of error or procedural lapses from the perspective of the registry's purpose. Such sources of error should be defined during inception and design phases. As described below, registries with different purposes may be at risk for different sources of error and focus on different practices and levels of assessment. Standardization of methods for particular purposes (e.g., national performance measurement) will likely become more common in the future if results are to be combined or compared between registries.

3.1 Assurance of Data Quality

Structures, processes, policies, and procedures need to be put in place to ascertain the quality of the data in the registry and to insure against several types of errors, including:

- *Errors in interpretation or coding*: An example of this type of error would be two abstracters looking for the same data element in a patient's medical record but extracting different data from the same chart. Variations in coding of specific conditions or procedures also fall under the category of interpretive errors. Avoidance or detection of interpretive error includes adequate training on definitions, testing against standard charts, testing and reporting on inter-rater reliability, and re-abstraction.
- *Errors in data entry, transfer, or transformation accuracy*: These occur when data are entered into the registry inaccurately—for example, a laboratory value of 2.0 is entered as 20. Avoidance or detection of accuracy errors can be achieved through upfront data quality checks (such as ranges and data validation checks), reentering samples of data to assess for accuracy (with the percent of data to be sampled depending on the study purpose), and rigorous attention to data cleaning.
- *Errors of intention*: Examples of intentional distortion of data (often referred to as "gaming") are inflated reporting of preoperative patient risk in registries that compare risk-adjusted outcomes of surgery or selecting only cases with good outcomes to report ("cherry-picking"). Avoidance or detection of intentional error can be challenging. Some approaches include checking for consistency of data between sites, assessing screening log information against other sources (e.g., billing data), and performing onsite audits (including monitoring of source records) either at random or "for cause."

Steps for assuring data quality include:

- *Provide training*: Educate data collectors/abstracters in a structured manner.
- *Ensure data completeness*: When possible, provide sites with immediate feedback on issues such as missing or out-of-range values and logical inconsistencies.

- *Maintain data consistency*: Compare data across sites and over time and apply consistent data transformation rules across secondary data sources.
- *Use automatic data quality monitoring and alerting:* Data quality control at scale is important for secondary data sources with vast amounts of patient data. Automatic data quality trending, variance, regression monitoring, and alerting based on set thresholds can be more cost efficient.
- *Complete onsite audits for a sample of sites*: Review screening logs and procedures and/or samples of data.
- *Complete for-cause audits*: Use both predetermined and data-informed methods to identify potential sites at higher suspicion for inaccuracy or intentional errors, such as discrepancies between enrollment and screening logs, narrow data ranges, and overly high or low enrollment.

To further minimize or identify these errors and to ensure the overall quality of the data, the following should be considered.

3.1.1 A Designated Individual Accountable for Data Quality at Each Site

Sites submitting data to a registry should have at least one person who is accountable for the quality of these data, irrespective of whether the person is collecting the data as well. The site coordinator should be fully knowledgeable of all protocols, policies, procedures, and definitions in a registry. The site coordinator should ensure that all site personnel involved in the registry are knowledgeable and that all data transmitted to registry coordinating centers are valid and accurate.

3.1.2 Assessment of Training and Maintenance of Competency of Personnel

Thorough training and documentation of maintenance of competency, for both site and registry personnel, are imperative to the quality of the registry. A detailed and comprehensive operations manual, as described earlier, is crucial for the proper training of all personnel involved in the registry. Routine cognitive testing (surveys) of healthcare provider knowledge of patient registry requirements and appropriate product use should be performed to monitor maintenance of the knowledge base and compliance with patient registry requirements. Retraining programs should be initiated when survey results provide evidence of lack of knowledge maintenance. All registry training programs should provide means by which the knowledge of the data collectors about their registries and their competence in data collection can be assessed on a regular basis, particularly when changes in procedures or definitions are implemented.

3.1.3 Data Quality Audits

As described above, the level to which registry data will be cleaned is influenced by the objectives of the registry, the type of data being collected (e.g., clinical data vs. economic data), the sources of the data (e.g., primary vs. secondary), and the timeframe of the registry (e.g., 3-month followup vs. 10-year followup). These registry characteristics often affect the types and number of data queries that are generated, both electronically and manually. In addition to

identifying missing values, incorrect or out-of-range values, or responses that are logically inconsistent with other responses in the database, specifically trained registry personnel can review the data queries to identify possible error trends and to determine whether additional site training is required. For example, such personnel may identify a specific patient outcome question or eCRF field that is generating a larger than average proportion of queries, either from one site or across all registry sites. Using this information, the registry personnel can conduct targeted followup with the sites to retrain them on the correct interpretation of the outcome question or eCRF field, with the goal of reducing the future query rate on that particular question or field. These types of "training tips" can also be addressed in a registry newsletter as a way to maintain frequent but unobtrusive communication with the registry sites.

If the registry purpose requires more stringent verification of the data being entered into the database by registry participants, registry planners may decide to conduct audits of the registry sites. Like queries discussed above, the audit plan for a specific registry will be influenced by the purpose of the registry, the type of data being collected, the source of the data, and the overall timeframe of the registry. In addition, registry developers must find the appropriate balance between the extensiveness of an audit and the impact on overall registry costs. Based on the objectives of the registry, a registry developer can define specific data fields (e.g., key effectiveness variables or adverse event data) on which the audit can be focused.

The term *audit* may describe examination or verification, may take place onsite (sometimes called monitoring) or offsite, and may be extensive or very limited. The audit can be conducted on a random sample of participating sites (e.g., 5 to 20 percent of registry sites); "for cause" (meaning only when there is an indication of a problem, such as one site being an outlier compared with most others); on a random sample of patients; or using sampling techniques based on geography, practice setting (academic center vs. community hospital), patient enrollment rate, or query rate ("risk-based" audit strategy).

The approach to auditing the quality of the data should reflect the most significant sources of error with respect to the purpose of the registry. This is true for both primary and secondary sources of data. For example, registries used for performance measurement may have a higher risk of exclusion of higher risk patients ("cherry-picking"), and the focus of an audit might be on external sources of data to verify screening log information (e.g., billing data) in addition to data accuracy. Finally, the timeframe of the registry may help determine the audit plan. A registry with a short followup period (e.g., 3 months) may require only one round of audits at the end of the study, prior to database lock and data analysis. For example, in the OPTIMIZE-HF registry, a data quality audit was performed, based on predetermined criteria, on a 5-percent random sample of the first 10,000 patient records verified against source documents.[16] For registries with multiyear followup, registry personnel may conduct site audits every 1 or 2 years for the duration of the registry.

In addition to the site characteristics mentioned above, sites that have undergone significant staffing changes during a multiyear registry should be considered prime audit targets to help confirm adequate training of new personnel and to quickly address possible inter-rater variability. To minimize any impact on the observational nature of the registry, the audit plan should be documented in the registry manual.

Subsequent to audits (onsite or remote), communication of findings with site personnel should be conducted face to face, along with followup written communication of findings and opportunities for improvement. As appropriate to meet registry objectives, the sponsor may request corrective actions from the site. Site compliance may also be enhanced with routine communication of data generated from the patient registry system to the site for reconciliation.

3.2 Registry Procedures and Systems

3.2.1 External Audits of Registry Procedures

If registry developers determine that external audits are necessary to ensure the level of quality for the specific purpose(s) of the registry, these audits should be conducted in accordance with pre-established criteria. Pre-established criteria could include monitoring of sites with high patient enrollment or with prior audit history of findings that require attention, or monitoring could be based on level of site experience, rate of serious adverse event reporting, or identified problems. The registry coordinating center may perform monitoring of a sample of sites, which could be focused on one or several areas. This approach could range from reviewing procedures and interviewing site personnel, to checking screening logs, to monitoring individual case records.

The importance of having a complete and detailed registry manual that describes policies, structures, and procedures cannot be overemphasized in the context of quality assurance of registry procedures. Such a manual serves both as a basis for conducting the audits and as a means of documenting changes emanating from these audits. As with data quality audits, feedback of the findings of registry procedure audits should be communicated to all stakeholders and documented in the registry manual.

3.2.2 Assurance of System Integrity and Security

All aspects of data management processes should fall under a rigorous life-cycle approach to system development and quality management. Each process is clearly defined and documented. The concepts described below are consistent across many software industry standards and healthcare industry standards (e.g., 21 CFR Part 11, legal security standards), although some specifics may vary. An internal quality assurance function at the registry coordinating center should regularly audit the processes and procedures described. When third parties other than the registry coordinating center perform activities that interact with the registry systems and data, they are typically assessed for risk and are subject to regular audits by the registry coordinating center.

3.2.3 System Development and Validation

All software systems used for patient registries should follow the standard principles of software development, including following one of the standard software development life-cycle (SDLC) models that are well described in the software industry.

In parallel, quality assurance of system development uses approved specifications to create a validation plan for each project. Test cases are created by trained personnel and systematically

executed, with results recorded and reviewed. Depending on regulatory requirements, a final validation report is often written and approved. Unresolved product and process issues are maintained and tracked in an issue tracking or CAPA (Corrective Action/Preventive Action) system.

Processes for development and validation should be similarly documented and periodically audited. The information from these audits is captured, summarized, and reviewed with the applicable group, with the aim of ongoing process improvement and quality improvement.

3.3 Security

All registries maintain health information, and therefore security is an important issue. Chapter 7 discusses applicable Federal laws and regulations. This section discusses some of the components of a security program. Security is achieved not simply through technology but by clear processes and procedures. Overall responsibility for security is typically assigned. Security procedures are well documented and posted. The documentation is also used to train staff. Some registries may also maintain personal information, such as information needed to contact patients to remind them to gather or submit patient-reported outcome information. Like any large databases, registries may be vulnerable to cybersecurity threats. Registries should assess these risks and develop appropriate mitigation strategies, which may include some of the security components described below. However, a full discussion of cybersecurity as it relates to registries is beyond the scope of this document.

3.3.1 System Security Plan

A system security plan consists of documented policies and standard operating procedures defining the rules of systems, including administrative procedures, physical safeguards, technical security services, technical security mechanisms, electronic signatures, and audit trails, as applicable. The rules delineate roles and responsibilities. Included in the rules are the policies specifying individual accountability for actions, access rights based on the principle of least privilege, and the need for separation of duties. These principles and the accompanying security practices provide the foundation for the confidentiality and integrity of registry data. The rules also detail the consequences associated with noncompliance.

3.3.2 Security Assessment

Clinical data maintained in a registry can be assessed for the appropriate level of security. Standard criteria exist for such assessments and are based on the type of data being collected. Part of the validation process is a security assessment of the systems and operating procedures. One of the goals of such an assessment is effective risk management, based on determining possible threats to the system or data and identifying potential vulnerabilities.

3.3.3 Education and Training

All staff members of the registry coordinating center should be trained periodically on aspects of the overall systems, security requirements, and any special requirements of specific patient

registries. Individuals should receive training relating to their specific job responsibilities and document that appropriate training has been received.

3.3.4 Access Rights

Access to systems and data should be based on the principles of least privilege and separation of duties. No individual should be assigned access privileges that exceed job requirements, and no individual should be in a role that includes access rights that would allow circumvention of controls or the repudiation of actions within the system. In all cases, access should be limited to authorized individuals.

3.3.5 Access Controls

Access controls provide the basis for authentication and logical access to critical systems and data. Since the authenticity, integrity, and auditability of data stored in electronic systems depend on accurate individual authentication, management of electronic signatures (discussed below) is an important topic.

Logical access to systems and computerized data should be controlled in a way that permits only authorized individuals to gain access to the system. This is normally done through a unique access code, such as a unique user ID and password combination that is assigned to the individual whose identity has been verified and whose job responsibilities require such access. The system should require the user to change the password periodically and should detect possible unauthorized access attempts, such as multiple failed logins, and automatically deauthorize the user account if they occur. The identification code can also be an encrypted digital certificate stored on a password-protected device or a biometric identifier that is designed so that it can be used only by the designated individual.

Rules should be established for situations in which access credentials are compromised. New password information should be sent to the individual by a secure method.

Intrusion detection and firewalls should be employed on sites accessible to the Internet, with appropriate controls and rules in place to limit access to authorized users. Desktop systems should be equipped with antivirus software, and servers should run the most recent security patches. System security should be reviewed throughout the course of the registry to ensure that management, operational, personnel, and technical controls are functioning properly.

3.3.6 Data Enclaves

With the growth of clinical data and demands for increasing amounts of clinical data by multiple parties and researchers, new approaches to access are evolving. Data enclaves are secure, remote-access systems that allow researchers to share respondents' information in a controlled and confidential manner.[17] The data enclave uses statistical, technical, and operational controls at different levels chosen for the specific viewer. This can be useful both for enhancing protection of the data and for enabling certain organizations to access data in compliance with their own organization or agency requirements. Data enclaves also can be used to allow other researchers

to access a registry's data in a controlled manner. With the growth of registries and their utility for a number of stakeholders, data enclaves have become increasingly important.[18]

3.3.7 Electronic Signatures

Electronic signatures provide one of the foundations of individual accountability, helping to ensure an accurate change history when used in conjunction with secure, computer-generated, time-stamped audit trails. Most systems use an electronic signature. For registries that report data to FDA, such signatures must meet criteria specified in 21 CFR Part 11 for general signature composition, use, and control (sections 11.100, 11.200, and 11.300). However, even registries that do not have such requirements should view these as reasonable standards. Before an individual is assigned an electronic signature, it is important to verify the person's identity and train the individual in the significance of the electronic signature. In cases where a signature consists of a user ID and a password, both management and technical means should be used to ensure uniqueness and compliance with password construction rules. Password length, character composition, uniqueness, and validity life cycle should be based on industry best practices and guidelines published by the National Institute of Standards and Technology. Passwords used in electronic signatures should abide by the same security and aging constraints as those listed for system access controls.

3.3.8 Validation

Systems that store electronic records (or depend on electronic or handwritten signatures of those records) that are required to be acceptable to FDA must be validated according to the requirements set forth in the 21 CFR Part 11 Final Rule,[19] dated March 20, 1997. The rule describes the requirements and controls for electronic systems that are used to fulfill records requirements set forth in agency regulations (often called "predicate rules") and for any electronic records submitted to the agency. FDA publishes nonbinding guidance documents from time to time that outline its current thinking regarding the scope and application of the regulation. The current guidance document is Guidance for Industry: Part 11, Electronic Records; Electronic Signatures – Scope and Application,[20] dated August 2003. In June 2017, FDA published draft guidance to clarify, update, and expand upon recommendations in the August 2003 guidance that pertain to clinical investigations conducted under 21 CFR 25 parts 312 and 812.[21]

Other documents that are useful for determining validation requirements of electronic systems are Guidance for Industry: Computerized Systems Used in Clinical Investigations,[22] dated May 2007; General Principles of Software Validation; Final Guidance for Industry and FDA Staff,[23] dated January 11, 2002; and Guidance for Industry: Electronic Source Data in Clinical Investigations, dated September 2013.[24]

4. Resource Considerations

Costs for registries can be highly variable, depending on the registry purpose and objectives. Each of the elements described in this chapter has an associated cost. Table 11-1 provides a list of some of the activities of the registry coordinating center as an example. Not all registries will

require or can afford all of the functions, options, or quality assurance techniques described in this chapter. Registry planners must evaluate benefit versus available resources to determine the most appropriate approach to achieve their goals.

Table 11-1. Data activities performed during registry coordination

Data Management	• Defines all in-process data quality control steps, procedures, and metrics. • Defines the types of edit checks that are run against the data. • Defines required file-format specifications for electronic files, as well as schedules and processes for transfers of data. • Defines quality acceptance criteria for electronic data, as well as procedures for handling exceptions. • Develops guidelines for data entry. • Identifies areas of manual review where electronic checks are not effective. • Develops and maintains process for reviewing, coding, and reporting adverse event data. • Develops and maintains archiving process. • Develops and documents the process for change management. • Develops and maintains process for query tracking and creates standard reports to efficiently identify outstanding queries, query types per site, etc. • Relates queries to processes and activities (e.g., CRF design) requiring process improvements. • Follows up on query responses and errors identified in data cleaning by performing accurate database updates. • Defines registry-specific dictionaries and code lists. • Performs database audits as applicable. • Conducts user testing of systems and applications per written specifications. • Establishes quality criteria and quality error rate acceptance limits. • Evaluates data points that should be audited and identifies potential sources of data errors for audits. • Identifies root cause of errors in order to recommend change in process/technology to ensure the error does not occur again (continuous improvement). • Ensures that sampling audit techniques are valid and support decisions made about data. • Outlines all other data flow, including external data sources.
	• Documents the process, procedures, standards, and checklist(s) and provides training. • Documents and maintains process and standards for identifying signals and trends in data. • Documents database quality control actions performed.
	• Generates standard reports of missing data from the patient database. • Creates tools to track and inventory CRFs, and reports anticipated vs. actual CRF receipts.

CRF = case report form.

References for Chapter 11

1. Clinical Data Interchange Standards Consortium. http://www.cdisc.org. Accessed June 10, 2019.

2. National Institutes of Health. National Institutes of Health Stroke Scale. https://www.stroke.nih.gov/documents/NIH_Stroke_Scale_508C.pdf. Accessed June 10, 2019.

3. Luck J, Peabody JW, Dresselhaus TR, et al. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. Am J Med. 2000;108(8):642-9. PMID: 10856412.

4. Reisch LM, Fosse JS, Beverly K, et al. Training, quality assurance, and assessment of medical record abstraction in a multisite study. Am J Epidemiol. 2003;157(6):546-51. PMID: 12631545. DOI: 10.1093/aje/kwg016.

5. Neale R, Rokkas P, McClure RJ. Interrater reliability of injury coding in the Queensland Trauma Registry. Emerg Med (Fremantle). 2003;15(1):38-41. PMID: 12656785.

6. eMERGE Network. Publications List. https://emerge.mc.vanderbilt.edu/publications/. Accessed June 10, 2019.

7. Phenotype KnowledgeBase. https://phekb.org/. Accessed June 10, 2019.

8. Apache cTAKES. http://ctakes.apache.org. Accessed June 10, 2019.

9. Clinical Language Annotation, Modeling, and Processing Toolkit. https://clamp.uth.edu. Accessed June 10, 2019.

10. MetaMap - A Tool For Recognizing UMLS Concepts in Text. National Library of Medicine. https://metamap.nlm.nih.gov. Accessed June 10, 2019.

11. Meystre SM, Savova GK, Kipper-Schuler KC, et al. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform. 2008:128-44. PMID: 18660887.

12. U.S. Food and Drug Administration. Guidance for Industry. E6 Good Clinical Practice: Consolidated Guidance. Apr, 1996. https://clinicalcenter.nih.gov/ccc/clinicalresearch/guidance.pdf. Accessed June 10, 2019.

13. HL7. http://www.hl7.org/. Accessed June 10, 2019.

14. National Institute of Standards and Technology. http://www.nist.gov. Accessed June 10, 2019.

15. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006;354(4):353-65. PMID: 16436767. DOI: 10.1056/NEJMoa051379.

16. Gheorghiade M, Abraham WT, Albert NM, et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA. 2006;296(18):2217-26. PMID: 17090768. DOI: 10.1001/jama.296.18.2217.

17. National Institutes of Health. NIH Data Sharing Policy and Implementation Guidance. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#enclave. Accessed June 10, 2019.

18. Platt R, Lieu T. Data Enclaves for Sharing Information Derived From Clinical and Administrative Data. JAMA. 2018;320(8):753-4. PMID: 30083726. DOI: 10.1001/jama.2018.9342.

19. U.S. Food and Drug Administration. CFR - Code of Federal Regulations Title 21, Part 11: Electronic Records; Electronic Signatures. https://www.ecfr.gov/cgi-bin/text-idx?SID=0e09bff792b1f80bc74436ed9d1eea4c&mc=true&tpl=/ecfrbrowse/Title21/21cfr11_main_02.tpl. Accessed June 10, 2019.

20. U.S. Food and Drug Administration. Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application. https://www.fda.gov/media/75414/download. Accessed June 10, 2019.

21. U.S. Food and Drug Administration. Use of Electronic Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 – Questions and Answers. DRAFT Guidance for Industry. https://www.fda.gov/media/105557/download. Accessed June 10, 2019.

22. U.S. Food and Drug Administration. Guidance for Industry Computerized Systems Used in Clinical Investigations. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf. Accessed June 10, 2019.

23. U.S. Food and Drug Administration. General Principles of Software Validation; Final Guidance for Industry and FDA Staff. https://www.fda.gov/media/73141/download. Accessed June 10, 2019.

24. U.S. Food and Drug Administration. Guidance for Industry Electronic Source Data in Clinical Investigations. https://www.fda.gov/media/85183/download. Accessed June 10, 2019.

Chapter 12. Adverse Event Detection, Processing, and Reporting

1. Introduction

Registries that collect information on specific drugs and medical devices must anticipate the need for adverse event (AE) detection, processing, and reporting. This chapter addresses the identification, processing, and reporting of AEs detected in situations in which a registry has contact with individual patients. This document is not a formal regulatory or legal document; therefore, any information or suggestions presented herein do not supersede, replace, or otherwise interpret Federal guidance documents that touch on these subjects. In addition, the discussion below focuses on U.S. regulations; different requirements may apply to AE detection, processing, and reporting in other countries. Registry sponsors are encouraged to discuss plans for AE collection and processing with local health authorities when planning a registry.

This chapter primarily focuses on AEs related to pharmaceutical products. Medical devices are significantly different from pharmaceutical products in the manner in which AEs and product problems (complaints) present themselves, in the etiology of their occurrence, and in the regulation governing the defining and reporting of these occurrences, as well as postapproval study requirements. Other sources provide more information about defining and reporting device-related AEs and product problems, and about postmarketing studies (including those involving registries).[1-3]

2. Identifying and Reporting Adverse Drug Events

The U.S. Food and Drug Administration (FDA) defines an adverse drug experience as any AE associated with the use of a drug in humans, whether or not considered drug related,[4] while the International Conference on Harmonisation (ICH) guideline ICH E2A similarly defines an AE as an untoward medical occurrence in a patient administered a pharmaceutical product, whether or not the occurrence is related to or considered to have a causal relationship with the treatment.[5]

For marketed products regulated by FDA, AEs are categorized for reporting purposes according to the seriousness and expectedness of the event (i.e., whether the event was previously observed and included in local product labeling), as it is presumed that all spontaneously reported events are potentially related to the product for the purposes of FDA reporting. Prior to marketing approval, relatedness is an additional determinant for reporting events occurring during clinical trials or preclinical studies associated with investigational new drugs and biologics. For AEs occurring in postapproval studies and reported during planned contacts and active solicitation of information from patients, as when registries collect data regarding one or more FDA-approved products,[6,7] the requirements for mandatory reporting also include whether there is a reasonable possibility that the drug caused the adverse experience.[4] For registries that do not actively solicit AEs, incidentally reported events (e.g., those reported during clinician or consumer contact for another purpose) should typically be handled and evaluated as spontaneously reported events.

The medical device reporting regulations differ from those for drugs and biologics in that reportable events include both AEs and problems with the device itself.[8] Medical device reporting is required for incidents in which the device may have caused or contributed to a death or serious injury, or may have malfunctioned and would likely cause or contribute to death or serious injury if the malfunction were to recur.[9]

Most registries have the opportunity to identify and capture information on AEs for biopharmaceutical products and/or medical devices. Since the passing of the FDA Amendments Act in September 2007 and the resulting increased emphasis on ongoing monitoring of safety profiles, evaluation of risks unknown at the time of product approval, and proactive detection of potential safety issues, registries increasingly are used to fulfill safety-related objectives.[10] Even when registries are not required to capture and process AE reports, there is an implicit requirement from the perspective of systematic data collection and promoting public health: any individual who believes a serious risk may be associated with exposure to a medical product should be encouraged to report this AE either to the product sponsor or directly to FDA. The FDA maintains MedWatch, a Web-based reporting system that allows consumers and health professionals to voluntarily report serious adverse events and other serious problems that they suspect are associated with the use of an FDA-regulated product.[11]

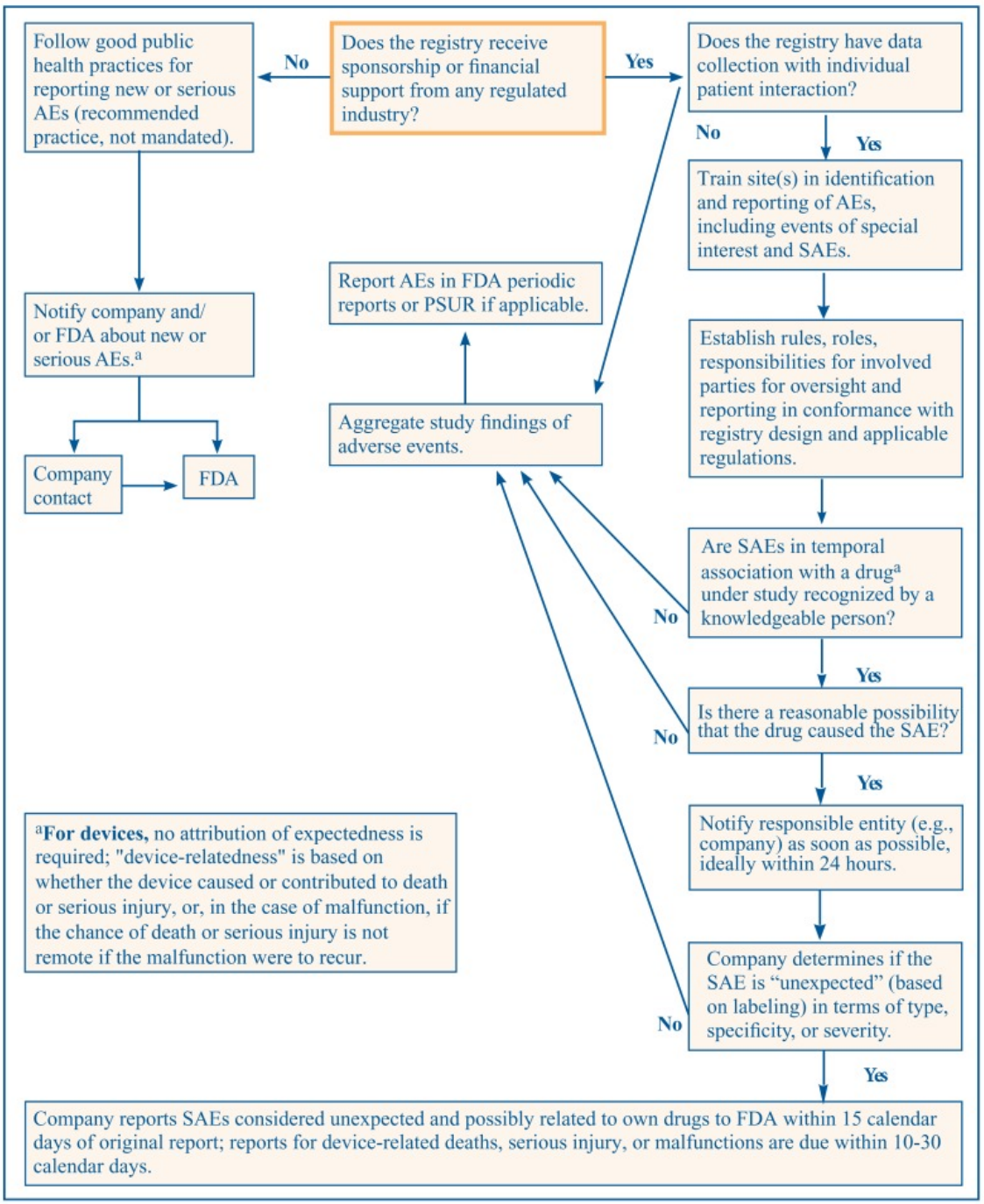
In the United States, the minimum dataset required to report an AE is (1) an identifiable patient, (2) an identifiable reporter, (3) suspect product, and (4) adverse event. However, in addition to direct data collection, AEs can be detected through retrospective analysis of a population database, where direct patient or healthcare provider contact does not occur. Patient interactions include clinical interactions and data collection by phone, Internet, or other means; perusal of electronic medical records or insurance claims data would not be considered direct patient interaction. Reporting is rarely required for individual AEs observed in aggregate population data, since there is no direct patient interaction where an association might be suggested or inferred. Nevertheless, if aggregate or epidemiologic analyses suggest that an AE is associated with exposure to a drug or medical product, it is desirable that the minimum dataset information be forwarded to the manufacturer of the product, who will determine any need for, and timing of, reporting of study results to the relevant regulatory authorities.

Figure 12-1 provides a broad overview of the reporting requirements for AEs and shows how the reporting differs according to whether the registry has direct patient interaction, and whether it receives sponsorship and/or financial support from a regulated industry.[12] These industries may include entities with products subject to FDA regulation, including products with FDA approval, an FDA-granted license, and investigational products; and other entities such as manufacturers, user facilities, and distributors.

All AE reporting begins with a suspicion by the physician (or responsible person who obtains or receives information) that a patient exposed to a medicinal product has experienced some AE and that the event has a reasonable possibility of being causally related to the product being used; this is referred to as the “becoming aware” principle. Some registries also collect and record AEs reported directly by the patients or their caregivers. It is important to develop a plan for detecting, processing, and reporting AEs for any registry that has direct patient contact. If the registry receives sponsorship in whole or part from a regulated industry (for drugs or devices),

the sponsor has mandated reporting requirements, including stringent timelines. AE reporting requirements for registry sponsors are discussed later in this chapter.

Figure 12-1. Overview of reporting requirements for registries related to drug/biologic therapies



Best practices for adverse event reporting to FDA by registries of postmarket products.

AE = adverse event; SAE = serious adverse event; FDA = U.S. Food and Drug Administration; PSUR = periodic safety update report.

Prior to registry launch, the process for detecting and reporting AEs should be established in collaboration with the sponsor and any oversight committees. (See Chapter 2.) Once the plans have been developed, the registry operator or sponsor should provide training to the physicians or other responsible parties (referred to as "sites" hereafter) on how to identify AEs and to whom they should be reported. AE reporting is based on categorization of the AE according to the seriousness of the event, its expectedness based on product labeling, and presumed causality or possible association with use of the product, as follows:

- *Seriousness*: Serious AEs (SAEs) include events that result in death, are life threatening (an event in which the patient was at risk of death at the time of the event), require or prolong inpatient hospitalization, result in persistent or significant disability or incapacity, or result in a congenital anomaly. Important medical events may also be considered serious when, based on medical judgment, they may jeopardize the person exposed and may require medical or surgical intervention to prevent one of the outcomes listed above (e.g., death or prolonged hospitalization).
- *Expectedness*: All AEs that are previously unobserved or undocumented are referred to as "unexpected," in that their nature and severity are not consistent with information provided in the relevant product information (e.g., approved professional package insert or product label). Determination of expectedness is made by the sponsor on a case-by-case basis. Expected events typically do not require expedited reporting to the regulatory authorities.
- *Relatedness*: Relatedness is a term intended to indicate that a determination has been made that the event had a reasonable possibility of being related to exposure to the product. This assessment of causality may be based on factors such as biological plausibility, prior experience with the product, and temporal relationship between product exposure and onset of the event, as well as dechallenge (discontinuation of the product to determine if the AE resolves) and rechallenge (reintroduction of the product to determine if the AE recurs). Many terms and scales are used to describe the degree of causality, including terms such as certainly, definitely, probably, possibly, or likely related or not related, but there is no standard nomenclature.[13] All spontaneous reports have an implied causal relationship as per regulatory guidance, regardless of the reporter's assessment.

The registry may use forms such as a structured questionnaire or an AE case report form to collect the information from providers or patients. When solicitation of AEs is not prespecified in the registry's operating plans, the registry may permit AE detection by asking general questions to solicit events, such as "Have you had any problems since your last visit or since we last spoke?" and then following up any such reports with probes as to what happened, diagnoses, and other documentation. This practice is not required.

3. Collecting AE Data in a Registry

There are two key considerations regarding AE collection as part of a registry: (1) what data need to be collected to meet the registry's safety-related objectives, and (2) what processes need to be in place to ensure that the registry is in compliance with regulations regarding expedited and periodic AE event reporting, if applicable. The data fields needed for the purpose of analysis by the registry may be minimal (e.g., event and onset date), whereas a complete SAE form for a

subset of events reported to the registry may be sought to fulfill the sponsor's reporting requirements. Due to the nature of registries, the goal of collecting enough data to meet the registry's objectives must constantly be balanced with the burden on sites. To this end, the processes for AE reporting should be streamlined as much as possible.

The collection of AE data by a registry is generally either intentionally solicited (meaning that the data are part of the uniform collection of information in the registry) or unsolicited (meaning that the AE information is volunteered or noted in an unsolicited manner and not as a required data element through a case report form). As described further below, it is good practice for a registry to specify when and how AE information (and any other events of special interest) should and should not be solicited from patients by a site and, if that information has been obtained, how and when the site should inform the appropriate persons.

While an AE may be reported to the manufacturer, to FDA (e.g., via MedWatch), or to the registry itself (and then from the registry to the manufacturer), it is strongly encouraged that the protocol describe the procedures that should be followed, and that the sites be trained in these procedures as well as in their general obligations and the relevant public health considerations. A separate safety reporting plan that fully identifies the responsible parties and describes the operational considerations may also be considered to ensure that potentially reportable information is evaluated in an appropriate timeframe, and, for manufacturer-sponsored registries, in accordance with any applicable standard operating procedures. This type of plan also should describe how deviations or systemic failures in detection and reporting processes will be identified, addressed, and considered for corrective action.

Determining whether a registry should use a case report form to collect AEs should be based on the principles described in Chapter 5, which refer to the scientific importance of the information for evaluating the specified outcomes of interest. This may mean that all, some, or no AEs are collected on the case report forms. However, if some AEs are collected in an intentional, solicited manner (e.g., routine collection of a primary or secondary outcome via an AE case report form) and others come to the registry's attention in an unsolicited, "spontaneous" way (e.g., when an AE is reported in the course of a registry contact, such as a call to the sponsor or to registry support staff), then from a practical perspective it is even more important to have a clear process, so that AEs that require reporting are identified. In this scenario, one best practice that is often used in electronic registry studies is to have a notification sent promptly to the sponsor's safety group when a case report form is submitted that contains specific or potential information indicating that a serious AE has occurred. This process allows for rapid followup by the sponsor, as needed.

4. AE Reporting by the Registry

Once suspicion has been aroused that an unexpected serious event has a reasonable possibility of being causally related to a drug, the AE should be reported to FDA through MedWatch, to the company that manufactures the product, or to the registry coordinating center. (See Chapter 11.) A system should be developed such that all appropriate events are captured and duplicate reporting is avoided to the extent possible. Generally, AE reports are submitted directly by the site or by the registry to the manufacturer, since they are often most efficient at evaluating,

processing, and reporting for regulatory purposes within the required time periods. Alternatively, sites could be instructed to report AEs directly to FDA according to their normal practices for marketed products; however, this often means that the companies are not notified of the AE and are not able to follow up or evaluate the event in the context of their safety database. In fact, companies are not necessarily notified by FDA if an AE report comes directly to FDA, since only certain reports are shared with industry, and reporters have an option to request that the information not be shared directly with the company.[14]

Systematic collection of all AEs provides a unique resource of consistent and contemporaneously collected comparison information that can be used at a later date to conduct epidemiologic assessments. Ideally, the practice for handling AEs and SAEs should be applied to all treatments (including comparators) recorded in the registry, so that all subjects are treated similarly. In fact, a strong advantage of registries with systematic data collection and internal comparators is that they provide both numerators and denominators for safety events; thus, reporting of comparative known AE rates in the context of a safety evaluation provides valuable information on real-world performance. The contrast with comparators helps promote clarity about whether the observed effects are unique to the product, unique to a class, or common to the condition being treated. Reporting AEs without denominator information is less useful from a surveillance perspective since events rates cannot be calculated without both numerators and denominators. The reliability of the denominator should always be judged, however, by considering the likelihood that all events were reported appropriately.

For postapproval registries not financially supported by pharmaceutical companies, healthcare providers at registry sites should be instructed that if they suspect or otherwise become aware of a serious AE that has a reasonable possibility of being causally related to a drug or product, they should report the event directly to the product manufacturer (who must then report to FDA under regulation) or to FDA's MedWatch program (or local health authority if the study is conducted outside of the United States). Reporting can be facilitated by providing the MedWatch Form 3500,[15] information regarding the process for submission, and MedWatch contact information.

For registries that are sponsored or financially supported in full or in part by a regulated industry and that study a single product, the most efficient monitoring system to avoid duplicate reporting is one in which all physicians participating in the registry report all AEs (or SAEs only) directly to the sponsor or centralized designated responsible personnel, who then reports to the regulatory authorities. However, when products other than those exclusively manufactured by the sponsor are involved, including other treatments, sponsors will need to determine how to process AE reports received for these other products. Since sponsors are not obligated to report AEs for their competitors, it is good practice from a public health perspective to specify how the site should address those AEs (e.g., whether to report directly to the other product's manufacturer or to FDA). Options for the sponsor include (1) recommending that sites report the AEs of comparators directly to the manufacturer or to FDA; (2) collecting all AEs and forwarding the AE report directly to the comparator's manufacturer (who would then, in turn, report to FDA); and (3) actually reporting the AE for the comparator product directly to FDA. As standard practice in pharmacovigilance, many sponsors report events potentially associated with another manufacturer's drug to that manufacturer's safety department as a courtesy, rather than report events directly to FDA, and choose to continue that practice when conducting a registry or other observational study.

Some disease registries are not focused on a specific product, but rather on conducting natural history studies or evaluating treatment patterns and outcomes in a particular patient population prior to marketing approval of the sponsor's product. In these situations, it is recommended that sites follow their own standard practices for spontaneous AE reporting, including reporting any events associated with a product known to be manufactured by the sponsor.

In most circumstances in which a serious drug-associated AE is suspected, sites are encouraged to submit supportive data to sponsors, such as laboratory values, vital signs, and examination results, along with the SAE report form. If the event is determined to be an AE, the sponsor will include it in the safety database, evaluate it internally, and transfer the AE report to the regulatory authorities if required. It should be noted that the regulations represent minimum requirements for compliance; special circumstances for a particular product may result in additional events being reportable (e.g., expected events of particular interest to regulators). It should not be expected that registry participants be aware of all the reporting nuances associated with a particular product. To the extent possible, guidance on reporting events of special interest should be provided in the protocol and in any safety training.

If an external party manages a registry, SAEs should be submitted to the sponsors as quickly as possible after the registry becomes aware of the event. In this situation, the registry is an agent of the sponsor, and FDA's 15-calendar-day reporting requirement starts as soon as the event has come to the attention of the registry. (See Section 7 below.) This submission can be accomplished by phone or fax, or by means of automated rules built into the vehicle used for data collection (such as automatic triggers that can be designed into electronic data capture programs). For direct regulatory submissions, the MedWatch Form 3500A[16] should be used for postapproval reporting for drugs and therapeutic biologics unless other means of submission are agreed upon. For vaccines, the Vaccine Adverse Event Reporting System should be consulted.[17] Foreign events may be submitted on a CIOMS form (the World Health Organization's Council for International Organizations of Medical Sciences),[8,18,19] or a letter can be generated that includes the relevant information in narrative format.

5. Coding

Coding AEs into a standard nomenclature should be done by trained experts to ensure accuracy and consistency. Reporters, patients, healthcare providers, and registry personnel should do their best to capture the primary data clearly, completely, and in as "natural" clinical language as possible. Since reporters may use different verbatim terms to describe the same event, it is recommended that sponsors apply coding conventions to code the verbatim terms. The Medical Dictionary for Regulatory Activities (MedDRA®) is customarily used throughout the product development cycle and as part of pharmacovigilance; however, other coding systems are also used. For example, SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) is used instead of MedDRA in some electronic health records. Coding the different verbatim language to preferred terms allows similar events to be appropriately grouped, creates consistency among the terms for evaluation, and maximizes the likelihood that safety signals will be detected.

Sponsors or their designees should review the accuracy of the coding of verbatim AEs into appropriate terms. If coding is performed by someone other than the sponsor, any applicable coding conventions associated with the underlying condition or product should be shared. Review of the coding process should focus on terms that do not accurately communicate the severity or magnitude of the AE or possibly mischaracterize the AE. Review of the coded terms compared with reported verbatim terms should be performed in order to ensure consistency and accuracy of the AE reporting and to minimize variability of coding of similar AE terms. Attention to consistency is especially important, as many different individuals may code AEs over time, and this situation contributes to variability in the coding process. In addition to monitoring AEs individually for complete clinical evaluation of the safety data, sponsors should consider grouping and analyzing clinically relevant coded terms that could represent similar toxicities or syndromes. Combining terms may provide a method of detecting less common and serious events that would otherwise be obscured. However, sponsors should be careful when combining related terms to avoid amplifying a weak signal or obscuring important overall findings when grouping is overly broad. In addition to monitoring individual AEs, sites and registry personnel should be attentive to toxicities that may cluster into syndromes.

6. Adverse Event Management

In some cases, such as when a safety registry is created as a condition of regulatory approval, a data safety monitoring board (DSMB), data monitoring committee (DMC), or adjudication committee may be established with the primary role of periodically reviewing the data as they are generated by the registry. Such activities are generally discussed directly with the regulatory authorities, such as FDA. These authorities are typically involved in the design and critique of protocols for postapproval studies. Ultimately, registry planning and the registry protocol should anticipate and clearly delineate the roles, responsibilities, processes, forms, and lines of communication for AE reporting for sites, registry personnel, the DSMB, DMC, or adjudication committee if one exists, and the sponsoring organization. Documentation should be provided for definitions and approaches to determining what is considered unexpected and possibly related to drug or device exposure. The management of AE reporting should be clearly specified in the registry protocol, including explanations of the roles, responsibilities, processes, and methods for handling AE reports by the various parties conducting the registry, and for performing followup activities with the site to ensure that complete information is obtained. Sponsors who are stakeholders in a registry should have a representative of their internal drug safety or pharmacovigilance group participate in the design and review of the registry protocol and have a role in the data collection and reporting process (discussed in Chapter 2) to facilitate appropriate and timely reporting and communication.

For postapproval studies financially sponsored by manufacturers, the overall company AE monitoring systems are usually operated by personnel experienced in drug safety (also referred to as pharmacovigilance, regulatory safety, product safety, and safety and risk management). If sites need to report or discuss an AE, they can call the contact number provided for the registry, and are then prompted to press a number if reporting an AE. This number then transfers them to drug safety surveillance so that they can interact directly with personnel in this division and bypass the registry coordinating group. These calls may or may not be tracked by the registry. Alternatively, the registry system can provide instructions to the site on how to report AEs

directly to the sponsor's drug safety surveillance division. By this method, the sponsor provides a separate contact number for AE reporting (independent of the registry support staff) that places the site in direct contact with drug safety personnel. This process minimizes the possibility of duplicate AE reports and the potentially complicated reconciliation of two different systems collecting AE information. Use of this process is critical when dealing with products that are available via a registry system as well as outside of a registry system, and it allows sites to have one designated drug safety representative for interaction.

Sponsors of registries designed specifically for surveillance of product safety are strongly encouraged to hold discussions with the regulatory authorities when considering the design of the AE monitoring system. These discussions should be focused on the purpose of the registry, the "best fit" model for AE monitoring, and the timing of routine registry updates. Health authorities expect compliance with the agreed-upon requirements with respect to the internal operations chosen by the sponsor to support the requirements of an AE monitoring system. Details regarding implementation are the responsibility of the sponsor.

It should also be noted that FDA's Proposed Rule for Safety Reporting Requirements for Human Drug and Biologics Products (68 FR 12405, March 14, 2003) suggests that the responsible point of contact for FDA should be provided for all expedited and periodic AE reports, and preferably, this individual should be a licensed physician. Although this proposed rule has never been finalized, the principle is similar to the Qualified Person for Pharmacovigilance (QPPV) in Europe, whereby a specific, qualified individual is identified to provide responses to health authorities, upon request, including those regarding AEs reported via the registry system.

7. Adverse Event Required Reporting for Registry Sponsors

The reporting requirements of the sponsor directly affect how registries should collect and report AEs. Sponsors that are regulated industries are subject to the requirements shown in Table 12-1. ICH guidelines describe standards for expedited reporting.[5,20] The guidelines also provide recommendations for periodic safety update reports[21] that are generally accepted globally.

Table 12-1. Overview of serious adverse event reporting requirements for marketed products

Type of Requirement	Drug and Biologics	Devices
U.S. postmarketing regulations	Primary: 21 CFR 314.80 (drugs), 21 CFR 600.80 (biologics) Other: 21 CFR 310.305, 21 CFR 314.98	21 CFR 803.20
Required reporting source	Regulated industries	Manufacturer, importer, user facility
Required reports	Serious, unexpected, and with a reasonable possibility of being related to drug exposure (with some exceptions)	Death or serious injury; device malfunction

Type of Requirement	Drug and Biologics	Devices
Alternative reports	Not applicable	Summary reports (periodic line-listing of reports of well-known events)
Timeframe for reporting	15 calendar days for expedited reports	5 workdays, 10 workdays, or 30 calendar days, depending on the source and action required
Standard reporting form	MedWatch 3500A (for mandatory reporting required of a regulated industry); MedWatch 3500 (for voluntary reporting by healthcare professionals); MedWatch 3500B (for voluntary reporting by consumers)	
Websites	http://www.fda.gov/Safety/MedWatch/default.htm	http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm#1

Requirements for regulated industries that sponsor or financially support a registry include expedited reporting of serious and unexpected AEs made known to them via spontaneous reports. For registries, the 15-calendar-day notification applies if the regulated industry believes there is a reasonable possibility that the unexpected SAE was causally related to product exposure. Best practices for international reporting are that all "affiliates" of a sponsor report serious, unexpected, and possibly related events to the sponsor in a timely fashion, ideally within 2 calendar days; this allows the sponsor, in turn, to complete notification to the responsible regulatory authority within a total of 15 calendar days. Events that do not meet the requirements of expedited reporting (such as nonserious events or serious events considered expected or not related) may require submission through inclusion in an appropriate safety update, such as the New Drug Application or Biologic Licensing Application Annual Report, Periodic Report, or Periodic Safety Update Report, as applicable.[4] In many cases, sponsors are also required to provide registry safety updates to the health authority. Thus, sponsors may coordinate registry safety updates (i.e., determining the date for creating the dataset—the data cutoff date) with the timing of the New Drug Application Annual Report, Periodic Report, Periodic Safety Update Report, or other agreed-upon periodic reporting format. Devices, however, have different reporting requirements (see http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm). In any event, sponsors should discuss safety reporting requirements for their specific registries with the applicable health authorities (such as FDA and European Medicines Agency) before finalizing their registry protocol.

In some cases, a registry sponsor may encourage sites to systematically report all potential SAEs to the sponsor. Given the potential for various assessments by different sites of the seriousness and relatedness of a particular AE—and therefore, inconsistency across sites in the evaluation of a particular AE—this method has certain advantages. In addition, expectedness is not always a straightforward assessment, and the expectedness of events can have significant variability depending on the local approved product labeling. For this reason, it is important that this determination be made by the sponsor and not the reporter of the event. Although this approach may result in substantially greater demands on the sponsor to evaluate all reports, it helps ensure compliance and avoid underreporting. Furthermore, sponsors must make their own assessments

regarding the causality of individual solicited events. This requirement typically does not affect the need for reporting, but allows the sponsor to provide its own evaluation in the full context of the safety database. For these reasons, planning for high-quality and consistent training in AE reporting requirements across sites is the preferred approach for a patient registry.

Regardless of who assesses presumed relatedness, sponsors should be prepared to manage the increased volume of AE reports, and sponsors' registry staff should be trained to understand company policy and regulations on AE reporting in order to ensure compliance with local regulations. This training includes the ability to identify and evaluate the attributes of each AE and determine whether the AE should be reported to the health authority in keeping with local regulation. Sponsors are encouraged to appoint a healthcare practitioner to this role in order to ensure appropriate assessment of the characteristics of an AE.

When biopharmaceutical or device companies are not sponsoring, financially supporting, or participating in a registry in any way, AE reporting is dependent upon the "become aware" principle. If any agent or employee of the company receives information regarding an AE report, the agent or employee must document receipt and comply with internal company policy and regulatory requirements regarding AE reporting, to ensure compliance with applicable drug and device regulations.

8. Special Case: Risk Evaluation and Mitigation Strategies (REMS)

Under the FDA Amendments Act (2007), FDA established a legally enforceable new framework for risk management of products with known safety concerns, called Risk Evaluation and Mitigation Strategies (REMS).[10,22] The purpose of REMS is to ensure that the benefits of a particular drug outweigh the risks. New REMS programs can be imposed by FDA during clinical development, as part of the approval process, or at any time postapproval, should a new safety signal be identified. Although each REMS is customized depending on the product and associated safety issues, potential components include some combination of—

- A medication guide and/or patient package insert. Medication guides are informational packets distributed with some prescription drugs, which provide important information to patients about possible side effects and drug-drug interactions. The FDA has indicated the situations in which a medication guide is required to be available and distributed to patients.[23] A medication guide alone can and frequently does constitute a REMS.
- A communication plan that specifies targeted education and outreach for physicians, pharmacists, and patients.
- Elements to assure safe use (ETASU), in some cases. ETASU may include restriction of prescribing to healthcare providers with particular training, experience, and certification; dispensing of the drug in restricted settings; documentation of safe use conditions (such as laboratory results or specific patient monitoring); and registries.[22]

Unlike the less structured disease or exposure registries discussed above, a restricted-access system associated with an ETASU is designed for approved products that have particular risk-benefit profiles that require more careful controls. The purpose of ETASU is to mitigate a certain known drug-associated risk by ensuring that product access is tightly linked to some preventive

and/or monitoring measure. Examples include systems that monitor laboratory values, such as white blood cell counts during clozapine administration to prevent severe leucopenia, or routine pregnancy testing during thalidomide administration to prevent in utero exposure of this known teratogenic compound. When these programs include registries, the registries often prospectively collect a battery of information using standardized instruments.

Data collection under ETASU may carry special AE reporting requirements, and as a result of the extensive contact with a variety of potential sources of safety information (e.g., pharmacists and patients), care should be taken to identify all possible routes of reporting. If special requirements exist, they should be made explicit in the registry protocol, with clear definitions of roles, responsibilities, and processes. Training of involved healthcare providers, such as physicians, nurses, and pharmacists, can be undertaken with written instructions, via telephone or with face-to-face counseling. Training of these healthcare providers should also extend beyond AE reporting to the specific requirements of the program in question. Such training may include the intended use and associated risk of the product, appropriate patient enrollment, and specific patient monitoring requirements, including guidelines for product discontinuation and management of AEs, as well as topics to cover during comprehensive counseling of patients. The objectives of the ETASU system and overall REMS should be clearly stated (e.g., prevention of in utero exposure during therapy via routine pregnancy testing), and registration forms that document the physician's and pharmacist's attestation of their commitment to requirements of the patient registry system should be completed prior to prescribing or dispensing the product.

9. Reporting Breaches of Confidentiality or Other Risks

In addition to addressing regulatory responsibilities for reporting adverse events, registries must also understand regulatory and ethical requirements and expectations regarding breaches of confidentiality or the reporting of other risks to patients that may arise during the course of a registry. The Health Information Technology for Economic and Clinical Health Act (HITECH Act) requires HIPAA-covered entities (entities covered by the Health Insurance Portability and Accountability Act of 1996) and their business associates to provide notification following a breach of unsecured protected health information.[24] See Chapter 7 for a detailed discussion of the HITECH Act. State breach notification laws may also apply to registry data.

Beyond these legal requirements, registries should establish clear notification procedures for breaches of confidentiality or other risks that become known during the course of the registry, whether or not they are governed by HIPAA or subject to State laws.

References for Chapter 12

1. Baim DS, Mehran R, Kereiakes DJ, et al. Postmarket surveillance for drug-eluting coronary stents: a comprehensive approach. Circulation. 2006;113(6):891-7. PMID: 16476863. DOI: 10.1161/CIRCULATIONAHA.105.569657.

2. U.S. Food and Drug Administration. Medical Device Reporting. http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm. Accessed June 10, 2019.

3. Gross TP, Witten CM, Uldriks C, et al. A view from the US Food and Drug Administration. In: Frank E, Johnson KSV, editors. The Bionic Human: Health Promotion for People with Implanted Prosthetic Devices. Vol. 5. Totowa, New Jersey: Humana Press, Inc.; 2006. pp. 61–87.

4. 21 CFR § 314.80 (2008). Revised April 1, 2018. Postmarketing reporting of adverse drug experiences.

5. ICH E2A: Clinical safety data management: definitions and standards for expedited reporting. 27 October 1994. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf. Accessed June 10, 2019.

6. U.S. Food and Drug Administration. Postapproval Pregnancy Safety Studies. DRAFT Guidance for Industry. May 2019. https://www.fda.gov/media/124746/download. Accessed June 4, 2019.

7. U.S. Food and Drug Administration. Postmarketing Adverse Experience Reporting for Human Drug and Licensed Biological Products: Clarification of What to Report. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071981.pdf. Accessed June 10, 2019.

8. 21 CFR § 312.32 (2008). Revised April 1, 2018. IND safety reporting.

9. 21 CFR § 803 (2008). Revised April 1, 2018. Medical Device Reporting.

10. Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85 (2007), Title VI. Sect. 603. Critical Path Public-Private Partnerships. https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf. Accessed June 4, 2019.

11. U.S. Food and Drug Administration. Reporting Serious Problems to the FDA. http://www.fda.gov/Safety/MedWatch/HowToReport/default.htm. Accessed June 10, 2019.

12. Dreyer NA, Sheth N, Trontell A, et al. Good practices for handling adverse events detected through patient registries. Drug Information Association Journal. 2008;5(42):421–8.

13. U.S. Food and Drug Administration. Guidance for Industry. Good pharmacovigilance practices and pharmacoepidemiologic assessment. March 2005. https://www.fda.gov/media/71546/download. Accessed June 3, 2019.

14. U.S. Food and Drug Administration. MedWatch Voluntary Reporting Form. https://www.accessdata.fda.gov/scripts/medwatch/medwatch-online.htm. Accessed June 10, 2019.

15. U.S. Food and Drug Administration. MedWatch Form FDA 3500 - Voluntary Reporting. https://www.fda.gov/media/76299/download. Accessed June 10, 2019.

16. U.S. Food and Drug Administration. MedWatch Form FDA 3500A - Mandatory Reporting. https://www.fda.gov/media/69876/download. Accessed June 10, 2019.

17. Vaccine Adverse Event Reporting System. http://vaers.hhs.gov. Accessed June 10, 2019.

18. Council for International Organizations of Medical Sciences. CIOMS Form. https://cioms.ch/wp-content/uploads/2017/05/cioms-form1.pdf. Accessed June 10, 2019.

19. 21 CFR § 314.80(f)(1) (2008). Postmarketing reporting of adverse drug experiences.

20. ICH Topic E2D. Post Approval Safety Data Management. European Medicines Agency; May 2004. CPMP/ICH/3945/03. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2D/Step4/E2D_Guideline.pdf. Accessed June 10, 2019.

21. ICH E2C R1: Clinical Safety Data Management: Periodic Updated Safety Reports for Marketed Drugs. European Medicines Agency; June 1997. CPMP/ICH/288/95. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2C/Step4/E2C_R1_Guideline.pdf. Accessed June 10, 2019.

22. U.S. Food and Drug Administration. DRAFT Guidance for Industry: Format and content of proposed risk evaluation and mitigation strategies (REMS), REMS assessments, and proposed REMS modifications. September 2009. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM184128.pdf. Accessed August 15, 2012.

23. U.S. Food and Drug Administration. Guidance: Medication Guides – Distribution Requirements and Inclusion in Risk Evaluation and Mitigation Strategies (REMS). https://www.fda.gov/media/79776/download. Accessed June 10, 2019.

24. 74 F.R. 42740 (August 24, 2009). Breach Notification for Unsecured Protected Health Information.

Chapter 13. Analysis, Interpretation, and Reporting of Registry Data To Evaluate Outcomes

1. Introduction

Registries have the potential to produce data that are an important source of information regarding healthcare patterns, decision making, and delivery, as well as the subsequent association of these factors with patient outcomes. Registries, for example, can provide valuable insight into the safety and/or effectiveness of an intervention or the efficiency, timeliness, quality, and patient centeredness of a healthcare system. Registry data may also be linked to other data sources, such as administrative health insurance claims databases, electronic health records (EHRs), or biorepositories, to investigate hypotheses or questions that are secondary to the original reason for data collection.

The utility and applicability of registry data and linked datasets (collectively, 'registry data') rely heavily on an understanding of how the data were derived and why they were recorded, and then on the quality of the data analysis plan and its users' ability to interpret the results. Analysis and interpretation of these data begin with a series of core questions:

- *Study purpose*: Is the intent descriptive or comparative and does it address natural history, effectiveness, safety or other characterization?
- *Patient population*: Who was studied and how did they come to be included in the registry?
- *Data quality*: How were the data collected and reviewed, and was any verification or validation performed?
- *Data completeness*: How were missing data handled for the main exposures and outcomes of interest and main confounders?
- *Data analysis*: What analyses were performed and how were differences in risk factors accounted for? Were any sensitivity analyses conducted to estimate the impact of bias on the observed results?

While registry data present many opportunities for meaningful analysis, there are inherent challenges to making appropriate inferences. Principal concerns include availability of the key variables of interest as well as data quality, since registries and secondary data sources vary in terms of their data quality assurance procedures and information on data quality or curation is not reported consistently. More importantly, nonrandomized studies of all designs are susceptible to systematic errors arising from mismeasurement of analytic variables,[1] unmeasured confounding,[2] and poor choice of reference group.[3,4] These factors must be considered when making inferences based on analyses of registry data.[5]

This chapter explains how analysis plans are constructed for registry data, how they differ depending on the purpose of the analysis, and how registry design and conduct and the characteristic(s) of any linked data can affect analysis and interpretation. The analytic techniques generally used for registry data are presented, addressing how conclusions may be drawn from

the data and what caveats are appropriate. The chapter also describes how timelines for data analysis can be built in at registry inception and how to determine when the registry data are sufficient to begin analysis. Case Examples 23, 24, and 25 provide examples of registry analyses. Chapters 6 and 11 provide more information on the use of secondary data sources in registries.

2. Research Questions and Registry Purposes

Every registry study should start with a research question or focus. For example, disease registries commonly provide descriptive information, such as the typical clinical features of individuals with a disease, variations in phenotype, and the path to diagnosis and clinical progression of the disease over time (i.e., natural history) but they may also compare the effectiveness and safety of various treatments. These registries play a particularly important role in the study of many conditions, and especially for rare diseases.

In the case of studies where the aim is to examine the associations between specific exposures and outcomes, there is a vocal school of thinking that requires prespecification (and registration announcing such prespecification) of the study hypotheses or research questions as a requirement of assuring credible results.[6] While aimed at avoiding publication bias, there are many strong counterarguments.[7-10] The key area of agreement centers around the importance of transparency and making study protocols including analytic plans available for review and publishing results with enough detail to allow replication for confirmation or refutation.

Regardless of whether the hypotheses are prespecified, most productive registry-based research begins with a clear statement of objectives. These objectives might be descriptive or may involve a comparison. Some examples of objectives include:

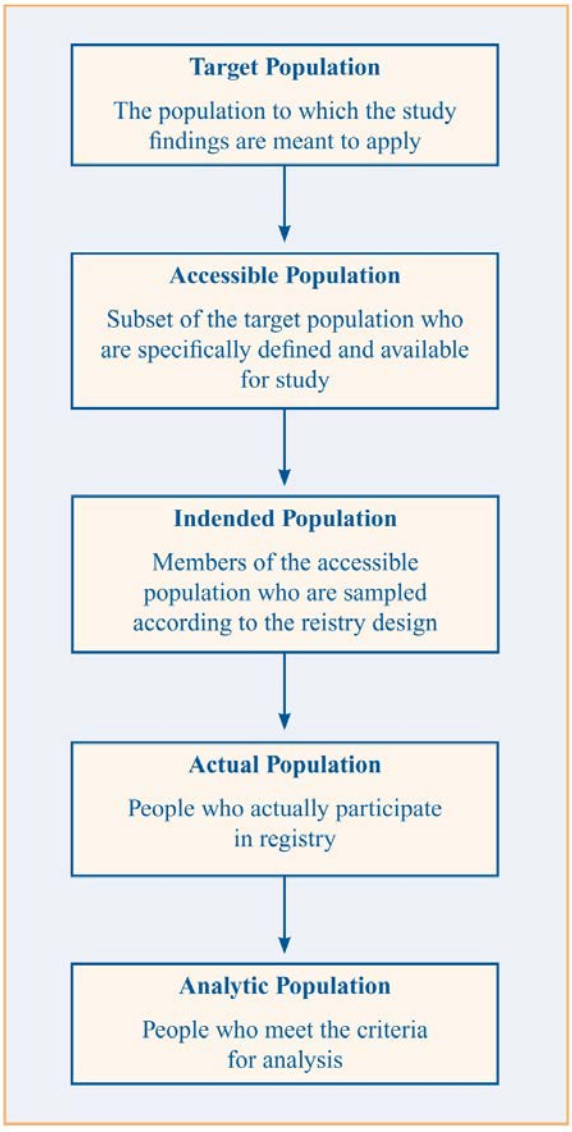
- Measure the incidence of a disease in a specific population,
- Characterize the patterns or costs of treatment for a disease in a specific population
- Measure the occurrence of outcomes among patients with a disease.
- Compare the incidence of a particular disease in two or more subgroups defined by common characteristics (e.g., etiologic research)
- Compare the cost or quality of care for a particular disease in two or more subgroups (e.g., health services research or disparities research)
- Compare the rate of outcomes among two or more subgroups of patients (often defined by different types or levels of treatment) with a particular disease (e.g., clinical research)

In all cases, the overarching objective is to obtain an accurate and valid estimate of the frequency of an outcome's occurrence, or its relative frequency compared across groups.[11] An additional objective may be to generalize study results to a broader population. A valid estimate is one that is likely to be true. A precise estimate is one that has little variability. A generalizable estimate is one that provides information pertinent to the target population, the population for which the study's information provides a basis for potential action, such as a public health or medical intervention. This is discussed further in the following section.

3. Patient Population

The purpose of registry-based research is to provide information about a specific patient population to which all study results are meant to apply. To determine how well the study results apply to the target population, five populations, each of which is a subset of the preceding population, need to be considered, along with how well each population represents the preceding population. These five subpopulations are shown in Figure 13-1.

Figure 13-1. Patient populations



The *target population* is defined by the study's purpose. To assess the appropriateness of the target population, one must ask the question, "Is this really the population that we need to know about?" For example, the target population for a registry of oral contraceptive users would include women of childbearing age who could become pregnant and are seeking to prevent

pregnancy. Studies often miss important segments of the population in an effort to make the study population more homogeneous. For example, a study to assess a medical device used to treat patients for cardiac arrhythmias that defines only men as its target population would be less informative than it could be, because the device is designed for use in both men and women. Studies using linked datasets may miss segments of the population because of limitations in the secondary data source; for example, a study that linked registry data to claims data from a private payer would miss patients covered by other payers (e.g., Medicare) as well as the uninsured.

The *accessible population* is defined using inclusion criteria and exclusion criteria. The inclusion criteria define the population that will be used for the study and generally include geographic (e.g., hospitals or clinics in the New England region), demographic, disease-specific, and temporal (e.g., specification of the included dates of hospital or clinic admission), as well as other criteria. Conversely, the exclusion criteria seek to eliminate specific patients from study and may be driven by an effort to assure an adequate-sized population of interest for analysis. The same goals may be said of inclusion criteria, since it is difficult to separate inclusion from exclusion criteria (e.g., inclusion of adults aged 18 and older vs. exclusion of children younger than 18).

The accessible population may lose representativeness to the extent that convenience plays a part in its determination, because people who are easy to enroll in the registry or capture in secondary data sources may differ in some critical respects from the population at large. Similarly, to the extent that homogeneity plays a part in determining the accessible population, it is less likely to be representative of the entire population because certain population subgroups will be excluded.

Factors to be considered in assessing the accessible population's representativeness of the target population include all the inclusion and exclusion criteria mentioned above. One method of evaluating representativeness is to describe the demographics and other key descriptors of the study population and to contrast its composition with patients with similar characteristics who are identified from another database, such as might be obtained from health insurers, health maintenance organizations, or the U.S. Surveillance Epidemiology and End Results (SEER) cancer registries.[12] For example, the Get With The Guidelines (GWTG)-Stroke registry was linked with Medicare claims data to examine the representativeness of the registry's population of Medicare beneficiaries admitted for ischemic stroke.[13]

However, simple numerical/statistical representativeness is not the main issue. Representativeness should be evaluated in the context of the purpose of the study—that is, whether the study results can reasonably be generalized or extrapolated to other populations of interest outside of those included in the accessible population. For example, suppose that the purpose of the study is to assess the effectiveness of a drug in U.S. residents with diabetes. If the accessible population includes no children, then the study results may not apply to children, since children often metabolize drugs very differently from adults.

On the other hand, consider the possibility that the accessible population is generally drawn from a geographically isolated region, whereas the target population may be the entire country or the world. In that case, the accessible population is not geographically representative of the target population, but that circumstance would have little or no impact on the representativeness of the study findings to the target population if the action of the drug (or its delivery) does not vary

geographically (which we would generally expect to be the case, unless pertinent racial/genetic or dietary factors were involved or if risk factors for the outcome differ geographically). Therefore, in this example, the lack of geographical representativeness would not affect interpretation of results.[14] In common practice, representativeness is interpreted to draw accessible populations that represent typical patients and practitioners, rather than referring to representative samples in the statistical sense of the term.

The reason for using an *intended population* rather than the whole accessible population for the study is simply a matter of convenience and practicality. The issues to consider in assessing how well the intended population represents the accessible population are similar to those for assessing how well the accessible population represents the target population. The main difference is that the intended population may be specified by a sampling scheme, which often tries to strike a balance among representativeness, convenience, and budget. If the study is designed to estimate a rate of occurrence of in a specific population, then a random sample of the accessible population would be considered representative of the accessible population. It is important to note that for many, if not most, registry-based studies, a complete roster of the accessible population does not exist outside of single health system. More commonly, the intended population is compared with the accessible population in terms of pertinent variables.

To the extent that convenience or other design (e.g., stratified random sample) is used to choose the intended population, one must consider the extent to which the sampling of the accessible population may affect any interpretations from the study. For example, suppose that, for the sake of convenience, only patients who attend clinic on Mondays are included in the study. If patients who attend clinic on Mondays are similar in every relevant respect to other patients, that may not constitute a limitation. But if Monday patients are substantially different from patients who attend clinic on other days of the week (e.g., well-baby clinics are held on Mondays) and if those differences affect the outcome that is being studied (e.g., proportion of baby visits for “well babies”), then that sampling strategy would substantially alter the interpretations from the study and would be considered a meaningful limitation.

The extent to which the *actual population* is not representative of the intended population (or typical patients and/or practitioners) is generally a matter of real-world issues that prevent recruitment of a more comprehensive population. It is important to consider the likely underlying factors that caused those subjects not to be included in the analysis of study results and how that might affect the interpretations from the registry. For example, consider a study of a newly introduced medication, such as an anti-inflammatory drug that is thought to be as effective as other products and to have fewer side effects but that is more costly. Inclusion in the actual population may be influenced by prescribing practices governed by a health insurer. For example, if a new drug is approved for reimbursement only for patients who have “failed” treatment with other anti-inflammatory products, the resulting actual population will be systematically different from the target population of potential anti-inflammatory drug users. The actual population may be refractory to treatment or may have more comorbidities (e.g., gastrointestinal problems), and may be specifically selected for treatment beyond the intention of the study-specified inclusion criteria. In fact, registries of newly introduced drugs and devices may often include patients who are different from the ultimate target population of broad interest.

A related issue is that bias that could result from recruitment of *early adopters*,[15] in which practitioners who are quick to use a novel healthcare intervention or therapy differ from those who use it only once it is well established. For example, a study of the use of a new surgical technique may initially enroll largely academic physicians and only much later enroll community-based surgeons. If the outcomes of the technique differ between the academic surgeons (early adopters) and community-based surgeons (later adopters), then the initial results of the study may not reflect the true effectiveness of the technique in widespread use. In fact, "operator experience" is an important factor in understanding the effectiveness of various surgical approaches or of devices that require surgical implantation. Patients selected for treatment with a novel therapy may also differ with regard to factors such as severity or duration of disease and prior treatment history, including treatment failures. For example, patients with more severe or late-stage disease who have failed other treatments might be more likely to use a newly approved product that has shown efficacy in treating their condition. Later on, patients with less severe disease may start using the product.

Finally, the *analytic population* includes all those patients who meet the criteria for analysis. In some cases, it becomes apparent that there are too few cases of a particular type, or too few patients with certain attributes, such that these subgroups do not contribute enough information for meaningful analysis. Analytic populations are also created to meet specific needs. For example, an investigator may request a dataset that will be used to analyze a subset of the registry population, such as those who had a specific treatment or condition.

Patients who are included in the analytic population for a given analysis may also be subject to selection or inclusion criteria (admissibility criteria), and these may affect interpretation of the resulting analyses. For example, if only patients who survive long enough to be admitted to hospital are included in a study, then immortal time bias may bias the results.[16] Another example is when patients who remain enrolled in a registry and attend followup visits through 2 years after registry initiation are included in analysis of adherence to therapy. In this event, it is possible or likely that adherence among those who remain enrolled in the registry will be different from adherence among those who do not. Differential loss to followup, whereby patients who are lost may be more likely to experience adverse outcomes, such as mortality, than those who remain under observation, is a related issue that may lead to biased results. (See Chapter 3.)

Selection of a study population inevitably involves balancing accuracy and generalizability concerns, as well as cost and feasibility considerations. For example, restriction is one of the most effective strategies for control of confounding through study design.[17] If one is concerned about confounding by sex, a simple and effective strategy to control that confounding is to restrict the study population to a single sex. However, such restriction reduces the study's precision by decreasing the sample size, and may also reduce the generalizability of the results (only applicable to a segment of the target population). An alternative would be to include both sexes and to stratify the analysis by sex. While this approach would improve the generalizability of the results and allow for an evaluation of confounding, the precision of the estimated association would be reduced, and perhaps substantially reduced, if the estimate of effect in men was substantially different from the estimate of effect in women. In this circumstance, the study becomes effectively two studies.

4. Data Quality

In addition to a full understanding of study design and methodology, analysis of registry events and outcomes will benefit from an assessment of data quality. When registry data are linked to secondary data sources, this quality assessment must consider both the quality of the registry data as well as the original purpose, inherent limitations, likelihood of differential followup, and quality of any data linked to the registry data; the accuracy of matching the data to specific patients should also be considered. One must examine whether most if not all important covariates were collected, how analytic variables were defined in registry and secondary data sources, data completeness for key variables of interest, how missing data were handled, and data accuracy.

Linked datasets may offer the opportunity to validate some registry data. For example, pharmacy data can be used to confirm that prescriptions were actually filled and may provide more accurate identification of medication use than the data recorded in a registry. The frequency of pharmacy refills is often an indicator of adherence and may be more accurate than patient-reported adherence. Similarly, registries that are derived primarily from patient-reported data may use record linkage for clinical validation of events of special interest or to supplement patient-reported information with clinical or administrative data.

4.1 Collection of Important Covariates

While registries are constructed, they generally serve a particular purpose that drives data collection strategies. However, registry information collected for one purpose (e.g., provider performance feedback) may later be used for another purpose, provided that the terms of data access, including informed consent, allows such additional uses.

For example, suppose the research question addresses the comparative effectiveness of two treatments for a given disease using an existing registry. To be meaningful, the registry should have accurate, well-defined, and complete information, including potential confounding and effect-modifying factors; population characteristics of those with the specified disease; exposures (whether patients received treatment A or B); and patient outcomes of interest. Confounding factors are variables that influence both the exposure (treatment selection) and the outcome in the analyses. These factors can include patient factors (age, gender, race, socioeconomic factors, disease severity, or comorbid illness); provider factors (experience, skills); and system factors (type of care setting, quality of care, or regional effects). It is generally not possible to identify all confounding factors in planning a registry, nor is it possible to collect all confounding factors of interest (e.g., genetic factors, complete family histories, occupational and environmental exposures and socioeconomic factors that may influence disease occurrence or treatment benefits and risks). However, it is desirable to give serious thought to what will be important and how the necessary data can be collected. While effect modification (i.e., when the magnitude of the effect of the primary exposure on an outcome differs depending on the level of a third variable) is not a threat to validity if properly accounted for in the analyses, it is important to consider potential effect modifiers for data collection and analysis to evaluate whether an association varies within specific subgroups.[18] Analysis of registry data requires information about such variables so that the confounding covariates can be accounted for, using one of

several analytic techniques covered in upcoming sections of this chapter. In addition, as described in Chapter 3 and above, eligibility for entry into the study may be restricted to individuals within a certain range of values for potential confounding factors in order to reduce the effects of these factors. Such restrictions may also affect the generalizability of the study.

4.2 Definition of Analytic Variables

Registries typically capture data using data elements with clear, unambiguous definitions that are determined in advance of data collection (see Chapters 3, 4, and 5). It is essential that such information is documented in an accessible manner and made available in the context of analysis files so that users of registry data understand the data definitions. Individual registry data elements may be transformed into composite endpoints for analytic purposes, and definitions of how these variables were created must be documented and made available to those who conduct registry study analyses and included in any reports. When registry data are linked to another data source for a specific study, it is equally important to define the analytic variables in the secondary data source. To enhance transparency and facilitate data curation and study replication, the protocol for the linked study should provide a clear, unambiguous definition of the exposures and outcome being studied, a description of how they will be measured, and a discussion of strengths and limitations of using these variables.

If the study objective is to compare the rates of outcome occurrence across subgroups, then the protocol should provide a definition of the exposure and comparator(s). It is critical that both the index condition (i.e., the “exposed” or “treated” group) and the reference group/condition (e.g., those not exposed to the study treatment of interest, those treated with another method, or those treated any other way). Attention should also be given to identifying and accurately measuring potential confounders and effect modifiers in the primary and secondary data source, as discussed in the prior section.

4.3 Data Completeness and Curation

Assuming that a registry or secondary data source has the necessary data elements, the next step is to characterize data completeness for the key variables needed for the primary objectives. Recognizing that registry-based studies, by definition, are observational in nature, completeness needs to be assessed in that context. For example, patients will present according to the dictates of routine care practices at a facility, tempered by patients’ ability to take time off from work or travel to the site. Thus, patients may not present for followup visits according to what is expected. Similarly, the practice of ordering specific diagnostic or laboratory tests will often vary by physician practice. The variable nature of real-world medical care also may influence how often patients complete surveys and patient-reported outcome (PRO) measures, if these questionnaires are completed when a patient present for care. Even during routine care, patients may miss a visit or decline to undergo a procedure or test, and providers may elect to forego expected tests for a few or a specific subset of their patients. Demographics, test results, and other key information may not be documented in the registry due to lack of availability, refusal to provide, or incorrect documentation (e.g., the values are inconsistent or out-of-range). These scenarios, among other potential issues, result in missing or inconsistent data in the registry databases and secondary data sources, but are to be expected.

Data curation describes the process of reviewing data for completeness and accuracy. For registry-based studies, it may be possible to query certain data elements that are missing or that fall outside of the standard expected range. Chapter 11 provides more information on queries. In addition, when incorporating secondary data into registries, it is essential to evaluate whether the data import performed according to specifications and whether any data transformations were made according to expectations. In some cases, external quality control checks may be useful for making these determinations (e.g., is the patient distribution (by age, ethnicity, etc.) similar to the distribution that would be expected for this population).

Recognizing the types of issues that are inherent in non-interventional research, evaluations of data should focus first on determining whether data are missing largely at random or if there is a systematic bias in the data.[19] For example, when looking at secondary data, a common key question is whether the patient is likely to obtain followup for the conditions of interest in the database. This question also has parallels for primary data collection, since healthcare providers will depend on patient reports for events not treated by that provider, and researchers should consider whether there are likely to be motivations that affect reporting completeness or accuracy(e.g., reporting motor vehicle accidents while impaired.)

4.4 Data Accuracy and Validation

While observational registry studies are usually not required to meet U.S. Food and Drug Administration and International Conference on Harmonisation standards of Good Clinical Practice developed for clinical trials, sponsors and contract research organizations that conduct registry studies are responsible for ensuring the accuracy of study data to the extent possible. Plans for data quality assurance, data verification, and site monitoring (if any) should be developed at the beginning of a study and adhered to throughout its lifespan. Chapter 11 discusses in detail approaches to data collection and quality assurance, including data management, site monitoring, and source data verification.

Ensuring the accuracy and validity of data and programming at the analysis stage also requires consideration. The Office of Surveillance and Epidemiology (OSE) of the Food and Drug Administration's Center for Drug Evaluation and Research uses the manual *Standards of Data Management and Analytic Process in the Office of Surveillance and Epidemiology* for database analyses conducted within OSE; the manual addresses many of these issues and may be consulted for further elaboration on these topics.[20] Topics addressed that pertain to ensuring the accuracy of data just before and during analysis include developing a clear understanding of the data at the structural level of the database and variable attributes. Creating analytic programs with careful documentation and an approach to variable creation and naming conventions that is straightforward and, when possible, consistent with the Clinical Data Interchange Standards Consortium (CDISC) initiative; are useful whether primary or secondary data are used. Similarly, some verification of programming and analytic dataset creation by a second analyst is also considered good practice.

4.5 Validation Substudies[i]

Validation substudies may be used to evaluate the accuracy of certain data elements or study assumptions and can inform estimates of the potential impact of bias on study results.[2,21] Registry-based research is often amenable to collection of internal validation data, for example by medical record review. In addition, many databases have internal protocols that constantly validate at least some aspects of the data. The validation data generated by these protocols, if accessible, may provide an initial indication of the data quality. To facilitate data collection for study-specific internal validation studies, investigators should consider the important threats to the validity of their research while designing their study, and should allocate project resources accordingly.

For example, in the study of statin use related to ALS and neurodegenerative diseases described above,[22] the ICD-10 code used to identify cases (G12.2) corresponded to diagnoses of ALS or other motor neuron syndromes. The investigators therefore selected a random sample of 25 individuals from among all those who satisfied the case definition, and a clinician investigator reviewed their discharge summaries. The proportion of these 25 who did not have ALS (32 percent) was used to inform a bias analysis to model the impact of these false-positive ALS diagnoses. Assuming a valid bias model, the bias analysis results showed that the null association was unlikely to result from the nondifferential misclassification of other diseases as ALS.

In this example, there was no effort to validate that non-cases of ALS were truly free of the disease. Non-cases are seldom validated, because false-negative cases, especially of rare diseases, occur very rarely. Furthermore, validating the absence of disease often requires a study-supported medical examination of the non-case patients, an expensive, time-consuming, and invasive procedure. Prevalent diseases with a lengthy preclinical period and relatively simple diagnostic tests, such as diabetes, are more amenable to validation of non-cases. The ALS example also illustrates that an internal validation study requires protocol planning and allocation of study resources to collect the validation data. A protocol should be written that specifies how participants in the validation sample will be selected from the study population. Participation in the validation substudy might require informed consent to allow medical record review, whereas the database data itself might be available without individual informed consent. These aspects should be resolved in the planning stage, and the analytic plan should include a section devoted to bias modeling and analysis.[2]

[i] *Adapted from Lash TL, Sorensen HT, Bradbury BD, et al. 'Analysis of Linked Registry Datasets Registries.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

4.6 Other Data Quality Issues Relevant to Linked Datasets[ii]

4.6.1 Changes in Coding Conventions Over Time

A common problem with secondary data sources is the impact of changes in coding conventions over the lifetime of the database. These changes can take the form of diagnostic drift,[23] changes in discharge coding schemes, changes in the definition of grading of disease severity, or even variations in the medications on formulary in one region but not others at different points in time. For example, the Danish National Registry of Patients (DNRP) is a database of patient contacts at Danish hospitals. From 1977 to 1993, discharge diagnoses were coded according to ICD-8, and from 1994 forward discharge diagnoses were coded according to ICD-10. ICD-10 included a specific code for chronic obstructive pulmonary disease (COPD, J44), whereas ICD-8 did not [ICD-8 496 (COPD not otherwise specified) did not appear in the DNRP]. In addition, from 1977 to 1994 the DNRP registered discharge diagnoses for only inpatient admissions, but from 1995 forward discharge diagnoses from outpatient admissions and emergency room contacts were also registered. COPD patients seen in outpatient settings before 1995 were therefore not registered; this excluded patients who likely had less severe COPD on average. The change in ICD coding convention in 1994 and the exclusion of outpatient admissions before 1995 presented a barrier to estimating the time trend for incidence of all admissions for COPD in any period that overlapped these two changes to the DNRP.[24]

The General Practice Research Database (GPRD) was a medical records database capturing information on approximately 5 percent of patients in the United Kingdom[25] (as of March 2012, the GPRD became the Clinical Practice Research Database). Information was directly entered into the database by general practitioners trained in standardized data entry. When the GPRD was initiated in 1987, diagnoses were recorded using Oxford Medical Information Systems (OXMIS) codes, which were similar to ICD-9 codes. In 1995, the GPRD adopted the Read coding system, a more detailed and comprehensive system that groups and defines illnesses using a hierarchical system. Without knowledge of this shift in coding and how to align codes for specific conditions across the different coding schemes, studies using multiple years of data could produce spurious findings.

4.6.2 Precision Considerations When Standard Errors Are Small

The large size of the study population that can often be included in a registry-based study is a strength, but it also requires special attention. The sample size allows adjustment for multiple potential confounders with little potential for over-fitting or sparse data bias,[26] and allows design features such as comparisons of different treatments for the same indication (comparative effectiveness research) to reduce the potential for confounding by indication.[27] Nonetheless, systematic errors remain a possibility, and these systematic errors dominate the uncertainty when estimates of association are measured with high precision as a consequence of a large sample

[ii] *Adapted from Lash TL, Sorensen HT, Bradbury BD, et al. 'Analysis of Linked Registry Datasets.' In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.*

size.[28] When confidence intervals are narrow, systematic errors remain, and/or inference or policy action will potentially result, investigators have been encouraged to employ quantitative bias analysis to more fully characterize the total uncertainty.[21] Bias analysis methods have been used to address unmeasured confounding,[29] selection bias,[30] and information bias[29,31] in registry-based research.

5. Data Analysis

Statistical methods commonly used for descriptive purposes include those that summarize information from continuous variables (e.g., mean, median) or from categorical variables (e.g., proportions, rates). Registries may describe a population using incidence (the proportion of the population that develops the condition over a specified time interval) and prevalence (the proportion of the population that has the condition at a specific point in time). Another summary estimate that is often used is an incidence rate. The incidence rate (also known as absolute risk) takes into account both the number of people in a population who develop the outcome of interest and the person-time at risk, or the length of time contributed by all people during the period when they were in the population and the events were counted.

For analytical studies, the association between a risk factor and outcome may be expressed as relative risk, odds ratio, or hazard ratio, depending on the nature of the data collected, the duration of the study, and the frequency of the outcome. The standard textbooks cited here have detailed discussions regarding epidemiologic and statistical methods commonly used for the various analyses supported by registries.[18,32-35]

It is always important to consider the role of confounding. Although those planning a study try to collect as much data as possible to address known confounders, there is always the chance that unknown or unmeasured confounders will affect the interpretation of analyses derived from observational studies. It is important to consider the extent to which bias (systematic error stemming from factors that are related to both the decision to treat and the outcomes of interest [confounders]) could have distorted the results. For example, a bias known as *confounding by indication*[36] results from the fact that physicians do not prescribe medicine at random: the reason a patient is put on a particular regimen is often associated with their underlying disease severity and may, in turn, affect treatment outcome. To detect such a bias, the distribution of various prognostic factors at baseline is compared for patients who receive a treatment of interest and those who do not. A related concept is *channeling bias*, in which drugs with similar therapeutic indications are prescribed to groups of patients who may differ with regard to factors influencing prognosis.[37] Other types of bias include detection bias[38] (e.g., when comparison groups are assessed at different points in time or by different methods), selective loss to followup in which patients with the outcomes of most interest (e.g., sickest) may differentially drop out of one treatment group than another, and performance bias (e.g., systematic differences in care other than the intervention under study, such as a public health initiative promoting healthy lifestyles directed at patients who receive a particular class of treatment). In addition to such biases, analyses need to account for effect modification.[39] The presence of effect modification may also be identified after the data are collected.

Confounding may be evaluated using stratified analysis, multivariable analysis, sensitivity analyses, and simple or quantitative bias analysis or simply by graphical comparison of characteristics and events between groups.[40] The extensive information and large sample sizes available in some registries also support use of more advanced modeling techniques for addressing confounding by indication, such as the use of propensity scores to create matched comparison groups, or for stratification or inclusion in multivariable risk modeling.[41-44] New methods also include the high-dimensional propensity score (hd-PS) for adjustment using administrative data.[45] Examples are too numerous for a few selections to be fully representative, but registries in nearly every therapeutic area, including cancer,[46] cardiac devices,[47] organ transplantation,[48] and rare diseases,[49] have published the results of analyses incorporating approaches based on propensity scores. Nonetheless, application of these and other methods may not fully control for unmeasured confounding.[50]

Groupings within a study population, such as patients seen by a single clinician or practice, residents of a neighborhood, or other "clusters," may themselves impact or predict health outcomes of interest. Such groupings may be accounted for in analysis through use of analytic methods including analysis of variance (ANOVA), and hierarchical or multilevel modeling.[51-54]

Heterogeneity of treatment effect is also an important consideration for comparative effectiveness research as the effect of a treatment may vary within subgroups of heterogeneous patients.[55] Stratification on the propensity score has been used to identify heterogeneity of treatment effect and may identify clinically meaningful differences between subgroups based on pre-treatment characteristics.

For economic analyses, the analytic approaches often encountered are cost-effectiveness analyses and cost-utility studies. To examine cost-effectiveness, costs are compared with clinical outcomes measured in units such as life expectancy or years of disease avoided.[56] Cost-utility analysis, a closely related technique, compares costs with outcomes adjusted for quality of life (utility) using measures known as quality-adjusted life years. Since most new interventions are more effective but also more expensive, another analytic approach examines the incremental cost-effectiveness ratio and contrasts that to the willingness to pay. (Willingness-to-pay analyses are generally conducted on a country-by-country basis, since various factors relating to national health insurance practices and cultural issues affect willingness to pay.) The use of registries for cost-effectiveness evaluations is a fairly recent development, and consequently, the methods are evolving rapidly. More information about economic analyses can be found in standard textbooks.[57-62]

A number of biostatistics and epidemiology textbooks cover in depth the issues raised in this section and the appropriate analytic approaches for addressing them—for example, "time-to-event" or survival analyses[63] and issues of recurrent outcomes and repeated measures, with or without missing data,[64] in longitudinal cohort studies. Other texts address a range of regression and nonregression approaches to analysis of case-control and cohort study designs[65] that may be applied to registries. For further information on how to quantify bias, please see Lash, Fox, and Fink.[2]

5.1 Factors To Consider in the Analysis

Registry results are most interpretable when they are specific to well-defined endpoints or outcomes in a specific patient population with a specific treatment status. Registry analyses may be more meaningful if variations of study results across patient groups, treatment methods, or subgroups of endpoints are reported. In other words, analysis of a registry should explicitly provide the following information:

- *Patient*: What are the characteristics of the patient population in terms of demographics, such as age, gender, race/ethnicity, insurance status, and clinical and treatment characteristics (e.g., history of significant medical conditions, disease status at baseline, and prior treatment history)?
- *Exposure (or treatment)*: Exposure could be therapeutic treatment such as medication or surgery; a diagnostic or screening tool; behavioral factors such as alcohol, smoking habits, and diet; or other factors such as genetic predisposition or environmental factors. What are the distributions of the exposure in the population? Is the study objective specific to any one form of treatment? Is a new user design being used?[66] Does the exposure definition (index and reference group) and analysis avoid immortal-time bias?[67] Are there repeated measures at predetermined intervals or is the exposure intermittent?
- *Endpoints (or outcomes)*: Outcomes of interest may encompass effectiveness or comparative effectiveness, the benefits of a healthcare intervention under real-world circumstances,[68] and safety—the risks or harms that may be associated with an intervention. Examples of effectiveness outcomes include survival, disease recurrence, symptom severity, quality of life, and cost-effectiveness. Safety outcomes may include infection, sensitivity reactions, cancer, organ rejection, and mortality. Endpoints must be precisely defined at the data collection and analysis stages. Are the study data on all-cause mortality or cause-specific mortality? Is information available on pathogen-specific infection (e.g., bacterial vs. viral)? Are there competing risks?[69]
- *Covariates*: As with all observational studies, comparative effectiveness research requires careful consideration, collection, and analysis of important confounding and effect modifying variables. For medication exposures, are dose, duration, and calendar time under consideration? Directed acyclic graphs (DAGs) can be useful tools to illustrate how the exposure (or treatment), outcome and covariates are related.[70,71]
- *Time*: For valid analysis of risk or benefit that occurs over a period of time following therapy, detailed accounting for time factors is required. For exposures, dates of starting and stopping a treatment or switching therapies should be recorded. For outcomes, the dates when followup visits occur, and whether or not they lead to a diagnosis of an outcome of interest, are required in order to take into account how long and how frequently patients were followed. Dates of diagnosis of outcomes of interest, or dates when patients complete a screening tool or survey, should be recorded. At the analysis stage, results must also be described in a time-appropriate fashion. For example, is an observed risk consistent over time (in relation to initiation of treatment) in a long-term study? If not, what time-related risk measures should be reported in addition to or instead of cumulative risk? When exposure status changes frequently, what is the method of

capturing the population at risk? Many observational studies of intermittent exposures (e.g., use of nonsteroidal anti-inflammatory drugs or pain medications) use time windows of analysis, looking at events following first use of a drug after a prescribed interval (e.g., 2 weeks) without drug use. Different analytic approaches may be required to address issues of patients enrolling in a registry at different times and/or having different lengths of observation during the study period.

- *Potential for bias*: Successful analysis of observational studies also depends to a large extent on the ability to measure and analytically address the potential for bias. Refer to Chapter 3 for a description of potential sources of bias. Directed acyclic graphs can also be useful for understanding and identifying the source of bias.[70,71] For details on how to quantify potential bias, see the textbook by Lash, Fox, and Fink.[2]

The choice of comparators is also a challenging issue. When participants in a cohort are classified into two or more groups according to certain study characteristics (such as treatment status, with the "standard of care" group as the comparator), the registry is said to have an *internal* or *concurrent* comparator. The advantage of an internal comparator design is that patients are likely to be more similar to each other, except for their treatment status, than patients in comparisons between registry subjects and external groups of subjects. When defining the comparator group, it is important not to introduce immortal time bias.[67] In addition, consistency in measurement of specific variables and in data collection methods make the comparison more valid. Internal comparators are particularly useful for treatment practices that change over time. Comparative effectiveness studies may often necessitate use of an internal comparator in order to maximize the comparability of patients receiving different treatments within a given study, and to ensure that variables required for multivariable analysis are available and measured in an equivalent manner for all patients to be analyzed.

Unfortunately, it is not always possible to have or sustain a valid internal comparator. For example, there may be significant medical differences between patients who receive a particularly effective therapy and those who do not (e.g., underlying disease severity or contraindications), or it may not be feasible to maintain a long-term cohort of patients who are not treated with such a medication. It is known that external information about treatment practices (such as scientific publications or presentations) can result in physicians changing their practice, such that they no longer prescribe the previously accepted standard of care. There may be a systematic difference between physicians who are early adopters and those who start using the drug or device after its effectiveness has been more widely accepted. Early adopters may also share other practices that differentiate them from their later-adopting colleagues.[15]

In the absence of a good internal comparator, one may have to leverage external comparators to provide critical context to help interpret data revealed by a registry. An external or historical comparison may involve another study or another database that has disease or treatment characteristics similar to those of registry subjects. Such data may be viewed as a context for anticipating the rate of an event. One widely used comparator is the U.S. SEER cancer registry data, because SEER provides detailed annual incidence rates of cancer stratified by cancer site, age group, gender, and tumor staging at diagnosis. SEER represents 28 percent of the U.S. population.[12] A procedure for formalizing comparisons with external data is known

as *standardized incidence rate or ratio*;[39] when used appropriately, it can be interpreted as a proxy measure of risk or relative risk.

Use of an external comparator, however, may present significant challenges. For example, SEER and a given registry population may differ from each other for a number of reasons. The SEER data cover the general population and have no exclusion criteria pertaining to history of smoking or cancer screening, for example. On the other hand, a given registry may consist of patients who have an inherently different risk of cancer than the general population, resulting from the registry's having excluded smokers and others known to be at high risk of developing a particular cancer. Such a registry would be expected to have a lower overall incidence rate of cancer, which, if SEER incidence rates are used as a comparator, may complicate or confound assessments of the impact of treatment on cancer incidence in the registry.

However, use of external comparators is becoming an important tool for regulators, and external comparators for phase II clinical trials may come from registries.[19]

Regardless of the choice of comparator, similarity between the groups under comparison should not be assumed without careful examination of the study patients. Different comparator groups may result in very different inferences for safety and effectiveness evaluations; therefore, analysis of registry findings using different comparator groups may be used in sensitivity analyses or bias analyses to determine the robustness of a registry's findings.

5.2 Developing a Statistical Analysis Plan

5.2.1 Need for a Statistical Analysis Plan

It is good practice develop a statistical analysis plan (SAP) that describes the analytical principles and statistical techniques to be employed to address the primary and secondary objectives, as specified in the study protocol or plan, before embarking on data analysis. A registry may require a primary "master SAP" as well as subsequent, supplemental SAPs. Supplemental SAPs might be triggered by new research questions emerging after the initial master SAP was developed or might be needed because the registry has evolved over time (e.g., additional data collected, data elements revised).

Although the evolving nature of data collection practices in some registries poses challenges for data analysis and interpretation, it is important to keep in mind that the ability to answer questions emerging during the course of the study is one of the advantages (and challenges) of a registry. In the specific case of long-term rare-disease registries, many of the relevant research questions of interest cannot be defined *a priori* but arise over time as disease knowledge and treatment experience accrue. Supplemental SAPs can be developed only when enough data become available to analyze a particular research question. At times, the method of statistical analysis may have to be modified to accommodate the amount and quality of data available.

To the extent that the research question and SAP are formulated *before* the data analyses are conducted and results are used to answer specific questions or hypotheses, such supplemental analysis retains much of the intent of prespecification rather than being wide-ranging exploratory analyses (sometimes referred to as "fishing expeditions"). The key to success is to provide

sufficient details in the SAP that, together with the study protocol and the case report forms, the overall process of the data analysis and reporting are well described.

5.2.2 Preliminary Descriptive Analysis To Assist SAP Development

During SAP development, one aspect of a registry that is somewhat different from a randomized controlled trial is the necessity to understand the distribution (sometimes referred to as the "shape") of the data collected to inform subsequent stratified analyses.[39] This may be crucial for a number of reasons.

Given the broad inclusion criteria that most registries tend to propose, there might be a wide distribution of patients, treatment, and/or outcome characteristics. The distribution of age, for example, may help to determine if more detailed analyses should be conducted in the "oldest old" age group (80 years and older) to help understand health outcomes in this subgroup that might be different from those of their younger counterparts.

Unless a registry is designed to limit data collection to a fixed number of regimens, the study population may experience many treatments, considering the possible combination of various dose levels, drug names, frequency and timing of medication use (e.g., acute, chronic, intermittent), and sequencing of therapies. The scope and complexity of these variations often constitute one of the most challenging aspects of analyzing a registry. Grouping of treatment into regimens for analysis should be done carefully, guided by clinical experts in that therapeutic area as well as by study purpose. The full picture of treatment patterns may become clear only after a sizable number of patients have been enrolled. Consequently, the treatment definition in an SAP may be refined during the course of a study. Furthermore, there may be occasions where a use of a therapeutic regimen of interest is much less frequent than anticipated, so that specific study objectives focusing on this group of patients might become unfeasible.

5.3 Timing of Analyses During the Study

Unlike a typical clinical trial, registries, especially those that take several years to complete, may conduct intermediate analyses before all patients have been enrolled and/or all data collection has been completed. Such midcourse analyses may be undertaken for several reasons. First, many of these registries focus on serious safety outcomes. For such safety studies, it is important for all parties involved to actively monitor the frequency of such events at regular predefined intervals so that further risk assessment or risk management can be considered. The timing of such analyses may be influenced by regulatory requirements. Second, it may be of interest to examine treatment practices or health outcomes during the study to capture any emerging trends. Finally, it may also be important to provide intermediate or periodic analysis to document progress, often as a requirement for continued funding.

While it is useful to conduct such periodic analysis, careful planning should be given to the process and timing and whether the need for such analyses are programmatic (e.g., for annual progress reports) or for scientific purposes. For scientific purposes, among the first questions generally asked is whether a sufficient number of patients have been enrolled, sufficient followup time has elapsed to observe events of interest and/or whether a sufficient number of

events have occurred. For example, some events, such as site reactions to injections, can be observed after a relatively short duration, compared with events like cancers, which may have a long induction or latency. If there are too few patients or insufficient time has elapsed, premature analyses may lead to the unreliable conclusions. However, it is inappropriate to delay analysis so long that an opportunity might be missed to observe emerging safety outcomes. Investigators should use sound clinical and epidemiological judgment when planning interim or periodic analyses and be sure that any resultant reports address both the strengths and limitations of such analyses.

5.3.1 Patient Censoring

At the time of a registry analysis, events may not have occurred for all patients. For these patients, the data are said to be *censored*, indicating that the observation period of the registry was stopped before all events occurred (e.g., mortality). In these situations, it is unclear when the event will occur, if at all. In addition, a registry may enroll patients until a set stop date, and patients entered into the registry earlier will have a greater probability of having an event than those entered more recently because of the longer followup. An important assumption, and one that needs to be assessed in a registry, is how patient prognosis varies with the time of entrance into the registry. This issue may be particularly problematic in registries that assess innovative (and changing) therapies. Patients and outcomes initially observed in the registry may differ from patients and outcomes observed later in the registry timeframe, either because of true differences in treatment options available at different points in time, or because of the shorter followup for people who entered later. Patients with censored data, however, contribute important information to the registry analysis. When possible, analyses should be planned so as to include all subjects, including those censored before the end of the followup period or the occurrence of an event. One method of analyzing censored data to estimate the conditional probability of the event occurring is to use the Kaplan-Meier method.[72] In this method, for each time period, the probability is calculated that those who have not experienced an event before the beginning of the period will still not have experienced it by the end of the period. The probability of an event occurring at any given time is then calculated from the product of the conditional probabilities of each time interval.

5.3.2 Sensitivity Analyses

Sensitivity analysis refers to a procedure used to determine how robust the study result is to alterations of various parameters. If a small parameter alteration leads to a relatively large change in the results, the results are said to be sensitive to that parameter. Sensitivity and bias analyses may be used to determine how the final study results might change when taking into account those lost to followup. A hypothetical simple sensitivity analysis is presented in Table 13-1.

Table 13-1. Impact of loss to followup on incidence rates per 1,000 in a hypothetical study of 1,000 patients in a registry

Assuming that the incidence rate among patients lost to followup is X times the rate of incidence estimated in those who stayed in the registry:

Various Assumptions of the Observed Incidence Rate	Assuming a 10% Loss to Followup	Assuming a 30% Loss to Followup
X=0.5	106	94
X=1	111	110
X=2	122	143
X=3	156	242

Table 13-1 illustrates the extent of change in the incidence rate of a hypothetical outcome assuming varying degrees of loss to followup, and differences in incidence between those for whom there is information and those for whom there is no information due to loss to followup. In the first example, where 10 percent of the patients are lost to followup, the estimated incidence rate of 111/1,000 people is reasonably stable; it does not change too much when the (unknown) incidence in those lost to followup changes from 0.5 times the observed to 3 times the observed, with the corresponding incidence rate that would have been observed ranging from 106 to 156 per 1,000. On the other hand, when the loss to followup increases to 30 percent, the corresponding incidence rates that would have been observed range from 94 to 242. This procedure could be extended to a study that has more than one cohort of patients, with one being exposed and the other being nonexposed. In that case, the impact of loss to followup on the relative risk could be estimated by using sensitivity analysis.

5.4 Missing Data[iii]

The intent of any analysis is to make valid inferences from the data. Missing data can threaten this goal both by reducing the information yield of the study and, in many cases, by introducing bias. Understanding the types of and reasons for missing data can help guide the selection of the most appropriate analytical strategy for handling the missing data, or the potential bias that may be introduced by such missing data. Typically, missing data result from item nonresponse, left truncation, and right censoring. The concepts below apply to both registry data and secondary data sources with an exception that data may be missing in secondary data sources simply because those data elements are not intended to be collected. For example, patient's race is not reported in administrative claims data.

[iii] *Adapted from Mack C, Su Z, Westreich D. Managing Missing Patient Data in Patient Registries. White Paper, addendum to Registries for Evaluating Patient Outcomes: A User's Guide, Third Edition. (Prepared by L&M Policy Research, LLC, under Contract No. 290-2014-00004-C.) AHRQ Publication No. 17(18)-EHC015-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2018.*

5.4.1 Reasons for Missing Data

5.4.1.1 Item Nonresponse

Item nonresponse, which occurs when a participant completes a case report form (CRF) or survey without providing a response for one or more of the data elements, may be the most common reason for missing data. As discussed in Chapter 11, CRFs may incorporate checks to ensure that complete, valid data are entered. These checks may prevent CRFs from being marked as complete if data are missing. However, item nonresponse may still occur, either because CRFs are not marked as complete or because some data elements are not considered critical to the study objectives. For example, a recent analysis of the characteristics of missing data in three patient registries found that 71 percent of patients in one registry were missing data for body mass index (BMI), an optional field.[73] Item nonresponse also occurs when patients complete PROs using paper forms and leave some fields blank or enter illegible data.

5.4.1.2 Threats From the Left: Truncation

The issue of left truncation, a form of selection bias, arises when events of interest occur prior to a patient's enrollment in the registry and (typically) pre-empt enrollment in the registry. Applebaum et al. define left truncation as occurring "when subjects who otherwise meet entry criteria do not remain observable for a later start of followup."[74] For example, in a study of miscarriage which enrolls pregnant women, some patients will be left truncated because "an unknown proportion of the source population experiences losses prior to enrollment."[75] Thus, left truncation results in data missing in the observed cohort due to non-enrollment, leading the study sample to not accurately reflect the underlying target population, in this example, pregnant women at risk for miscarriage.

A related bias can be introduced due to entry of already-exposed individuals into a registry or other data source. Consider, for example, a registry designed to study disease progression over several years in patients with a rare disease. Ideally, the registry would enroll only patients at the time of diagnosis, with the goal of collecting detailed baseline and diagnostic information for all patients. However, limiting the registry enrollment to only those newly diagnosed patients would reduce the sample size significantly, and, in the case of a rare disease, likely render the registry infeasible. To enroll sufficient patients, the registry may include both existing (prevalent) patients and newly diagnosed (incident) patients. This enrollment strategy, while practical, has the potential to introduce significant bias for numerous reasons, including under-ascertainment of early events. Examples of the latter include venous thromboembolism risk in women taking third generation over-the-counter drugs relative to earlier products, falls after initiating benzodiazepines, and nonsteroidal anti-inflammatory drugs (NSAIDs) and peptic ulcers.[66]

The concept of "baseline" will be different for patients who are newly diagnosed versus those with an existing diagnosis at the time of enrollment, and comparisons of symptoms, treatment effectiveness, and disease progression would need to account for these differences. In particular, the patients with existing diagnoses may be missing information on symptoms at diagnosis or other tests or procedures related to their diagnosis that occurred prior to study enrollment.[76] Ray gives an overview of this issue in the context of medication effects, suggesting that focusing on

new users (or newly exposed people, generally) is a strategy which can minimize bias, and should be considered whenever logistically feasible.[66]

5.4.1.3 Threats From the Right: Loss to Followup, Censoring, Competing Risks

Loss to followup and right censoring occur when information is missing during the study period or at the conclusion rather than the inception of the registry. In studies that collect long-term followup data, participants may be lost to followup if they formally withdraw from the registry or simply stop completing surveys or coming for scheduled visits. Attrition of this nature occurs for many reasons, including factors both related to the study objectives (e.g., the participant becomes too ill to complete study visits) and unrelated (e.g., the participant moves or changes his/her email address without notifying study staff). Broadly speaking, if the attrition is associated with the study outcomes, it introduces a form of selection bias into the registry that must be described and accounted for in analyses to the extent possible (known as informative censoring in the context of randomized clinical trials).[77] Whether it introduces bias or not, loss to followup can limit the ability of the registry to examine long-term outcomes and can have an impact on statistical power. Registries that aim to collect long-term followup data are encouraged to develop retention targets, actively monitor retention against those targets, and take proactive measure to minimize loss to followup, as needed. Strategies to retain participants and minimize loss to followup are discussed extensively in Chapters 3, 10, and 11.

A related concept to loss-to-followup is administrative right censoring, which occurs when the registry ends before an outcome of interest occurs for all subjects (which is typically the case). This is especially common in pregnancy registries, which are designed to assess outcomes of pregnancies during which the mother (or, in some cases, the father) was exposed to medical products. Pregnancy registries typically collect information on congenital defects that are ascertained at birth or shortly after birth (e.g., 30-day followup or, often at most, one year), but are not designed to detect defects or developmental delays that are diagnosed later in life.[78] Right censoring occurs in other types of registries as well. For example, a registry designed to study the effectiveness of a cancer treatment may conduct survival analyses after following patients for five years. Some patients will have died during that period, and their survival after treatment will be known. However, for patients who are still alive at the conclusion of the study, survival after treatment will be right censored due to the close of the registry. In general, missing data due to administrative right censoring will not introduce bias in analysis, but bias is possible if there are strong temporal trends in risk of the outcome.

Finally, competing risks must be considered. A competing risk is an event that prevents the outcome or outcomes of interest not merely from being observed, but from happening in the first place. For example, in a study of incidence of heart attack, death (by any cause besides heart attack) prevents incident heart attack from occurring; in a study of breast cancer, preventive double mastectomy likewise may be considered a competing risk for breast cancer. Competing risks can lead to missing data in certain settings; sometimes a study may be interested in the risk of breast cancer in all individuals – including those who, due to beliefs about their personal risks breast cancer status that these women would have had, had they not gotten a mastectomy, can be regarded as a variety of missing data; in other cases, competing risks do not lead to such clear

instances of missing data. See Lau et al. for a more involved discussion of competing risks and missing data, as well as analytic approaches.[79]

5.4.1.4 Data Gaps in Secondary Data Sources

Item nonresponse, left truncation, and right censoring are specific examples of the more general problem of data gaps in secondary data sources. While registries collect data continuously, secondary data sources may only pertain to a particular subgroup of a larger population, and membership in that subgroup may be dynamic. Examples include individuals covered by Medicaid and members enrolled in managed care plans. In both examples, the databases pertain to participants in a health insurance program, and membership in those programs can change frequently. Data are collected only while the participants are members. If membership is lost and restored again later, there will be a data gap. Importantly, membership in these plans might be related to other characteristics that affect health, such as socioeconomic status or employment.[80] Similar problems can arise when there are gaps in residency and the database is based on national healthcare data, or when individuals have health insurance from more than one source.

Data gaps in secondary data sources can also arise when medications are dispensed in the hospital, since many databases do not capture in-hospital medication use, leading to a form of information bias. In drug safety studies examining mortality risk related to the use of a particular medication, missing in-hospital medication use can result in spurious estimates of treatment effects.[81] This bias was illustrated in a case-control study examining mortality risk related to inhaled corticosteroid use from the Saskatchewan, Canada, database. Analyses that failed to account for missed corticosteroid use during hospitalization events preceding death or the matched date for controls showed a beneficial effect (RR=0.6; 95% CI, 0.5 to 0.73). The RR estimates changed markedly once the missing in-hospital corticosteroid use was included (RR=0.93; 95% CI, 0.76 to 1.14 and RR=1.35; 95% CI, 1.14 to 1.60).[81] This bias has also been observed in studies of injectable medications in dialysis patients where hospitalization events preceding death resulted in spuriously low effect estimates.[82]

5.4.2 Analytic Implications and Management Strategies for Missing Data

When considering the potential impact of the missing data on the study findings, it is may be helpful to consider whether the data are missing largely at random or due to systematic reasons (e.g., followup is provided outside the health system or by other providers). The type, pattern and amount of missing data will help guide the selection of appropriate management strategies for an analysis.

5.4.2.1 Complete Case Strategy

The *complete case* strategy limits the analysis to patients with complete information for all variables. A simple deletion of all incomplete observations, however, is not appropriate or efficient in all circumstances, and it introduces significant bias if the deleted cases are substantively different from the retained, complete cases (i.e., not missing completely at random). Therefore, the complete case strategy is inefficient and not generally used. For example, patients with diabetes who were hospitalized because of inadequate glucose control

might not return for a scheduled followup visit at which HbA1c was to be measured. Those missing values for HbA1c would probably differ from the measured values because of the reason for which they were missing. Similarly, the availability of the results of certain tests or measurements may depend on what is covered by patients' health insurance (a known value), since registries do not typically pay for testing. Patients without this particular measurement may still contribute meaningfully to the analysis. In order to include patients with missing data, one of several *imputation* techniques may be used to estimate the missing data.

5.4.2.2 Single and Multiple Imputation

Unlike complete case analysis, patients with missing data are retained in the analysis when imputation methods are used. Imputation methods replace missing observations with values predicted in some manner, often from a model.

Single imputation can be useful when dates are partially missing. For example, if the day element of an adverse event (AE) date is missing and is not retrievable, consideration can be given to imputation of the missing day element on the middle (e.g., 15th) day of the month. This date would need to be constrained by underlying study issues (for example, the date must be after the study enrollment date, and study discontinuation date). If treatment changes during this month, the relationship with the timing of the treatment change should be considered when discussing the appropriate imputation method. This issue is closely related to that of interval censoring.

In general, however, use of multiple imputation methods is strongly preferred to single imputation. In multiple imputation, multiple datasets are produced with different values imputed for each missing variable per dataset, thus reflecting the uncertainty around the true values of the missing variables. The multiple values may be derived from the posterior probability distributions for the missing values.[83] As noted, the result of a multiple imputation process is multiple complete datasets for analysis from which a single summary finding is estimated. Standard errors are obtained through a combination of the between-model variance and the within-model standard errors, using Rubin's Rules for Imputation.[84] If data are missing at random (MAR), multiple imputation will generally produce unbiased results if the model includes the correct set of covariates, and unlike single imputation will propagate error correctly. In the presence of data that are missing systematically, multiple imputation in general cannot fully correct any bias due to missing data.

5.4.2.3 Maximum Likelihood Methods

Maximum likelihood estimation (MLE) is an analytic maximization procedure which provides the values of the model parameters that maximize the sample likelihood, i.e., the values that make the observed data "most probable." MLE has the advantage of using all available data and does not require data to be sorted by a fixed number of study visits. Under the assumption of that data are missing at random, MLE is efficient and provides unbiased estimates. Calculating MLE's and fitting them into regression models for statistical inference often requires specialized software, especially when data are missing for predictor variables. This is still a challenge today, but as time goes by, more statistical packages are upgrading to contain MLE analysis capability.

When data are missing for dependent variables only, likelihood-based methods including the well-known mixed models for repeated measurements can be used for analyzing data with monotone or non- monotone missingness patterns.[85]

In observational research, not all studies have meaningful "visits," or have data collected at a given visit for all subjects. For example, it is quite common that a subject can have unscheduled visits, and a lab test can be "missing" simply because the test was not ordered by the physician. Longitudinal data of this nature can be analyzed by the random intercepts model.[86] In this model, each subject is assumed to have a random effect, which follows a normal distribution. The time variable can be modeled as a random effect, a fixed effect, or both.[87] The model has the flexibility of allowing linear, quadratic or other forms of the time effect, and inclusion of the interaction effects between other covariates and time as fixed effects. If the repeated-measures are for a binary dependent variable, or count data, the above-mentioned benefits can be obtained by fitting the generalized liner mixed effects model. The generalized linear mixed model also assumes MAR.

5.4.2.4 Sensitivity Analyses

In addition to the above approaches, "scenario-based" sensitivity analyses should be considered for missing data. Investigators can identify "worst case" scenarios for the missing data: for missing outcomes, one such "worst case" scenario might be to assume that all exposed missing outcomes are events, while all unexposed missing outcomes are nonevents (or vice-versa). Such scenario-based approaches can help set boundaries on effect size in ways that are useful for contextualizing main results. However, since scenario-based analyses are by their nature specific to the data and situation under study, it is important to consider carefully what questions are of most substantive relevance to the study question at hand. Ideally, sensitivity analyses using different analytic approaches for missing data should be pre-specified in the protocol or a separate data analysis plan, and not done post-hoc.

Readers interested in learning more about methods for handling missing data and the potential for bias are directed to other useful resources by Greenland,[70] Greenland and Finkle,[88] Hernán and Robins,[89] Hernán and colleagues,[40] Daniel and colleagues,[77] Westreich,[90] and Lash, Fox, and Fink.[2] It is important to keep in mind that the impact of data completeness will differ, depending on the extent of missing data and the intended use of the registry. It may be less problematic with regard to descriptive research than research intended to support decision making. For all registry-based studies, it is important to have a strategy for how to identify and handle missing data as well as how to explicitly report on data completeness to facilitate interpretation of study results. (See Reporting section below.)

5.5. Machine Learning and Natural Language Processing

Machine learning (ML) uses computer algorithms to identify patterns in large datasets with a multitude of variables, and has emerged as a highly effective method for prediction and decision making in a multitude of disciplines, including natural language processing (NLP), predicting patient outcomes, and identifying eligible patients for registry analyses.[91-94]

NLP describes the computerized approach to analyzing text,[95] for example, doctors' notes for patients. With the increasing adoption of electronic medical record (EMR) systems, more of patients' text data are becoming electronic and therefore available for computer processing. However, the data in medical narrative documents are unstructured and cannot be directly utilized for analyses. Two basic approaches have been used in attempts to provide structure to the text reports a clinician creates. One of these involves avoiding the creation of unstructured data altogether through the use of a form-based user interface so that the clinician chooses to mark appropriate concepts and values from a list of possible entries. This approach has been criticized by clinicians because of its frequent inability to capture many of the nuances represented in free text documentation, as well as the increased time it requires. The other approach has been to use NLP to perform information retrieval on narrative medical documents and to convert the data in documents into a form suitable for various applications. A medical NLP system is one that is applied to processing clinical documents (such as radiology or pathology reports, or discharge summaries) that are produced during the actual clinical care of a patient.[95]

ML methods have potential pitfalls including overfitting,[96] bad sample data or training sets not being representative. Thus, results from ML need to be vigorously validated. A common and important validation strategy is cross-validation. In cross-validation, a model is fitted using only a portion of the sample data. The model is then applied to another portion of the data to test performance. Ideally, a model will perform equally well on both portions of the data. If it does not, it is likely that the model has been over fit. Another validation is manual review of patients' records, which usually applies to validation of NLP.[97]

6. Interpretation and Reporting of Registry Data

Proper interpretation of registry data is grounded in a strong understanding of the strengths and limitations of the registry methods including the analyses. Interpretation also should also be tempered, in part, by whether the analyses are attempting to confirm or refute other studies or if the registry is providing first reports. If the purpose of the registry is explicit, the actual population studied is reasonably representative of the target population, the data have been curated to enhance quality, and the analyses performed so as to reduce potential biases, then the interpretation of the registry data should allow a realistic picture of the quality of medical care, the natural history of the disease studied, or the safety, effectiveness, or value of a clinical evaluation. Each of these topics needs to be discussed in the interpretation of the registry data, and potential shortcomings should be explored.

In interpreting the findings, the precision of the estimated effect measure from the study should be discussed. Confidence intervals are important tools that provide a range of the effect measures consistent with the study findings. Statistical significance alone does not exclusively determine the clinical importance of the findings because some registries include large amounts of healthcare data, and very small effect measures can be statistically significant. Considerations should also be given to the clinical significance of the effect estimates and potential biases.[98]

Assumptions or biases that could have influenced the outcomes of the analyses should be highlighted and separated from those that do not affect the interpretation of the registry results.

Documentation of how data were collected and coded, data completeness and how missing data were addressed when reporting registry findings is important to provide transparency and to allow readers to accurately interpret registry findings. To this end, three useful guidelines are the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement,[99] the Patient-Centered Outcomes Research Institute (PCORI) Methodology Report,[100] and the GRACE checklist for observational studies of comparative effectiveness.[101]

7. Summary

In summary, a meaningful analysis requires careful consideration of study design features and the nature of the data collected. Most typical epidemiological study analytical methods can be applied, and there is no one-size-fits-all approach for registry-based research. Efforts should be made to carefully evaluate the presence of biases and to control for identified potential biases during data analysis. This requires close collaboration among clinicians, epidemiologists, statisticians, study coordinators, and others involved in the design, conduct, and interpretation of the study.

References for Chapter 13

1. Rothman K, Greenland S, Lash TL. Validity in Epidemiologic Studies. In: Rothman K, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 128-47.

2. Lash TL, Fox MP, Fink AK. Applying quantitative bias analysis to epidemiologic data: Springer; 2009.

3. Rothman K, Greenland S, Poole C, et al. Causation and Causal Inference. In: Rothman K, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 5-31.

4. Rothman K, Greenland S, Lash TL. Case-Control Studies. In: Rothman K, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 111-27.

5. Sedrakyan A, Marinac-Dabic D, Normand SL, et al. A framework for evidence evaluation and methodological issues in implantable device studies. Med Care. 2010;48(6 Suppl):S121-8. PMID: 20421824. DOI: 10.1097/MLR.0b013e3181d991c4.

6. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017;26(9):1033-9. PMID: 28913966. DOI: 10.1002/pds.4297.

7. Editors. The registration of observational studies--when metaphors go bad. Epidemiology. 2010;21(5):607-9. PMID: 20657291. DOI: 10.1097/EDE.0b013e3181eafbcf.

8. Poole C. A vision of accessible epidemiology. Epidemiology. 2010;21(5):616-8. PMID: 20657293. DOI: 10.1097/EDE.0b013e3181e9be3f.

9. Lash TL. Preregistration of study protocols is unlikely to improve the yield from our science, but other strategies might. Epidemiology. 2010;21(5):612-3. PMID: 20657295. DOI: 10.1097/EDE.0b013e3181e9bba6.

10. Lash TL, Collin LJ, Van Dyke ME. The Replication Crisis in Epidemiology: Snowball, Snow Job, or Winter Solstice? Curr Epidemiol Rep. 2018;5(2):175-83.

11. Greenland S, Rothman K. Measures of Occurrence. In: Rothman K, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 32-50.

12. National Cancer Institute. Surveillance Epidemiology and End Results. https://seer.cancer.gov/. Accessed June 10, 2019.

13. Reeves MJ, Fonarow GC, Smith EE, et al. Representativeness of the Get With The Guidelines-Stroke Registry: comparison of patient and hospital characteristics among Medicare beneficiaries hospitalized with ischemic stroke. Stroke. 2012;43(1):44-9. PMID: 21980197. DOI: 10.1161/STROKEAHA.111.626978.

14. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol. 2013;42(4):1012-4. PMID: 24062287. DOI: 10.1093/ije/dys223.

15. Schneeweiss S, Gagne JJ, Glynn RJ, et al. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. Clin Pharmacol Ther. 2011;90(6):777-90. PMID: 22048230. DOI: 10.1038/clpt.2011.235.

16. Suissa S. Effectiveness of inhaled corticosteroids in chronic obstructive pulmonary disease: immortal time bias in observational studies. Am J Respir Crit Care Med. 2003;168(1):49-53. PMID: 12663327. DOI: 10.1164/rccm.200210-1231OC.

17. Rothman K, Greenland S, Lash TL. Design strategies to improve study accuracy. In: Rothman K, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 162-82.

18. Rothman K, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998.

19. Dreyer NA. Advancing a Framework for Regulatory Use of Real-World Evidence: When Real Is Reliable. Ther Innov Regul Sci. 2018;52(3):362-8. PMID: 29714575. DOI: 10.1177/2168479018763591.

20. U.S. Food and Drug Administration. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research. Standards for Data Management and Analytic Processes in the Office of Surveillance and Epidemiology (OSE). March 3, 2008. hhttps://www.fda.gov/media/72761/download. Accessed June 10, 2019.

21. Greenland S, Lash TL. Bias Analysis. In: Rothman K, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 345-80.

22. Sorensen HT, Riis AH, Lash TL, et al. Statin use and risk of amyotrophic lateral sclerosis and other motor neuron disorders. Circ Cardiovasc Qual Outcomes. 2010;3(4):413-7. PMID: 20530788. DOI: 10.1161/CIRCOUTCOMES.110.936278

23. Anderson IB, Sorensen TI, Prener A. Increase in incidence of disease due to diagnostic drift: primary liver cancer in Denmark, 1943-85. BMJ. 1991;302(6774):437-40. PMID: 2004170. DOI: 10.1136/bmj.302.6774.437.

24. Lash TL, Johansen MB, Christensen S, et al. Hospitalization rates and survival associated with COPD: a nationwide Danish cohort study. Lung. 2011;189(1):27-35. PMID: 21170722. DOI: 10.1007/s00408-010-9274-z.

25. Rodriguez LA, Perez-Gutthann S, Jick SS. The UK General Practice Research Database. In: Strom BL, editor. Pharmacopepidemiology. 3rd ed. Chichester, UK: John Wiley & Sons, LTD; 2000. p. 375-85.

26. Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. Am J Epidemiol. 2000;151(5):531-9. PMID: 10707923. DOI: 10.1093/oxfordjournals.aje.a010240.

27. Sturmer T, Jonsson Funk M, Poole C, et al. Nonexperimental comparative effectiveness research using linked healthcare databases. Epidemiology. 2011;22(3):298-301. PMID: 21464649. DOI: 10.1097/EDE.0b013e318212640c.

28. Greenland S. Randomization, statistics, and causal inference. Epidemiology. 1990;1(6):421-9. PMID: 2090279.

29. Lash TL, Schmidt M, Jensen AO, et al. Methods to apply probabilistic bias analysis to summary estimates of association. Pharmacoepidemiol Drug Saf. 2010;19(6):638-44. PMID: 20535760. DOI: 10.1002/pds.1938.

30. Fink AK, Lash TL. A null association between smoking during pregnancy and breast cancer using Massachusetts registry data (United States). Cancer Causes Control. 2003;14(5):497-503. PMID: 12946045.

31. Lash TL, Fox MP, Thwin SS, et al. Using probabilistic corrections to account for abstractor agreement in medical record reviews. Am J Epidemiol. 2007;165(12):1454-61. PMID: 17406006. DOI: 10.1093/aje/kwm034.

32. Kleinbaum DG, Kupper LL, Miller KE, et al. Applied regression analysis and other multivariable methods. Belmont, CA: Duxbury Press; 1998.

33. Hennekens CH, Buring JE, Mayrent SL. Epidemiology in medicine. 1st ed. Boston: Little, Brown and Company; 1987.

34. Aschengrau A, Seage G. Essentials of epidemiology in public health. 2003.

35. Rosner B. Fundamentals of biostatistics. 5th ed. Boston: Duxbury Press; 2000.

36. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol. 1999;149(11):981-3. PMID: 10355372. DOI: 10.1093/oxfordjournals.aje.a009758.

37. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med. 1991;10(4):577-81. PMID: 2057656.

38. Higgins J, Green S.The Cochrane Collaboration. The Cochrane handbook for systematic reviews of interventions. 2006. http://www.cochrane.org/sites/default/files/uploads/Handbook4.2.6Sep2006.pdf. Accessed August 15, 2012.

39. Swihart BJ, Caffo B, James BD, et al. Lasagna plots: a saucy alternative to spaghetti plots. Epidemiology. 2010;21(5):621-5. PMID: 20699681. DOI: 10.1097/EDE.0b013e3181e5b06a.

40. Hernan MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002;155(2):176-84. PMID: 11790682. DOI: 10.1093/aje/155.2.176.

41. Mangano DT, Tudor IC, Dietzel C, et al. The risk associated with aprotinin in cardiac surgery. N Engl J Med. 2006;354(4):353-65. PMID: 16436767. DOI: 10.1056/NEJMoa051379.

42. Cepeda MS, Boston R, Farrar JT, et al. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol. 2003;158(3):280-7. PMID: 12882951. DOI: 10.1093/aje/kwg115.

43. Sturmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol. 2006;59(5):437-47. PMID: 16632131. DOI: 10.1016/j.jclinepi.2005.07.004.

44. Glynn RJ, Schneeweiss S, Sturmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006;98(3):253-9. PMID: 16611199. DOI: 10.1111/j.1742-7843.2006.pto_293.x.

45. Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009;20(4):512-22. PMID: 19487948. DOI: 10.1097/EDE.0b013e3181a663cc.

46. Reeve BB, Potosky AL, Smith AW, et al. Impact of cancer on health-related quality of life of older Americans. J Natl Cancer Inst. 2009;101(12):860-8. PMID: 19509357. DOI: 10.1093/jnci/djp123.

47. Brodie BR, Stuckey T, Downey W, et al. Outcomes with drug-eluting stents versus bare metal stents in acute ST-elevation myocardial infarction: results from the Strategic Transcatheter Evaluation of New Therapies (STENT) Group. Catheter Cardiovasc Interv. 2008;72(7):893-900. PMID: 19016465. DOI: 10.1002/ccd.21767.

48. Shuhaiber JH, Kim JB, Hur K, et al. Survival of primary and repeat lung transplantation in the United States. Ann Thorac Surg. 2009;87(1):261-6. PMID: 19101309. DOI: 10.1016/j.athoracsur.2008.10.031.

49. Grabowski GA, Kacena K, Cole JA, et al. Dose-response relationships for enzyme replacement therapy with imiglucerase/alglucerase in patients with Gaucher disease type 1. Genet Med. 2009;11(2):92-100. PMID: 19265748. DOI: 10.1097/GIM.0b013e31818e2c19.

50. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006;17(4):360-72. PMID: 16755261. DOI: 10.1097/01.ede.0000222409.00878.37.

51. Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. J Epidemiol Community Health. 2005;59(6):443-9. PMID: 15911637. DOI: 10.1136/jech.2004.023473.

52. Holden JE, Kelley K, Agarwal R. Analyzing change: a primer on multilevel models with applications to nephrology. Am J Nephrol. 2008;28(5):792-801. PMID: 18477842. DOI: 10.1159/000131102.

53. Diez-Roux AV. Multilevel analysis in public health research. Annu Rev Public Health. 2000;21:171-92. PMID: 10884951. DOI: 10.1146/annurev.publhealth.21.1.171.

54. Leyland AH, Goldstein H. Multilevel modeling of health statistics. Chichester, UK: John Wiley & Sons, LTD; 2001.

55. Varadhan R, Segal JB, Boyd CM, et al. A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. J Clin Epidemiol. 2013;66(8):818-25. PMID: 23651763. DOI: 10.1016/j.jclinepi.2013.02.009.

56. Palmer AJ. Health economics--what the nephrologist should know. Nephrol Dial Transplant. 2005;20(6):1038-41. PMID: 15840678. DOI: 10.1093/ndt/gfh824.

57. Neumann PJ. Opportunities and barriers. Using cost-effectiveness analysis to improve health care. New York, NY: Oxford University Press; 2004.

58. Edejer TT-T, Baltussen R, Adam T, et al. Making choices in health: WHO guide to cost-effectiveness analysis. Geneva: World Health Organization; 2004.

59. Drummond M, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. 3rd ed. New York: Oxford University Press; 2005.

60. Muennig P. Designing and conducting cost-effectiveness analyses in medicine and health care. New York: John Wiley & Sons, LTD; 2002.

61. Haddix AC, Teutsch SM, Corso PS. Prevention effectiveness: a guide to decision analysis and economic evaluation. New York: Oxford University Press; 2003.

62. Gold MR, Siegel JE, Russell LB, et al. Cost-effectiveness in health and medicine: the Report of the Panel on Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.

63. Kleinbaum DG, Klein M. Survival analysis: a self-learning text. 2nd ed. New York: Springer; 2005.

64. Twisk JWR. Applied longitudinal data analysis for epidemiology – a practical guide. Cambridge, UK: Cambridge University Press; 2003.

65. Newman SC. Biostatistical methods in epidemiology. New York: John Wiley & Sons, LTD; 2001.

66. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003;158(9):915-20. PMID: 14585769. DOI: 10.1093/aje/kwg231.

67. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007;16(3):241-9. PMID: 17252614. DOI: 10.1002/pds.1357.

68. Haynes RB, Sackett DL, Guyatt GH, et al. Clinical epidemiology. 3rd ed. New York: Lippincott Williams & Wilkens; 2005.

69. Andersen PK, Geskus RB, de Witte T, et al. Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol. 2012;41(3):861-70. PMID: 22253319. DOI: 10.1093/ije/dyr213.

70. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37-48. PMID: 9888278.

71. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615-25. PMID: 15308962.

72. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572. PMID: 9836663. DOI: 10.1136/bmj.317.7172.1572.

73. Mendelsohn AB, Dreyer NA, Mattox PW, et al. Characterization of Missing Data in Clinical Registry Studies. Therapeutic Innovation & Regulatory Science. 2015;49(1):146-54. PMID: 30222467. DOI: 10.1177/2168479014532259.

74. Applebaum KM, Malloy EJ, Eisen EA. Left truncation, susceptibility, and bias in occupational cohort studies. Epidemiology. 2011;22(4):599-606. PMID: 21543985. DOI: 10.1097/EDE.0b013e31821d0879.

75. Howards PP, Hertz-Picciotto I, Poole C. Conditions for bias from differential left truncation. Am J Epidemiol. 2007;165(4):444-52. PMID: 17150983. DOI: 10.1093/aje/kwk027.

76. Cain KC, Harlow SD, Little RJ, et al. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. Am J Epidemiol. 2011;173(9):1078-84. PMID: 21422059. DOI: 10.1093/aje/kwq481.

77. Daniel RM, Kenward MG, Cousens SN, et al. Using causal diagrams to guide analysis in missing data problems. Stat Methods Med Res. 2012;21(3):243-56. PMID: 21389091. DOI: 10.1177/0962280210394469.

78. Hernandez-Diaz S, Chambers C, Ephross S, et al. "Pregnancy Registries." In: Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014. http://www.effectivehealthcare.ahrq.gov.

79. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009;170(2):244-56. PMID: 19494242. DOI: 10.1093/aje/kwp107.

80. Riley GF. Administrative and claims records as sources of health care cost data. Med Care. 2009;47(7 Suppl 1):S51-5. PMID: 19536019. DOI: 10.1097/MLR.0b013e31819c95aa.

81. Suissa S. Immeasurable time bias in observational studies of drug effects on mortality. Am J Epidemiol. 2008;168(3):329-35. PMID: 18515793. DOI: 10.1093/aje/kwn135.

82. Bradbury BD, Wang O, Critchlow CW, et al. Exploring relative mortality and epoetin alfa dose among hemodialysis patients. Am J Kidney Dis. 2008;51(1):62-70. PMID: 18155534. DOI: 10.1053/j.ajkd.2007.09.015.

83. Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley & Sons; 1987.

84. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, LTD; 1987.

85. A Guide to Planning for Missing Data. Clinical Trials with Missing Data. p. 71-129.

86. Allison PD. Handling Missing Data by Maximum Likelihood. Paper 312-2012, SAS Global Forum 2012. http://www.statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf. Accessed June 10, 2019.

87. Loughin TM. SAS® for Mixed Models, 2nd edition Edited by Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. Biometrics. 2006;62(4):1273-4. DOI: 10.1111/j.1541-0420.2006.00596_6.x.

88. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol. 1995;142(12):1255-64. PMID: 7503045. DOI: 10.1093/oxfordjournals.aje.a117592.

89. Hernán MA, Robins JM (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming.

90. Westreich D. Berkson's bias, selection bias, and missing data. Epidemiology. 2012;23(1):159-64. PMID: 22081062. DOI: 10.1097/EDE.0b013e31823b6296.

91. Motwani M, Dey D, Berman DS, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. European heart journal. 2017;38(7):500-7. PMID: 27252451. DOI: 10.1093/eurheartj/ehw188.

92. Linden A, Yarnold PR. Combining machine learning and matching techniques to improve causal inference in program evaluation. J Eval Clin Pract. 2016;22(6):864-70. PMID: 27353301. DOI: 10.1111/jep.12592.

93. Oermann EK, Rubinsteyn A, Ding D, et al. Using a Machine Learning Approach to Predict Outcomes after Radiosurgery for Cerebral Arteriovenous Malformations. Sci Rep. 2016;6:21161. PMID: 26856372. DOI: 10.1038/srep21161.

94. Lotsch J, Ultsch A. Machine learning in pain research. Pain. 2018;159(4):623-30. PMID: 29194126. DOI: 10.1097/j.pain.0000000000001118.

95. Al-Haddad MA, Friedlin J, Kesterson J, et al. Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. HPB (Oxford). 2010;12(10):688-95. PMID: 21083794. DOI: 10.1111/j.1477-2574.2010.00235.x.

96. Glowacki J, Reichoff M. Effective Model Validation using Machine Learning. Milliman White Paper. May 2017. http://us.milliman.com/uploadedFiles/insight/2017/effective-model-validation-machine-learning.pdf. Accessed June 10, 2019.

97. Jones BE, South BR, Shao Y, et al. Development and Validation of a Natural Language Processing Tool to Identify Patients Treated for Pneumonia across VA Emergency Departments. Appl Clin Inform. 2018;9(1):122-8. PMID: 29466818. DOI: 10.1055/s-0038-1626725.

98. U.S. Food and Drug Administration. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. May 2013. https://www.fda.gov/files/drugs/published/Best-Practices-for-Conducting-and-Reporting-Pharmacoepidemiologic-Safety-Studies-Using-Electronic-Healthcare-Data-Sets.pdf, June 10, 2019.

99. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9. PMID: 25046131. DOI: 10.1016/j.ijsu.2014.07.013.

100. Patient-Centered Outcomes Research Institute. PCORI Methodology Standards. April 2018. https://www.pcori.org/sites/default/files/PCORI-Methodology-Standards.pdf. Accessed June 4, 2019.

101. Dreyer NA, Bryant A, Velentgas P. The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness. J Manag Care Spec Pharm. 2016;22(10):1107-13. PMID: 27668559. DOI: 10.18553/jmcp.2016.22.10.1107.

Case Examples for Chapter 13

Case Example 23. Understanding Baseline Characteristics of Combined Datasets Prior to Analysis

Description	The Kaiser Permanente Anterior Cruciate Ligament Reconstruction (KP ACLR) Registry was established to collect standardized data on ACLR procedures, techniques, graft types, and types of fixation and implants. The objectives of the registry are to identify risk factors that lead to degenerative joint disease, graft failure, and meniscal failure; determine outcomes of various graft types and fixation techniques; describe the epidemiology of ACLR patients; determine and compare procedure incidence rate at participating sites; and provide a framework for future studies tracking ACLR outcomes.
Sponsor	Kaiser Permanente
Year Started	2005
Year Ended	Ongoing
No. of Sites	42 surgical centers and 240 surgeons
No. of Patients	>40,000

Challenge

The KP ACLR Registry aimed to collaborate with the Norwegian Ligament Reconstruction Registry on a series of studies to proactively identify patient risk factors as well as surgical practices and techniques associated with poor surgical outcomes. Combining data from these two registries would allow for faster identification of certain risk factors and evaluation of low frequency events.

Proposed Solution

The first step was to compare the patient cohorts of the registries and the surgical practices of the two countries. Aggregate data were shared between the registries in tabular form. Analysis was conducted to identify differences that would be important to consider when making inferences about a population other than that covered by the registry. Commonalities were also identified to determine when inferences could be made from each other's analysis and when data do not need to be adjusted.

Results

The analysis found that the registries generally had similar distributions of age, gender, preoperative patient-reported knee function, and knee-related quality of life. Differences were observed between the two registries in race, sports performed at the time of injury, time to surgery, graft use, and fixation type. While these differences needed to be accounted for in future analyses of combined datasets from both registries, the results indicated that analyses of the combined datasets were likely to produce findings that could be generalized to a wider population of ACLR patients.

Following this comparison, two hypothesis-driven analyses were conducted, investigating questions using the combined registry datasets.

Key Point
Combining or pooling registry data can be a valuable approach to achieving a larger sample size for data analysis. However, it is important to identify cohort and practice differences and similarities between registries before making generalizations of registry findings to other populations or sharing data for collaboration projects.

For More Information

- https://national-implantregistries.kaiserpermanente.org/
- Granan LP, Inacio MC, Maletis GB, et al. Intraoperative findings and procedures in culturally and geographically different patient and surgeon populations. ACTA Orthop. 2012;83:577–82. PMID: 23116436. DOI: 10.3109/17453674.2012.741451.
- Maletis G, Granan LP, Inacio M, et al. Comparison of a community based anterior cruciate ligament reconstruction registry in the United States and Norway. The Journal of Bone and Joint Surgery. 2011 December;93(Supplement 3):31–6. PMID: 22262420. DOI: 10.2106/JBJS.K.00905.

Case Example 24. Using Registry Data To Evaluate Outcomes by Practice

Description	The Epidemiologic Study of Cystic Fibrosis (ESCF) Registry was a multicenter, encounter-based, observational, postmarketing study designed to monitor product safety, define clinical practice patterns, explore risks for pulmonary function decline, and facilitate quality improvement for cystic fibrosis (CF) patients. The registry collected comprehensive data on pulmonary function, microbiology, growth, pulmonary exacerbations, CF-associated medical conditions, and chronic and acute treatments for children and adult CF patients at each visit to the clinical site.
	Genentech, Inc.
	1993
	Patient enrollment completed in 2005; followup complete.
	215 sites over the life of the registry
	32,414 patients and 832,705 encounters recorded

Challenge
Although guidelines for managing cystic fibrosis patients have been widely available for many years, little is known about variations in practice patterns among care sites and their associated outcomes. To determine whether differences in lung health existed between groups of patients attending different CF care sites, and to determine whether these differences were associated with differences in monitoring and intervention, data on a large number of CF patients from a wide variety of CF sites were necessary.

As a large, observational, prospective registry, ESCF collected data on a large number of patients from a range of participating sites. At the time of the outcomes study, the registry was estimated to have data on over 80 percent of CF patients in the United States, and it collected data from more than 90 percent of the sites accredited by the U.S. Cystic Fibrosis Foundation. Because the registry contained a representative population of CF patients, the registry database offered strong potential for analyzing the association between practice patterns and outcomes.

Proposed Solution

In designing the study, the team decided to compare CF sites using lung function (i.e., FEV1 [forced expiratory volume in 1 second] values), a common surrogate outcome for respiratory studies. Data from 18,411 patients followed in 194 care sites were reviewed, and 8,125 patients from 132 sites (minimum of 50 patients per site) were included. Only sites with at least 10 patients in a specified age group (ages 6–12, 13–17, and 18 or older) were included for evaluation of that age group. For each age group, sites were ranked in quartiles based on the median FEV1 value at each site. The frequency of patient monitoring and use of therapeutic interventions were compared between upper and lower quartile sites after stratification for disease severity.

Results

Substantial differences in lung health across different CF care sites were observed. Within-site rankings tended to be consistent across the three age groups. Patients who were cared for at higher-ranking sites had more frequent monitoring of their clinical status, measurements of lung function, and cultures for respiratory pathogens. These patients also received more interventions, particularly intravenous antibiotics for pulmonary exacerbations. The study concluded that frequent monitoring and increased use of appropriate medications in the management of CF are associated with improved outcomes.

Key Point

Stratifying patients by quartile of lung function, age, and disease severity allowed comparison of practices among sites and revealed practice patterns that were associated with better clinical status. The large numbers of patients and sites allowed for sufficient information to create meaningful and informative stratification, and resulted in sufficient information within those strata to reveal meaningful differences in site practices.

For More Information

- Johnson C, Butler SM, Konstan MW, et al. Factors influencing outcomes in cystic fibrosis: a center-based analysis. Chest. 2003;123:20–7. PMID: 12527598. DOI: 10.1378/chest.123.1.20.
- Padman R, McColley SA, Miller DP, et al. Infant care patterns at Epidemiologic Study of Cystic Fibrosis sites that achieve superior childhood lung function. Pediatrics. 2007;119:E531–7. PMID: 17332172. DOI: 10.1542/peds.2006-1414.

Case Example 25. Using Registry Data To Study Patterns of Use and Outcomes

Description	The Palivizumab Outcomes Registry was designed to characterize the population of infants receiving prophylaxis for respiratory syncytial virus (RSV) disease, to describe the patterns and scope of the use of palivizumab, and to gather data on hospitalization outcomes.
	MedImmune, LLC
	2000
	2004
	256
	19,548 infants

Challenge

RSV is a leading cause of serious lower respiratory tract disease in infants and children and hospitalizations nationwide for infants under 1 year of age. Palivizumab was approved by the U.S. Food and Drug Administration (FDA) in 1998 and is indicated for the prevention of serious lower respiratory tract disease caused by RSV in pediatric patients at high risk of RSV disease. Two additional large retrospective surveys conducted after FDA approval studied the effectiveness of palivizumab in infants, again showing that it reduces the rate of RSV hospitalizations. To capture postlicensure patient demographic outcome information, the manufacturer wanted to create a prospective study that identified infants receiving palivizumab. The objectives of the study were to better understand the population receiving the prophylaxis for RSV disease and to study the patterns of use and hospitalization outcomes.

Proposed Solution

A multicenter registry study was created to collect data on infants receiving palivizumab injections. No control group was included. The registry was initiated during the 2000–2001 RSV season. Over 4 consecutive years, 256 sites across the United States enrolled infants who had received palivizumab for RSV under their care, provided that the infant's parent or legally authorized representative gave informed consent for participation in the registry. Data were collected by the primary healthcare provider in the office or clinic setting. The registry was limited to data collection related to subjects' usual medical care. Infants were enrolled at the time of their first injection, and data were obtained on palivizumab injections, demographics, and risk factors, as well as on medical and family history.

Followup forms were used to collect data on subsequent palivizumab injections, including dates and doses, during the RSV season. Compliance with the prescribed injection schedule was determined by comparing the number of injections actually received with the number of expected doses, based on the month that the first injection was administered. Infants who received their first injection in November were expected to receive five injections, whereas infants receiving their first injection in February would be expected to receive only two doses through March. Data were also collected for all enrolled infants hospitalized for RSV and were

directly reported to an onsite registry coordinator. Testing for RSV was performed locally, at the discretion of the healthcare provider. Adverse events were not collected and analyzed separately for purposes of this registry. Palivizumab is contraindicated in children who have had a previous significant hypersensitivity reaction to palivizumab. Cases of anaphylaxis and anaphylactic shock, including fatal cases, were reported following initial exposure or re-exposure to palivizumab. Other acute hypersensitivity reactions, which might have been severe, were also reported on initial exposure or re-exposure to palivizumab. Adverse reactions occurring greater than or equal to 10 percent and at least 1 percent more frequently than placebo are fever and rash. In postmarketing reports, cases of severe thrombocytopenia (platelet count <50,000/microliter) and injection site reactions were reported.

Results

From September 2000 through May 2004, the registry collected data on 19,548 infants. The analysis presented injection rates and hospitalization rates for all infants by month of injection and by site of first dose (pediatrician's office or hospital). The observed number of injections per infant was compared with the expected number of doses based on the month the first injection was given. Over 4 years of data collection, less than 2 percent (1.3%) of enrolled infants were hospitalized for RSV. This analysis confirmed a low hospitalization rate for infants receiving palivizumab prophylaxis for RSV in a large nationwide cohort of infants from a geographically diverse group of practices and clinics. The registry data also showed that the use of palivizumab was mostly consistent with the 2003 guidelines of the American Academy of Pediatrics for use of palivizumab for prevention of RSV infections. As the registry was conducted prospectively, nearly complete demographic information and approximately 99 percent of followup information was captured on all enrolled infants, an improvement compared with previously completed retrospective studies.

Key Point

A simple stratified analysis was used to describe the characteristics of infants receiving injections to help prevent severe RSV disease. Infants in the registry had a low hospitalization rate, and these data support the effectiveness of this treatment outside of a controlled clinical study. Risk factors for RSV hospitalizations were described and quantified by presenting the number of infants with RSV hospitalization as a percentage of all enrolled infants who were hospitalized. These data supported an analysis of postlicensure effectiveness of RSV prophylaxis, in addition to describing the patient population and usage patterns.

For More Information

- Leader S, Kohlhase K. Respiratory syncytial virus-coded pediatric hospitalizations, 1997-1999. Ped Infect Dis J. 2002;21(7):629–32. PMID: 12237593 DOI: 10.1097/01.inf.0000019891.59210.1c.
- Frogel M, Nerwen C, Cohen A, et al. Prevention of hospitalization due to respiratory syncytial virus: Results from the Palivizumab Outcomes Registry. J Perinatol. 2008;28:511–7. PMID: 18368063. DOI: 10.1038/jp.2008.28.
- American Academy of Pediatrics—Committee on Infectious Disease. Red Book 2003: Policy Statement: Revised indications for the use of palivizumab and respiratory syncytial virus immune globulin intravenous for the prevention of respiratory syncytial virus infections. Pediatrics. 2003;112:1442–6. PMID: 14654627.

Section IV

Evaluating Registries

Chapter 14. Assessing Quality

1. Introduction

As described throughout this guide, registries are created for many purposes, including scientific, clinical, and policy, may serve more than one purpose, and may add or change purposes over time. This leads to variations in design, operations, or quality assurance that are sometimes viewed as inadequacies. It is not generally appreciated that the attributes important for some purposes may be less important for others. As a result, it is necessary to distinguish these purposes with respect to recommending particular practices.

For example, in patients for whom there is little other systematic information available, some relevant and accurate data from a registry are better than no data. Further, even registries that fall short of including all the essential elements of good registry practice described in this chapter may still provide valuable insights for some purposes. As a general rule, quality should be evaluated by elements that directly impact the ability of the registry to achieve the purpose for which it is being used. In other words, a registry should be evaluated in the context of its "fit" for a given use.

Nonetheless, there are levels of rigor that enhance validity, and make some registries more useful for generating robust evidence than others. For example, there are certain practices that enhance the validity and reliability of registries intended to be used to characterize benefits and risks and comparative effectiveness in terms of design and validity of key exposures and outcomes.

Prior to the publication of the first edition of this User's Guide,[1] no criteria had been developed to guide evaluation of registries. Research related to the quality aspects of registries, whatever their purpose, remains relatively sparse, especially when compared with the rich information available to guide quality in clinical trials. The aim of this chapter is to provide a simple and user-friendly framework of attributes and practices that allow registries to be described and evaluated in terms of their essential elements as well as potential enhancements in the context of a given purpose. Information is presented to guide reviewers to distinguish between—

- Essential registry practices that are desirable for every study.
- Practices that could enhance scientific rigor and that are important for certain purposes, but may not be achievable because of practical constraints.

The items listed as "essential" elements of good practice are applicable to all patient registries. While it may not be practical or feasible to achieve all essential elements of good practice in all registries, it is useful to consider these characteristics in planning and evaluating registries.

It is also important to remind readers about some of the fundamental differences between clinical trials and registries, and how those may drive different measures of quality. For example, a clinical trial will have a rigorously maintained schedule of visits and assessment. A clinical trial patient who does not adhere to the schedule may be viewed as noncompliant. For most registries, treatments and assessments may be recommended, but are not mandated if the registry is adhering to a noninterventional design. In such cases, not every patient will have the same

assessments, and assessments that are performed may be done at different time intervals, making analyses challenging. Nonetheless, some argue that the kind of data and evidence produced by these registries may be more useful for inferences needed in clinical decision making because the data reflect assessments customarily used by clinicians and the constraints – such as lack of health insurance coverage for expensive tests or treatments – experienced by both clinicians and patients in real-world situations.[2] Further, these registries have the added benefit of greater generalizability since few exclusion criteria are used, supporting broader inferences for many subgroups of interest.

The information described in this User's Guide, and particularly in this chapter, is also designed for use in reporting registry study results, much like the Consolidated Standards of Reporting Trials (CONSORT) guidelines have been used for reporting of clinical trials,[3] Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational studies in general, and the Good ReseArch for Comparative Effectiveness (GRACE) checklist for observational studies of comparative effectiveness.[4-7]

2. Defining Quality

This chapter has adapted a definition of quality that was developed for randomized controlled trials.[8] The term "quality" is used here to refer to the confidence that the design, conduct, and analysis of the registry can be shown to protect against bias (systematic error) and errors in inference—that is, erroneous conclusions drawn from a study.[9] As used here, quality refers both to the data and to the conclusions drawn from analyses of these data. For more information about the types of biases that can affect observational studies, as well as strategies for addressing and even avoiding these biases to the extent feasible, see Chapters 3 and 13.

3. Measuring Quality

Quality must be evaluated in the context of the data elements themselves and the methods used to generate evidence. For example, high-quality data could yield lower quality evidence without appropriate curation and analyses.

Evaluations of the quality of any registry must be done with respect to the essential elements of the registry and those aspects that are important in the context of the purpose for which the registry data are being used. Both the internal and external validity of the data should be assessed, with the assessment tempered by considerations of cost, feasibility, and the context of other evidence available for the products, conditions, and target populations of interest.

The most commonly used method to assess quality of studies is a quality scale; there are numerous quality scales of varying length and complexity in existence, with strong opinions both for and against their use.[3,10] Each scale emphasizes distinctive dimensions of quality and therefore can yield disparate results when applied to a given study. Some scales use a summary score, derived by adding individual item scores with or without weighting. The weakness of most summary scoring approaches is that they ignore whether the items exert a bias toward the null (suggesting the erroneous interpretation that there is no effect) or tend to exaggerate the appearance of an effect when none really exists.[11] Furthermore, validation of the scales is

difficult; studies have found wide variation in the scores for a particular study both by different reviewers and the same reviewers at different times.[12]

A quality component analysis is recommended here for assessing the quality of both the data and evidence from registries. This approach uses two domains: research quality, which pertains to the scientific process (in this instance, the design and framework of registry operations) used to generate the registry data, and evidence quality, which relates to the findings derived from the registry and processes used, including data collection, site and patient recruitment, followup, data curation, safety reporting, etc., in the context of a given study purpose.[13,14] According to Lohr,[15] "[t]he level of confidence one might have in evidence turns on the underlying robustness of the research and the analysis done to synthesize that research." The individual items described as essential elements of good research practice and evidence quality can be used to guide both the creation and evaluation of registries, though there are no criteria as yet as to what proportion of elements must be satisfied in order to be considered "good enough" for various purposes.

To select the quality components for analysis, key elements identified in previous research studies and quality initiatives were reviewed; among the many consulted were Guidelines for Good Pharmacoepidemiology Practice,[16] the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Guideline on Good Clinical Practice,[17] the Council for International Organizations of Medical Sciences (CIOMS) International Guidelines for Ethical Review of Epidemiological Studies,[18] standards developed for the conduct of registry studies for patient-centered-outcomes research,[19] various reports on rating scientific evidence from observational studies[10,20] and surveillance systems,[21,22] Goldberg's review of registry evaluation methods,[23] the Meta-analysis of Observational Studies in Epidemiology (MOOSE) proposal,[24] the European League Against Rheumatism (EULAR) task force on biologic registers,[25] and Guidance for Reporting Observational Studies maintained by the Equator Network.[26] Special purpose quality guidance documents, including the Food and Drug Administration's guidance document on real-world evidence,[27] the National Medical Device Registry Task Force report,[28] the Regulators Forum Registry Working Group report,[29] the Clinical Trials Transformation Initiative (CTTI) recommendations[30] for registry trials, GRACE principles for observational studies of comparative effectiveness,[6,7,31] the Registry Evaluation and Quality Standards Tool (REQueST®) to support health technology assessments,[32] and other published literature[33-36] were also reviewed.

The results of the quality component analysis should be considered in terms of the registry purpose and in the context of the disease area (See Table 14-1). For example, a disease-specific registry that has been designed to look at natural history should not be deemed low quality simply because it is not large enough to detect rare or delayed treatment effects.

Table 14-1. Overview of registry purposes

Purpose	Description
Effectiveness	Determine clinical effectiveness, cost effectiveness, or comparative effectiveness of a test or treatment, including evaluating the acceptability of drugs, devices, or procedures for reimbursement.
Safety	Measure or monitor safety and harm of specific products and treatments, including delayed risks and comparative evaluation of safety and effectiveness.
Quality	Measure or improve quality of care, including conducting programs to measure and/or improve the practice of medicine and/or public health.
Natural history	Assess natural history, including estimating the magnitude of a problem, quantifying the underlying incidence rate or prevalence of a condition or exposure; examining trends of disease over time; conducting surveillance; assessing service delivery; identifying groups who respond well or poorly to treatments, or who are at high risk; documenting the types of patients served by a health provider; and describing and estimating survival.

4. Quality Domains

Quality domains address research methods and evidence separately.

For research methods, the quality domains are design, processes and procedures which should be considered in planning, design, selection of data elements and data sources, and ethics, privacy, and governance. Table 14-2 shows the essential elements of good registry practice for research as well as the other indicators of quality that may enhance registry validity and reliability.

For evidence, the quality domains are data and research execution, and analysis and reporting, with a focus on transparency of reporting. Table 14-3 shows the essential elements of good registry practice for evidence as well as the optional further indicators of quality, including those important for selected purposes. It is important to weigh efforts to promote the accuracy and completeness of evidence in balance with the burden of reporting, the types of interventions that are available, and the risks to public health from coming to a wrong conclusion. These lists of components are most likely incomplete, but the level of detail provided should be useful for high-level quality distinctions.

Most importantly, the essential elements of good practice, as well as the optional further indicators of quality depend, to a great extent, on the resources and budget available to support registry-based research and the feasibility of collecting the data of interest.

Table 14-2. Research quality—good practice for establishing and operating registries

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
Registry Design	GOALS	Develop goals, objectives and/or research questions (main and supporting, as needed).	Formalize the study plan as a research protocol. It may be helpful for external stakeholders to have input to ensure clinical relevance and feasibility.
	TARGET POPULATION	Describe the target population, including what requirements are needed to be eligible to participate in the registry and any exclusion criteria. For registries that are intended to study effectiveness and safety, it is often desirable to study typical patients that would use the treatments, procedure, etc. of interest. For registries where practice characteristics may influence outcome, seek to include diverse clinical practices.	Where feasible, it is desirable to study diverse patients (few exclusion criteria) to facilitate analyses of subgroups. Confirm eligibility (inclusion and exclusion criteria) upon enrollment. For studies of effectiveness and safety, use concurrent comparators, since they generally offer an advantage over historical comparison groups, especially in situations where treatments and/or diagnostics have changed over time. The comparator cohort should be as similar as possible to the exposed cohort, aside from the exposure under study.
	OBSERVATION PERIOD	Describe the followup time required to detect events of interest. If it is not feasible to conduct as much followup as desired (e.g., 20-year outcomes following hip replacement), consider whether the registry can help shape what is known about more intermediate events (e.g., 5-year outcomes following hip replacement).	Consider whether longer-term followup can be achieved through linkage with external data sources, and if it is feasible to collect appropriate identifiers required for accurate data linkage.

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
	SIZE	Determine the desired number of patients and observation time required to detect an effect should one exist, or to achieve a desired level of precision. Temper considerations about ideal study size with budgetary and feasibility constraints.	For studies of effectiveness and safety, use formal statistical calculations to estimate the number of patients or patient-years of observation needed to measure an effect with a certain level of precision or to achieve a specified statistical power to detect an effect should one exist, although the desired size may not be achievable within practical study constraints.
	DATA	Determine which variables are critical to the registry purpose and which are desirable but not critical. Focus on including those "must-have" variables that are reasonably feasible to collect and likely to be reliable, including effect-modifiers, confounders, and safety events of interest in addition to essential exposure and outcome measures. Use existing common data elements or other data standards, where appropriate, in the registry. Evaluate whether data in existing sources are of sufficient quality to achieve the registry's purpose or if existing data can be used to supplement or minimize active data collection.	Use open standard approaches to interoperability when health information systems are used for active data collection to permit more efficient collection of data from multiple systems. Consider the collection of information to permit linkage with external databases for data supplementation (e.g., using pharmacy data to assure that prescriptions were filled) or followup. Use the literature to inform the choice of data elements. Consider the value of exploratory data elements for which little published literature exists.
	EXPOSURE	Determine appropriate exposure assessments to accommodate registry purposes. For studies of a specific product(s), collect sufficient information to identify the product of interest, e.g., drug or biologic brand name or generic, code, device product and its universal device identifier (UDI), etc., as appropriate and feasible.	Collect information on start and stop dates of treatments of interest and dose (if relevant) or other means to discriminate between high and low exposure.

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
	OUTCOMES	Choose outcomes that can be identified in typical situations and that are clinically meaningful and relevant to patients and providers. Define patient outcomes clearly, especially for conditions or outcomes that may not have uniformly established criteria (e.g., define "injection site reaction" in operational terms). Consider whether existing core sets of outcome measures or other standardized measures are available and relevant for the registry purpose, and use where feasible. Use validated scales and tests when such tools exist for the purpose needed, including measures for patient-reported outcomes. When capturing composite scores, collect and record core components, if possible. Consider where these outcomes will be collected, e.g., from medical care providers, patients, or other observers, and the likelihood that such reporters will be accurate and specific enough.	Endpoints that can be confirmed by an unbiased observer, such as death and test results, are more generalizable than endpoints that are clinical impressions without established guidance for detection and recording. Consider potential sources of error relating to accuracy and falsification of data for safety and other reporting. Any such errors should be rigorously evaluated and quantified to the extent feasible (e.g., through database and/or site reviews).

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
	EFFECT MODIFIERS & CONFOUNDERS	Identify important factors or characteristics that may influence response (effect modifiers or potential confounding factors), such as other important exposures (treatment), medical history, other risk factors including personal habits, and mitigating (or protective) factors. Collect these characteristics with sufficient detail for the study purpose (e.g., current smokers vs. number of packs per day and type of tobacco product for tobacco-related study purposes.) Temper this list with the feasibility of data collection and burden of reporting.	
	SAFETY	Consider what safety events, if any, need to be reported to satisfy regulatory requirements and develop appropriate reporting plans. Consider the timing of any such reporting obligations.	Maintain appropriate documentation, such as an audit trail, to ensure proper handling of safety information.
	ANALYSIS PLAN	Create a high-level data analysis plan to address the key objectives or research questions, e.g., how exposures and outcomes will be compared and what comparative information, if any, will be used. Determine how missing data will be handled for key variables. Describe how composite variables will be created.	Develop formal analysis plans.

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
Framework	ETHICS & DATA PROTECTION	Evaluate the issues of protection of human subjects—including privacy, informed consent, data security, and study ethics—and address them in accordance with local, national, and international regulations. Obtain review and approval by any required oversight committees (e.g., ethics committee, privacy committee, or institutional review board as applicable). Identify appropriate personnel and facilities, including those for secure data storage.	For any data linkage activities, determine appropriate methods for collecting and storing such protected health information. Identify the individual(s) responsible for the integrity of the data, computerized and hard copy, and make sure these individuals have the training and experience to perform the assigned tasks.
	GOVERNANCE	Develop a clear, written plan for registry governance that specifies how registry decisions will be made and describes the roles of any external advisors. Define the role of any external sponsor, including data access, use, and rights to review, participate or approve any publications.	Consider using an advisory committee(s) to guide registry objectives, data collection methods, analysis, interpretation, and dissemination. Advisory committee members who are external to the clinician, center, or company that sponsors the research may bring added value for scientific and methodologic purposes, and may also assist with internal or external communications. If using an advisory committee, consider how decisions or recommendations will be agreed on (consensus or voting) and whether term limits and rotation would be helpful.

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
	TRANSPARENCY	Consider if, when, and how to allow third parties access to data, if feasible, and the process for any such data access. Assure that any data transfers are accurate, only provide the requisite data, and maintain the privacy of patients, clinicians and health systems. Plan how study results will be communicated on completion and whether the results will be made public, and if so, by whom. Consider posting information on a public registry of patient registries (e.g., at the Registry of Patient Registries).	Specify publication policies in advance of collecting data and reevaluate as needed. It may be desirable to make key elements of the protocol, analytic methods, and results publicly accessible to promote transparency and allow other researchers to know what data might be accessible through the registry, or to consider using similar methods and study populations to confirm, refute, or extend studies derived from the registry.
	CHANGE PROCESS	Establish a process for documenting any modifications to the research plan, since the main objective(s) and analytic plans may change over time as knowledge accumulates, and the plan for data collection and followup may need to be adapted.	Develop plans for periodic review of analytic plans and frequency of analyses to extract the greatest value from the registry by learning through experience and remaining current with external scientific and public health advances and needs. Develop a plan for stopping or transitioning the registry, including any archiving or transferring of data and notifying participants, as appropriate.

Table 14-3. Evidence quality—indicators of good evidence quality for registries

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
Methods: Data Collection, Curation, and Documentation	DATA COLLECTION	Use an efficient, reliable, and affordable means to collect data consistently of sufficient quality to meet the registry's purpose. Prioritize simplicity and accuracy to the extent feasible. Consider using tools for automated data extraction from existing records when they are available, affordable and likely to be repeatable on a consistent basis.	
	SITE AND PATIENT RECRUITMENT AND FOLLOWUP	For primary data collection, develop plans for patient and site enrollment. Determine how to obtain followup data, using similar methods for all subjects in each group, to the extent feasible. Plan to expend reasonable efforts to ensure that appropriate patients are enrolled and followed systematically. Methods used for followup, including efforts to minimize loss to followup, should be documented. Use similar methods for following all study subjects to the extent feasible.	Depending on study purpose, it may be desirable to include researchers with a range of experience and expertise. A feasibility study or pilot test (e.g., when studying hard-to-reach populations, when sensitive data are sought, and when critical registry methods are new or have not otherwise been tested) can be useful to assure that the study plan is feasible and sufficiently appealing to contributors and participants. Enrollment logs are useful to document sequential enrollment. Loss to followup should be evaluated periodically to see if there is differential loss to followup, and, if it is remediable.
	DATA COLLECTION GUIDANCE	Methods for data collection should be documented. Provide clear, operational definitions of outcomes and other data elements. Documentation should provide explicit	Whenever possible, use standardized data dictionaries, such as the International Classification of Diseases, and use coding that is consistent with nationally or internationally approved coding systems to

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
		definitions of data elements and their coding. Develop standard instructions for use in training data collectors. For safety studies, create a process for identifying and reporting serious events that is consistent with regulatory requirements. For studies with direct patient contact, plan training for study personnel about how to identify or recognize serious events, including (1) asking about complaints or adverse events in a manner that is clear and specific (e.g., solicited vs. unsolicited), and (2) knowing if, how and when information should be reported to manufacturers and health authorities. For studies using existing data sources, use uniform and systematic methods for collecting and curating the data to assure that the appropriate data have been extracted and linked (where applicable).	promote comparability of information among studies. Methods used for data transformations should be recorded. Uniform written guidance for any active medical chart review and abstraction will enhance validity and reliability. For studies linking to or integrating existing data sources, document the process for record linkage and whether probabilistic or deterministic matching was used. Consider any additional requirements that may influence successful linkage, such as selection of data elements used to assure accurate matching.
	QUALITY ASSURANCE	Develop a data handling and analysis plan that describes any quality assurance and data curation activities that will be implemented. Any quality assurance procedures used must be fit for purpose and should be focused on variables that are essential for analysis, such as endpoints of primary interest. Data checks should use range and consistency checks and may also include a review of consistency and comparability of	Quality assurance (QA) may include review or monitoring of a sample of data and/or data review by an adjudication committee for complex conditions or endpoints for which established procedures and/or coding are not used. For primary data collection, a sample of data collected should be compared with patient records, (e.g., 5–20% of patients' records) to assure validity and reproducibility of abstraction and coding.

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
		data across sites, and with external data sources. Methods should be described for data curation, e.g., quality control procedures to enhance internal validity, review of consistency and comparability of data across sites, and any comparisons that will be made with external data sources.	For some studies and some outcomes, validation of endpoints may be desirable depending on the study purpose (e.g., may be necessary for some regulatory uses). If the registry chooses to implement a system of periodic monitoring for quality assurance, a risk-based strategy should be used to focus on detecting and quantifying the most likely causes of error and the types of error most likely to affect the registry purpose, with QA activities adapted based on observed performance (e.g., increase QA for sites that appear to be having difficulty in study conduct or data entry). Establish processes and standards for creating analytic data files and maintaining such files to support publications and presentations, since registry analyses may be performed on live data (data that may change as the registry continues to collect and verify information through various quality control procedures) or on data that have been locked and have undergone formal review and editing.
Reporting	OVERALL REPORTING	Registry reports or publications should describe the methods, including target population and selection of sites and study subjects, compliance with applicable regulatory rules and regulations, data collection and curation/quality control methods,	Completeness of information on eligible patients should be evaluated and described for key variables of interest for the main exposures and/or outcomes of primary interest.

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
		statistical methods for data analysis, and any circumstances that may have affected the quality or integrity of the data. The information should be reported with enough detail to allow replication of the methods in another study. Followup time should be described to enable assessment of the impact of the observation period on the conclusions drawn.	
	ANALYTICS	Results should be reported for all the main objectives, including estimates of effect for each (where relevant) and confidence intervals where feasible. The data elements used in any models should be described. For safety studies, the risks and/or benefits of products, devices, or processes under study should be quantitatively evaluated beyond simply evaluating statistical significance (e.g., rates, proportions, and/or relative risks, as well as confidence intervals, were reported). The role and impact of missing data and potential confounding factors should be considered. The findings should be compared and contrasted with other relevant research.	Information about data analyses, including transformation of variables and/or construction of composite endpoints and how missing data were handled should be reported with sufficient detail to permit replication. Appropriate analytic approaches should be used to address confounding. Sensitivity analyses should be used to examine and quantify the potential impact on the association between the exposure(s) of interest and the outcome(s) by, for example, quantifying how the results would change if all the missing data were the same, e.g., respondents lied about medical adherence or certain risk factors likely to have a strong effect on the outcome, or all missing data on smokers was from people who smoke. Selection bias should be evaluated by describing the representativeness of the actual population in terms of how it was selected, how well the

Domain Category	Domain	Essential Elements of Good Practice	Enhancements
			characteristics of the actual population match those of the target population, and to whom the results apply.
	COMPARISONS	For comparative studies, comparators reflect medical practice for the appropriate reference time period. When other reasonably accurate and relevant contemporaneous comparative data are not available, historical data should be used with appropriate justification. Contemporaneous comparators in similar health systems or sites are preferable when feasible.	External validity should be described by showing how registry participants compare to known characteristics of the target population in terms of key characteristics and other factors that may be used to characterize the patients, providers and healthcare settings that were studied.

References for Chapter 14

1. Gliklich RE, Dreyer NA, editors. Registries for Evaluating Patients Outcomes: A User's Guide. Rockville, MD: Agency for Healthcare Research and Quality; Apr, 2007.

2. Dreyer NA, Garner S. Registries for robust evidence. JAMA. 2009;302(7):790-1. PMID: 19690313. DOI: 10.1001/jama.2009.1092.

3. Moher D, Schulz KF, Altman D, et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA. 2001;285(15):1987-91. PMID: 11308435.

4. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9. PMID: 25046131. DOI: 10.1016/j.ijsu.2014.07.013.

5. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4(10):e297. PMID: 17941715. DOI: 10.1371/journal.pmed.0040297.

6. Dreyer NA, Bryant A, Velentgas P. The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness. J Manag Care Spec Pharm. 2016;22(10):1107-13. PMID: 27668559. DOI: 10.18553/jmcp.2016.22.10.1107.

7. Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. Am J Manag Care. 2010;16(6):467-71. PMID: 20560690.

8. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995;16(1):62-73. PMID: 7743790.

9. Rothman K, editor. Causal inference. Chestnut Hill, MA: Epidemiology Resources, Inc.; 1988.

10. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evid Rep Technol Assess (Summ). 2002(47):1-11. PMID: 11979732.

11. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. Am J Epidemiol. 1994;140(3):290-6. PMID: 8030632. DOI: 10.1093/oxfordjournals.aje.a117248.

12. Berkman ND, Lohr KN, Ansari M, et al. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Methods for Effective Health Care. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.

13. Mosteller F, Boruch R, editors. Evidence matters: randomized trials in education research. Washington, DC: The Brookings Institute; 2002.

14. Shavelson RJ, Towne L, editors. Scientific research in education. Washington, DC: National Research Council, National Academy Press; 2002.

15. Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. Int J Qual Health Care. 2004;16(1):9-18. PMID: 15020556. DOI: 10.1093/intqhc/mzh005.

16. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf. 2008;17(2):200-8. PMID: 17868186. DOI: 10.1002/pds.1471.

17. European Medicines Agency. ICH Topic E6(R2) Guideline for Good Clinical Practice. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r1-guideline-good-clinical-practice-step-5_en.pdf. Accessed June 10, 2019.

18. CIOMS. 1991 International Guidelines for Ethical Review of Epidemiological Studies. http://www.cioms.ch/publications/guidelines/1991_texts_of_guidelines.htm. Accessed June 10, 2019.

19. Gliklich R, Dreyer NA, Leavy MB, et al. Standards in the conduct of registry studies for patient-centered outcomes research. Prepared for the Patient-Centered Outcomes Research Institute, Washington, DC, 2012. http://www.pcori.org/assets/Standards-in-the-Conduct-of-Registry-Studies-for-Patient-Centered-Outcomes-Research.pdf. Accessed June 10, 2019.

20. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62(10):1013-20. PMID: 19230606. DOI: 10.1016/j.jclinepi.2008.10.009.

21. Centers for Disease C. Guidelines for evaluating surveillance systems. MMWR Suppl. 1988;37(5):1-18. PMID: 3131659.

22. German RR, Lee LM, Horan JM, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001;50(RR-13):1-35; quiz CE1-7. PMID: 18634202.

23. Goldberg J, Gelfand HM, Levy PS. Registry evaluation methods: a review and case study. Epidemiol Rev. 1980;2:210-20. PMID: 7000537. DOI: 10.1093/oxfordjournals.epirev.a036224.

24. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283(15):2008-12. PMID: 10789670.

25. Dixon WG, Carmona L, Finckh A, et al. EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. Ann Rheum Dis. 2010;69(9):1596-602. PMID: 20525843. DOI: 10.1136/ard.2009.125526.

26. Equator Network. http://www.equator-network.org/. Accessed June 10, 2019.

27. U.S. Food and Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. Guidance for Industry and Food and Drug Administration Staff. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM513027.pdf. Accessed June 3, 2019.

28. Recommendations for a National Medical Device Evaluation System. A Report from the Medical Device Registry Task Force & the Medical Devices Epidemiology Network. August 20, 2015. http://mdepinet.org/wp-content/uploads/2015-08-Recommendations-for-a-National-Medical-Device-Evaluation-System.pdf. Accessed June 10, 2019.

29. International Medical Device Regulators Forum. Patient Registry: Essential Principles. October 2, 2015. http://www.imdrf.org/docs/imdrf/final/consultations/imdrf-cons-essential-principles-151124.pdf. Accessed June 10, 2019.

30. Clinical Trials Transformation Initiative. CTTI Recommendations: Registry Trials. May 2017. https://www.ctti-clinicaltrials.org/files/recommendations/registrytrials-recs.pdf. Accessed June 10, 2019.

31. Dreyer NA, Velentgas P, Westrich K, et al. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. J Manag Care Spec Pharm. 2014;20(3):301-8. PMID: 24564810. DOI: 10.18553/jmcp.2014.20.3.301.

32. Registry Evaluation and Quality Standards Tool (REQueST®). European Network for Health Technology Assessment (EUnetHTA). https://www.eunethta.eu/request-tool-and-its-vision-paper/. Accessed February 13, 2020.

33. National Academies of Sciences, Engineering, and Medicine. 2017. Real-world evidence generation and evaluation of therapeutics: Proceedings of a workshop. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/24685.

34. Berger M, Daniel G, Frank K., et al. A Framework for Regulatory Use of Real-World Evidence. The Margolis Center for Health Policy at Duke University. https://healthpolicy.duke.edu/sites/default/files/atoms/files/rwe_white_paper_2017.09.06.pdf. Accessed June 10, 2019.

35. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017;26(9):1033-9. PMID: 28913966. DOI: 10.1002/pds.4297.

36. Miksad RA, Abernethy AP. Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. Clin Pharmacol Ther. 2018;103(2):202-5. PMID: 29214638. DOI: 10.1002/cpt.946.

Contributors and Reviewers

Contributors

The findings and conclusions in this document are those of the contributors and reviewers. The contributors and reviewers participated as individuals and not necessarily as representatives of their organizations. No statement in this document should be construed as an official position of any organization represented here.

**Chapter Coordinator*

1. Patient Registries

*Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1
Boston, MA

2. Planning a Registry

Thomas Gross, M.D., M.P.H.
Director
Office of Surveillance and Biometrics
Center for Devices and Radiological Health
U.S. Food and Drug Administration
Silver Spring, MD

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1
Boston, MA

3. Registry Design

Nancy A. Dreyer, Ph.D., M.P.H., FISPE
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions
Cambridge, MA

Hesha J. Duggirala, Ph.D., M.P.H.
Senior Epidemiologist
Office of Surveillance and Compliance
Center for Veterinary Medicine
U.S. Food and Drug Administration
Silver Spring, MD

*Vandana Menon, M.D.
Vice President, Research
OM1
Boston, MA

4. Selecting and Defining Outcome Measures for Registries

Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

*Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1
Boston, MA

5. Data Elements for Registries

Irene Katzan, M.D., M.S.
Director, Neurological Institute Center for Outcomes Research & Evaluation
Cleveland Clinic Foundation
Cleveland, OH

*Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1
Boston, MA

Christina Mack, Ph.D., M.S.P.H.
Sr. Director, Epidemiology and Clinical Evidence
Head, Injury Surveillance and Analytics
IQVIA Real-World & Analytic Solutions
Chapel Hill, NC

John Spertus, M.D., M.P.H., FACC, FAHA
Director of CV Education and Outcomes Research
Saint Luke's Hospital
Kansas City, MS

6. Data Sources

Fang Li, M.D.
Formerly Sr. Director, Clinical Informatics
OM1
Boston, MA

*Keith Marsolo, Ph.D.
Associate Professor, Department of Population Health Sciences
Member, Duke Clinical Research Institute
Duke University School of Medicine
Durham, NC

7. Principles of Registry Ethics, Data Ownership, and Privacy

*Jiayan Chen, J.D.
Partner
McDermott Will & Emery LLP
Washington, DC

Andy DeMayo, J.D.
Senior Counsel and Head, Strategic Partnerships
OM1
Boston, MA

Deepali Doddi, J.D.
Associate
McDermott Will & Emery LLP
Chicago, IL

Scott A. Weinstein, J.D.
Partner
McDermott Will & Emery LLP
Washington, DC

8. Informed Consent

*Jessica Berg, J.D., M.P.H.
Dean, School of Law
Tom J.E. and Bette Lou Walker Professor of Law
Professor of Bioethics & Public Health
Case Western Reserve University
Cleveland, OH

9. Registry Governance

*Nancy A. Dreyer, Ph.D., M.P.H., FISPE
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions
Cambridge, MA

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1
Boston, MA

10. Recruiting and Retaining Participants in the Registry

*Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

Ralph G. Brindis, M.D., M.P.H., MACC, FSCAI
Clinical Professor of Medicine
Department of Medicine & the Philip R. Lee Institute for Health Policy Studies
University of California, San Francisco
Senior Medical Officer, External Affairs – ACC National Cardiovascular Registry
San Francisco, CA

Robert E. Harbaugh, M.D., FAANS, FACS, FAHA
Senior Vice President, Penn State Health
Distinguished Professor and Chair, Department of Neurosurgery
Professor, Department of Engineering Science and Mechanics
Penn State University, M.S. Hershey Medical Center
Hershey, PA

11. Obtaining Data and Quality Assurance

*Keith Marsolo, Ph.D.
Associate Professor in Population Health Sciences
Duke University School of Medicine
Durham, NC

Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

12. Adverse Event Detection, Processing, and Reporting

Nancy A. Dreyer, M.P.H., Ph.D.
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions
Cambridge, MA

Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

*Marnie Hall, Ph.D., M.P.H., M.A.
Vice President, Center for Advanced Evidence Generation
IQVIA
Washington, DC

13. Analysis, Interpretation, and Reporting of Registry Data To Evaluate Outcomes

Nancy A. Dreyer, M.P.H., Ph.D.
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions
Cambridge, MA

*Zhaohui Su, Ph.D.
Vice President, Biostatistics
OM1
Boston, MA

14. Assessing Quality

Nancy A. Dreyer, M.P.H., Ph.D.
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions
Cambridge, MA

Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

We would also like to acknowledge the authors who contributed to previous editions of the user's guide. These authors are listed below in alphabetical order:

Amy P. Abernethy, M.D.
Susan Adams, J.D.
Fred Anderson, Ph.D.
Anthony L. Asher, M.D., FACS
Landen Bain
Ethan Basch, M.D., M.Sc.
Marc L. Berger, M.D.
Maya A. Bernstein, J.D.
Stella Blackburn, M.B., B.S., M.Sc., FRCP (Edin), FISPE
Jaclyn L. F. Bosco, Ph.D., M.P.H.
Brian D. Bradbury, M.A., D.Sc.
Jeanette M. Broering, R.N., Ph.D., M.P.H.
Laura D. Buccini, Dr.P.H., M.P.H.
Christina Chambers, Ph.D., M.P.H.
J. Alexander Cole, D.Sc., M.P.H.
Deborah Covington, Dr.P.H.
Janet D. Cragan, M.D., M.P.H.
David C. Dale, M.D.
Andy DeMayo, J.D.
Wei Dong, M.D., Ph.D.
Paul Eggers, Ph.D.
Alexander Elbert, Ph.D.
Benjamin C. Eloff, Ph.D.
Sara A. Ephross, Ph.D.
Sarah Garner, Ph.D., B.Pharm., M.R.Pharm.S.
Eric Gemmen, M.A.
Lisa Griffin Vincent, M.A., Ph.D.
Melissa M. Goldstein, J.D.
Stephen C. Groft, Pharm.D.
Stephen E. Fienberg, Ph.D., FRSC
Joanna Haas, M.D., M.Sc.
Melanie J. Harrison, M.D., M.S.C.E.
Sonia Hernández-Díaz, M.D., M.P.H., Dr.P.H.
Adrian F. Hernandez, M.D., M.H.S.
Kathleen Hewitt, M.S.N., R.N., A.A.C.C.
Holly L. Howe, Ph.D.
Alan F. Karr, Ph.D.
David W. Kaufman, Sc.D.
Stuart Kim, J.D., M.S.
James K. Kirklin, M.D.
Catherine Koepper
Joel M. Kremer, M.D.
Harlan M. Krumholz, M.D., S.M.
Kimary Kulig, Ph.D., M.P.H.

Richard E. Kuntz, M.D., M.Sc
Debra R. Lappin, J.D.
Timothy L. Lash, D.Sc., M.P.H.
Daniel Levy, M.S.
David J. Malenka, M.D.
Bradley Malin, Ph.D.
Keith Mandel, M.D.
Neal Mantick, M.S.
Danica Marinac-Dabic, M.D., Ph.D.
Bruce C. Marshall, M.D.
David Matcher, M.D.
C. Geoffrey McDonough
Benjamin Miriovsky, M.D.
Barry Mirrer, M.S., CMQ/OE
David C. Naftel, Ph.D.
Sharon-Lise T. Normand, Ph.D.
Elizabeth Paxton, M.A.
Eric Peterson, M.D., M.P.H.
Rachel L. Richesson, Ph.D., M.P.H.
Sara Rosenbaum, J.D.
Kenneth J. Rothman, Dr.P.H.
Yaffa Rubinstein, Ph.D.
Gregory Samsa, Ph.D.
Gillian D. Sanders, Ph.D.
Lee H. Schwamm, M.D., FAHA
Art Sedrakyan, M.D., Ph.D.
David M. Shahian, M.D.
Neha Sheth, Pharm.D.
Henrik Toft Sørensen, M.D., Ph.D., Dr.Med.Sci.
Paul Stang, Ph.D.
Kathryn Starzyk, Sc.M.
Charlie Strange, M.D.
Michelle Tarver, M.D., Ph.D.
Melissa S. Tassinari, Ph.D.
Dale R. Tavris, M.D., M.P.H.
Joseph Terdiman, M.D., Ph.D.
Jane Hyatt Thorpe, J.D.
Stephen Toovey, M.D., Ph.D.
Anne E. Trontell, M.D., M.P.H.
Priscilla Velentgas, Ph.D.
Marianne Ulcickas Yood, D.Sc., M.P.H.

Reviewers

The findings and conclusions in this document are those of the contributors and reviewers. The contributors and reviewers participated as individuals and not necessarily as representatives of their organizations. No statement in this document should be construed as an official position of any organization represented here.

Amisha Barochia, M.B.B.S., M.H.S.
Staff Clinician
National Heart, Lung, and Blood Institute
Bethesda, MD

Regan W. Bergmark, M.D.
Associate Surgeon, Otolaryngology-Head and Neck Surgery
Patient Reported Outcomes, Value and Experience (PROVE) Center
Center for Surgery and Public Health
Brigham and Women's Hospital
Instructor, Department of Otolaryngology
Harvard Medical School
Boston, MA

Seth Blumenthal, M.B.A.
Director, Clinical Review Group
Integrated Health Model Initiative
American Medical Association
Chicago, IL

Allison Bryant, M.P.H.
Epidemiologist
OM1
Boston, MA

Gary Curhan, M.D., Sc.D.
Professor of Medicine
Harvard Medical School
Professor, Department of Epidemiology
Harvard T.H. Chan School of Public Health
Boston, MA

Nancy A. Dreyer, M.P.H., Ph.D.
Chief Scientific Officer
IQVIA Real-World & Analytic Solutions
Cambridge, MA

Paula DiStabile, R.N., M.S.N., J.D.
Health Scientist Administrator
Center for Quality Improvement and Patient Safety (CQuIPS)
Agency for Healthcare Research and Quality
Rockville, MD

Chris Crowe, Ph.D.
National Mental Health Director-Psychotherapy
Office of Mental Health and Suicide Prevention
U.S. Department of Veterans Affairs
Washington, DC

Diane Cousins, R.Ph.
Center for Quality Improvement and Patient Safety (CQuIPS)
Agency for Healthcare Research and Quality
Rockville, MD

Andy DeMayo, J.D.
Senior Counsel and Head, Strategic Partnerships
OM1
Boston, MA

Richard E. Gliklich, M.D.
Chief Executive Officer and Chairman
OM1
Boston, MA

Julia G. Gorey, J.D.
Executive Director, SACHRP
U.S. Department of Health and Human Services
Rockville, MD

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

Daniel Levy, M.S.
Vice President, Data Solutions
OM1
Boston, MA

Theodore Lystig, Ph.D.
Director, Corporate Biostatistics
Strategic & Scientific Operations
Medtronic
Minneapolis, MN

Elisha Malanga, M.S.
Chief Research Officer
COPD Foundation
Washington, DC

Sonia Maruti, Ph.D.
Director of Epidemiology
Sanofi Genzyme
Cambridge, MA

Aaron B. Mendelsohn, Ph.D., M.P.H.
Research Scientist
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute
Boston, MA

Vandana Menon, M.D.
Vice President, Research
OM1
Boston, MA

Sanghamitra Mohanty, M.D., M.S., FHRS
Director, Translational Research
Texas Cardiac Arrhythmia Institute
Austin, TX

Kathleen Mortimer, Ph.D.
Director of Epidemiology
OM1
Boston, MA

Marissa Gordon-Nguyen, J.D.
Senior Advisor for HIPAA Policy
Office for Civil Rights
U.S. Department of Health and Human Services
Washington, D.C.

Bryan Olin, Ph.D.
Senior Vice President, Clinical, Quality Assurance & Regulatory Affairs
LivaNova PLC
London, UK

Sarah Rittner, M.A.
Director of Operations
Center for Health Equity Transformation
Northwestern University
Chicago, IL

Kathryn Starzyk, Sc.M.
Senior Director, Real-World Evidence
OM1
Boston, MA

Zhaohui Su, Ph.D.
Vice President, Biostatistics
OM1
Boston, MA

Paul Wallace, M.D.
Senior Scholar
AcademyHealth
Washington, DC

David S. Wendler, M.A., Ph.D.
Head, Unit on Vulnerable Populations Department of Bioethics
National Institutes of Health Clinical Center
Bethesda, MD

Albert Wu, M.D., M.P.H.
Professor of Health Policy and Management
Director, Center for Health Services and Outcomes Research
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD

We would also like to acknowledge the reviewers who contributed to previous editions of the user's guide. These reviewers are listed below in alphabetical order:

Susan Adams, J.D.
Paul S. Appelbaum, M.D.
Diane J. Aschman, M.S.
Steven J. Atlas, M.D., M.P.H.
Stephen W. Bernstein, J.D.
Marcella Blackledge, B.A.
Kathleen Blake, M.D., M.P.H.
Jaclyn L. F. Bosco, Ph.D., M.P.H.
Jeanette M. Broering, R.N., Ph.D., M.P.H.
Jeffrey Brown, Ph.D.
Daniel Caños
Michael A. Carome, M.D.
J. Alexander Cole, D.Sc., M.P.H.
Jack L. Cronenwett, M.D.
Jay Crowley
Theresa Cullen, M.D., M.S.
David C. Dale, M.D.
Leonard D'Avolio, Ph.D.

Joseph A.C. Delaney, Ph.D.
Patrice Desvigne-Nickens, M.D.
Douglas S. Diekema, M.D., M.P.H.
Rachel S. Dokholyan, M.P.H.
Pamela S. Douglas, M.D., FACC
Joseph P. Drozda, Jr., M.D., FACC
Hesha J. Duggirala, Ph.D., M.P.H.
Brenda K. Edwards, Ph.D.
Paul Eggers, Ph.D.
Jonathan S. Einbinder, M.D., M.P.H.
Benjamin C. Eloff, Ph.D.
Barbara Evans, J.D., Ph.D.
Shamiram Feinglass, M.D., M.P.H.
T. Bruce Ferguson, M.D., FAHA
Leslye K. Fitterman, Ph.D.
Sally Flanzer, Ph.D., CIP
Mark J. Fox
Lori Frank, Ph.D.
Gregg C. Fonarow, M.D.
Sarah Garner, Ph.D., B.Pharm., M.R.Pharm.S.
Yvonne Geissbuehler, Ph.D.
Eric Gemmen, M.A.
Julian M. Goldman, M.D.
Marissa Gordon-Nguyen, J.D.
Julia G. Gorey, J.D.
Marie R. Griffin, M.D., M.P.H.
Thomas Gross, M.D., M.P.H.
Claudia Grossmann, Ph.D.
Joanna Haas, M.D., M.Sc.
Rosemarie Hakim, Ph.D.
John D. Halamka, M.D., M.S.
Stephen C. Hammill, M.D.
Melanie J. Harrison, M.D., M.S.C.E.
Christina M. Heide, J.D.
Kathleen Hewitt, M.S.N., R.N., A.A.C.C., CPHQ
Janet Hiller, M.P.H., Ph.D.
Barrie J. Hurwitz, M.B., MRCP
Michael Ibara, Pharm.D.
Solomon Iyasu, M.D., M.P.H.
Emma James, Ph.D.
Robert A. Jenders, M.D., M.S., FACP, FACMI
Peter I. Juhn, M.D., M.P.H.
Michael Kahn, M.D., Ph.D
Clifford Ko, M.D.
Catherine Koepper
Rebecca D. Kush, Ph.D.

Joseph A. Ladapo, M.D., Ph.D.
Julia I. Lane
Wendy Lipworth, M.D., Ph.D.
Peter Littlejohns, M.D.
Danica Marinac-Dabic, M.D., Ph.D.
Bruce C. Marshall, M.D.
Susan McAndrew, J.D.
Newell McElwee, Pharm.D., M.S.P.H.
J. Michael McGinnis, M.D., M.P.P.
Deven McGraw, J.D., M.P.H., L.L.M.
Charles R. McKay, M.D., FACC
Aaron B. Mendelsohn, Ph.D., M.P.H.
Marissa A. Miller, D.V.M., M.P.H.
Barry Mirrer, M.S., CMQ/OE
Kristi R. Mitchell, M.P.H.
Lawrence H. Muhlbaier, Ph.D.
Barbara Muth
David C. Naftel, Ph.D.
Steven E. Nissen, M.D.
J. Marc Overhage, M.D.Ph.D.
Dimitrios Pappas, M.D., M.P.H.
Gareth Parry, Ph.D.
Steven Pearson, M.D.
Eric Peterson, M.D., M.P.H.
Steve E. Phurrough, M.D., M.P.A.
Richard Platt, M.D., M.Sc.
Richard L. Prager, M.D.
Victoria M. Prescott, J.D.
Michael T. Rapp, M.D., J.D., FACEP
Robert Reynolds, Sc.D.
Gerald F. Riley, M.S.P.H.
Sara Rosenbaum, J.D.
Kenneth J. Rothman, Dr.P.H.
Yaffa Rubinstein, Ph.D.
Patricia Russo, Ph.D.
Susan T. Sacks, Ph.D.
Grannum Sant, M.D., FRCS, FACS
Brian C. Sauer, Ph.D., M.S.
William B. Saunders, Ph.D., M.P.H.
Judy A. Staffa, Ph.D., R.Ph.
Juergen Schmider, M.D., Ph.D.
Art Sedrakyan, M.D., Ph.D.
James G. Snell, J.D.
Paul Stang, Ph.D.
Michael Steinbuch, Ph.D.
Walter L. Straus, M.D., M.P.H.

Elizabeth J. Swain
Michelle Tarver, M.D., Ph.D.
Hugh Tilson, M.D., M.P.H., Dr.P.H.
Anne E. Trontell, M.D., M.P.H.
Kelly Trout, B.S.N., R.N.
Jack V. Tu, M.D., Ph.D.
Paul D. Varosy, M.D
Priscilla Velentgas, Ph.D.
Cunlin Wang, M.D., Ph.D.
Jennifer Weisman, Ph.D.
Suzanne L. West, M.P.H, Ph.D.
Marcy Wilder, J.D.
Marcus D. Wilson, Pharm.D.
Albert Wu, M.D., M.P.H.
John Yee, M.D., M.P.H.

Case Example Contributors

The findings and conclusions in this document are those of the contributors and reviewers. The contributors and reviewers participated as individuals and not necessarily as representatives of their organizations. No statement in this document should be construed as an official position of any organization represented here.

1. Creating a Registry To Fulfill Multiple Purposes and Using a Publications Committee To Review Data Requests

Edna Stoehr
Global Study Manager
Genentech, Inc.
South San Francisco, CA

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

2. Determining When To Stop an Open-Ended Registry

Jessica Albano, Ph.D., M.P.H.
Executive Director, Epidemiology & Health Outcomes
Syneos Health Clinical Solutions
Sewickley, PA

Ann Mallard, M.P.H.
Executive Director, Non-interventional Studies
Evidera
Wilmington, NC

Pattra Mattox, S.M.
Senior Manager
Aetion
Boston, MA

3. Modifying a Registry Due To Changes in Standards of Care

Dawn Eng, R.N., B.S.N.
Formerly Senior Clinical Trial Head, Oncology Global Medical Affairs
Novartis Pharmaceuticals Corporation
East Hanover, NJ

Alexandra Parnell, M.S.
Regulatory Affairs Manager, Real World Analytics Solutions
IQVIA
Boston, MA

4. Using Registries To Understand Rare Diseases

Betsy Bogard, M.S.
Head of Real-World Evidence
bluebird bio
Cambridge, MA

Catherine Koepper
Vice President, Head of Alliance Management
Sanofi Genzyme
Cambridge, MA

Nicole Robichaud
Senior Project Manager, Global Registry Program
Sanofi Genzyme
Cambridge, MA

John Yee, M.D., M.P.H.
Senior VP, Medical Affairs
Flexion Therapeutics
Burlington, MA

5. Using a Registry To Assess Long-Term Product Safety

Pattra Mattox, S.M.
Senior Manager
Aetion
Boston, MA

Deborah Symmons, M.D.
Professor of Rheumatology and Musculoskeletal Epidemiology (retired)
University of Manchester
Manchester, United Kingdom

Kimme Hyrich, M.D., Ph.D., F.R.C.P.C.
Professor of Epidemiology and Honorary Consultant in Rheumatology
Centre for Musculoskeletal Research
Division of Musculoskeletal and Dermatological Sciences
School of Biological Sciences
Faculty of Biology Medicine and Health
The University of Manchester
Manchester, United Kingdom

6. Expanding an Ongoing Pregnancy Registry

Jessica Albano, Ph.D., M.P.H.
Executive Director, Epidemiology & Health Outcomes
Syneos Health Clinical Solutions
Sewickley, PA

Deborah Covington, Dr.P.H.
Senior Research Leader, Real-World Evidence
Evidera
Wilmington, NC

7. Designing a Registry To Address Unique Patient Enrollment Challenges

Karen Domino, M.D., M.P.H.
Professor, Department of Anesthesiology and Pain Medicine
University of Washington
Seattle, WA

Laura Khurana, M.P.H.
Senior Scientific Advisor
ERT
Boston, MA

Karen L. Posner, Ph.D.
Research Professor, Department of Anesthesiology and Pain Medicine
University of Washington
Seattle, WA

8. Using Registries To Drive Quality Improvement in Chronic Conditions

Pattra Mattox, S.M.
Senior Manager
Aetion
Boston, MA

Peter Schmidt, Ph.D.
Senior Vice President, Chief Research & Clinical Officer
National Parkinson's Foundation
Miami, FL

9. Developing and Validating a Patient-Administered Questionnaire

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

Grannum Sant, M.D., FRCS, FACS
Formerly Vice President and Head
Oncology and Urology Medical Affairs
Sanofi-aventis U.S.
Bridgewater, NJ
Adjunct Professor of Urology, Tufts University School of Medicine
Boston, MA

10. Using Validated Measures To Collect Patient-Reported Outcomes

Kathleen M. Fox, Ph.D., M.H.S.
President
Strategic Healthcare Solutions, LLC
Aiken, SC

Susan Grandy, Ph.D., M.B.A.
Senior Director, Global Product and Portfolio Strategy,
AstraZeneca
Gaithersburg, MD

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

11. Using Recognition Measures To Develop a Dataset

Kenneth A. LaBresh, M.D.
Senior Fellow
RTI International
Waltham, MA

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

Louise Morgan, M.S.N., CPHQ
Director, International Quality Improvement
American Heart Association
Scotch Plains, NJ

Lee H. Schwamm, M.D., FAHA
Executive Vice Chairman, Department of Neurology
C. Miller Fisher Endowed Chair
Director, MGH Stroke Services
Massachusetts General Hospital
Professor of Neurology
Harvard Medical School
Boston, MA

12. Patient-Powered Registries: Developing Scalable and Reusable Infrastructure To Support Harmonized Data Collection Across Rare Diseases

Vanessa Boulanger, M.Sc.
Director of Research Programs
National Organization for Rare Disorders
Danbury, CT

Suzanne Rossov
Research Programs Manager
National Organization for Rare Disorders
Danbury, CT

Allison Seebald
Research Programs Manager
National Organization for Rare Disorders
Danbury, CT

13. Using Claims Data Along With Patient-Reported Data To Identify Patients

D. Kevin Horton, Dr.P.H., M.S.P.H.
Chief
Environmental Health Surveillance Branch
Agency for Toxic Substances & Disease Registry
Centers for Disease Control and Prevention
Atlanta, GA

Paul Mehta, M.D.
Medical Epidemiologist
Environmental Health Surveillance Branch
Agency for Toxic Substances & Disease Registry
Centers for Disease Control and Prevention
Atlanta, GA

Pattra Mattox, S.M.
Senior Manager
Aetion
Boston, MA

14. Using a Patient-Centered Study Design To Collect Informed Consent, Maximize Recruitment and Retention, and Provide Meaningful Clinical Data

Patricia D. Franklin, M.D., M.B.A., M.P.H.
Professor and Co-Director, Outcomes and Measurement Hub
Department of Medical Social Sciences
Northwestern University Feinberg School of Medicine
Senior Scientist, FORCE-TJR
Chicago, IL

Laura Khurana, M.P.H.
Senior Scientific Advisor
ERT
Boston, MA

15. Linking a Procedure-Based Registry With Claims Data To Study Long-Term Outcomes

Mark J. Fox
Privacy and Research Compliance Officer
American College of Cardiology
Washington, DC

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

16. Issues With Obtaining Informed Consent

Moira K. Kapral, M.D., M.Sc., FRCPC
Professor of Medicine
University of Toronto
Senior Scientist, ICES
Toronto, Canada

Janice A. Richards, R.N., B.A.
Formerly Project Manager
Registry of the Canadian Stroke Network
Toronto, Canada

Jack V. Tu, M.D., Ph.D.
Professor of Medicine, University of Toronto
Staff Physician, Sunnybrook Schulich Heart Centre
Senior Scientist, Institute for Clinical Evaluative Sciences
Toronto, Canada

17. Operationalizing Informed Consent for Children

Catherine Birken, M.D., M.Sc., FRCPC
Associate Professor, Department of Pediatrics
University of Toronto
Staff Paediatrician
The Hospital for Sick Children
Associate Scientist
Hospital for Sick Children Research Institute
Toronto, Canada

Laura Khurana, M.P.H.
Senior Scientific Advisor
ERT
Boston, MA

Jonathon Maguire, M.D., M.Sc., FRCPC
Associate Professor of Paediatrics
University of Toronto
Scientist, Li Ka Shing Knowledge Institute
Department of Pediatrics
St. Michael's Hospital
Toronto, Canada

Magda Nunes de Melo, B.Sc.Pharm., M.Sc.
Formerly Manager, Observational Epidemiology and Qualitative Research Unit
Applied Health Research Centre
St. Michael's Hospital
Toronto, Canada

Patricia Parkin, M.D., FRCPC
Research Director
Division of Pediatric Medicine and the Pediatric Outcomes Research Team
The Hospital for Sick Children
Toronto, Canada

18. Obtaining a Waiver of Informed Consent

Mark J. Fox
Privacy and Research Compliance Officer
American College of Cardiology
Washington, DC

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

19. Developing a Public-Private Partnership for Comparative Effectiveness Research

Elise Berliner, Ph.D.
Director, Technology Assessment Program
Center for Evidence and Practice Improvement (CEPI)
Agency for Healthcare Research and Quality
Rockville, MD

Flora Lum, M.D.
Vice President, Quality and Data Science
American Academy of Ophthalmology
San Francisco, CA

Pattra Mattox, S.M.
Senior Manager
Aetion
Boston, MA

20. Developing a Patient-Centered Governance Structure

Ben Nowell, Ph.D.
Director, Patient-Centered Research
Global Healthy Living Foundation and Patient Powered Arthritis Research Network
Nyack, NY

21. Using Registry Tools To Recruit Sites

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

22. Using a Scientific Advisory Board To Support Investigator Research Projects

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

Michael D. Taylor, Pharm.D., Ph.D.
Global Head of Epidemiology for Oncology
Genentech, A Member of the Roche Group
South San Francisco, CA

23. Understanding Baseline Characteristics of Combined Datasets Prior to Analysis

Maria Inacio, M.S.
Epidemiologist
Surgical Outcomes and Analysis
Kaiser Permanente
San Diego, CA

Laura Khurana, M.P.H.
Senior Scientific Advisor
ERT
Boston, MA

Gregory B. Maletis, M.D.
Chief of Orthopedics
Baldwin Park Medical Center
Kaiser Permanente
Baldwin Park, CA

Elizabeth Paxton, M.A.
Director, Kaiser Permanente National Implant Registries
Southern California Permanente Medical Group
San Diego, CA

24. Using Registry Data To Evaluate Outcomes by Practice

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

Jeffrey Wagener, M.D.
Professor of Pediatrics
University of Colorado Medical School
Aurora, CO

25. Using Registry Data To Study Patterns of Use and Outcomes

Marnie L. Boron, Pharm.D.
Formerly Director, Scientific Communications
MedImmune
Gaithersburg, MD

Michelle B. Leavy, M.P.H.
Head, Healthcare Research and Policy
OM1, Inc.
Boston, MA

Appendixes

Appendix A. An Illustration of Sample Size Calculations

As a general principle, sample size calculations depend on the study design, the study question, and the scale of measurement of the variables being measured. Indeed, one of the benefits of performing a sample size calculation is the requirement that each of these elements be specified, thus increasing the likelihood that the proper variables will be measured on the proper patients in the proper manner.

For concreteness, assume that the outcome of interest is a dichotomous variable measured for each patient, such as the presence/absence of a complication associated with carotid endarterectomy (CE). Typically, this literature considers complications within 30 days of the procedure. Nothing essential changes for outcome variables measured on other scales, such as continuous or survival data. The dichotomous outcome (i.e., presence or absence of a complication) is then aggregated across patients into a complication rate (e.g., 9 complications for 300 patients equals a 3-percent complication rate).

For CE, some registry-based designs and study questions that might be of interest include the following. For the purpose of this discussion, case-mix adjustment is the incorporation of various patient characteristics believed to influence complications of CE into a mathematical model used to predict the likelihood of these complications. The most natural such model is a logistic regression.

Design 1: For patients at high risk of stroke, perhaps using an operational definition of "symptomatic with 70–99 percent stenosis of the carotid artery," the study question is whether the surgeons within a larger entity (e.g., a national chain of hospitals) are, in aggregate, experiencing complication rates similar to those who participated in the randomized trials demonstrating the efficacy of CE. The reason this is an open question is that the surgeons and institutions in these randomized trials have undergone a high degree of selection, so that there is a concern that their surgical outcomes may be better than could be expected in usual practice.

The patient inclusion criteria for the registry are selected to be as close as possible to those of the randomized trials; thus, while various characteristics might be collected on each patient, no formal case-mix adjustment is required.

Further, suppose that the 30-day complication rate of CE in the randomized trials was 3 percent. The study question can then be translated into a statistical hypothesis of a one-sample comparison of an observed complication rate versus a prespecified value. In other words, the null hypothesis is that surgeons within the larger entity are, in aggregate, experiencing complication rates that are the same (3%) as those of surgeons who participated in the randomized trials. The final input required to perform the sample size calculation is the complication rate under the alternative hypothesis. For example, if it is determined that the goal of the registry is to have high power to flag results as statistically significant if the true complication rate is 6 percent or higher, then the complication rate under the alternative hypothesis is 6 percent.

In general, the value of the complication rate under the alternative hypothesis is derived using a combination of quantitative and qualitative reasoning. The precise methods used are context dependent and thus not discussed in detail here. In the present example, a cost-effectiveness analysis might suggest that complication rates of 6 percent and above would call into question the efficacy of CE. Given these inputs, it can be shown that the effect size is 0.21, and the sample size required for 80-percent power is approximately 370.

Design 2: Continuing to follow patients at high risk of stroke, now suppose that the goal of the registry is to compare complication rates across hospitals. For simplicity, we continue to assume that patients are sufficiently similar to the comparator patients that no explicit adjustment for case mix is required.

Design 2 is a simple form of benchmarking application. For example, the CE complication rates for each hospital might be reported to a regulatory agency and/or the general public, on the presumption that statistically significant differences between complication rates can be used to identify hospitals with differences in quality of care. The particular danger in this design is that the complication rate for any particular hospital might be estimated with relatively little precision, thus generating results that have more noise than signal. Another danger, discussed later, is that case-mix adjustment is required and not performed, or performed, but not adequately.

We assume that the benchmarking will focus on comparing specific hospitals—i.e., in the underlying statistical model, hospital will represent a fixed rather than random effect. The null hypothesis is that the complication rates for all the hospitals are identical, and the alternative hypothesis is that the complication rates follow some pattern other than being identical. In this design, specifying the alternative hypothesis of interest is a potentially formidable task. One way to formulate this hypothesis is to focus on outlier hospitals. For example, suppose that there are 10 hospitals in the registry, the overall complication rate among 9 of these is expected to be 3 percent, and the complication rate at the tenth hospital is 10 percent. This information, along with expected number of cases in each hospital, is sufficient to calculate an effect size and thus perform the sample size calculation.

When comparing complication rates among specific hospitals, some adjustment may be made for multiple comparisons—that is, in any group of hospitals, there will always be a hospital with the highest complication rate, and focusing on differences between the outcomes of this particular hospital versus outcomes of the others will overstate the level of statistical significance. The initial statistical test used to assess the homogeneity of complication rates across all the hospitals in the registry implicitly takes this multiple-comparison problem into account. Subsequent tests, in particular those tests that compare apparent outlier hospitals with others, should include an explicit adjustment for multiple comparisons, and the sample size calculations should reflect the fact that an adjusted comparison is being made.

In practice, the approach to this design might reasonably depend on whether registry data are being collected electronically or manually. If data are being collected electronically, the most sensible policy is to collect information on all CE procedures performed within each hospital and to use the sample size formula as an assessment of whether the registry as a whole is likely to

produce results that are sufficiently accurate to support decision making. This assessment can be framed in terms of statistical power, as discussed above, or in terms of precision.

Considering precision, a 95-percent confidence interval for a nonzero complication rate for any hospital is $p \pm 1.96 \text{ sqrt } (pq/n)$, where p is the observed complication rate, $q = 1- p$, and n is the sample size. Supposing that $p = 3$ percent and $n = 300$ per hospital, within any particular hospital, the width of this confidence interval is expected to be approximately ±1.9 percent. If data are being collected manually, and thus the marginal cost of data collection per patient is high, a reasonable policy would be to collect data on a sufficient number of patients in each hospital so that the precision of the estimates of the complication rate within a given hospital would be considered adequate.

As with hypothesis testing, the analysis to derive the width of the confidence interval usually applies a combination of qualitative and quantitative insights. In particular, the question can be reframed as the following: For what values of the complication rate will my decision (whether taken from the perspective of clinical medicine, public health, etc.) be the same? For example, if the decision is the same regardless of where the complication rate falls within the range of 2 to 4 percent, an interval of this width is sufficiently precise.

Unless sample sizes are large, using registries to compare individual hospitals is potentially quite problematic. Although determining the inputs to the power calculations is not always a straightforward task, performing this analysis is quite useful, even if the result is only to suggest extreme caution in the interpretation of differences between hospitals.

Design 3: Continuing to follow patients undergoing CE, now suppose that the goal of the registry is to compare two different versions of the surgical procedure. For simplicity, continue to assume that patients are sufficiently similar to the comparator patients that no explicit adjustment for case mix is required. The following discussion (after including an adjustment for case mix, if appropriate) also applies to comparisons of two different versions of a medical device and similar applications. The key distinctions between this design and Design 2 are that in Design 3 the primary comparison or comparisons can be stated ahead of time and the number of comparisons is relatively small, so that the issue of multiple comparisons can be ignored.

The analytic approach to this design is a logistic regression, with the input file having one record per patient. The outcome variable is the presence or absence of a complication, the categorically scaled control variable is the hospital, and the primary predictor is the categorically scaled coding of the type of surgical procedure (i.e., CE using version A vs. CE using version B). The null hypothesis is that, after accounting for any differences in hospitals, the two different versions of the procedure have identical complication rates. The alternative hypothesis is that the rates differ by a specified amount, this amount being the minimum clinically significant difference interpreted to be of concern. Power calculations proceed in the same fashion as for logistic regression with multiple predictors.

The main pitfall in this design is that patients who receive version A of the surgical procedure might differ from those who receive version B of the procedure along some dimension that has an impact on outcomes. (This pitfall is discussed in more detail under Design 4.)

In this application, the null and alternative hypotheses are sometimes structured the same way as in an equivalence trial—that is, differences in complication rates are not expected, and the goal of the study is to demonstrate that complication rates for the two versions of the surgical procedure are similar within a certain level of precision. The structure of the analysis is not fundamentally different. Indeed, sample size calculations for equivalence trials are sometimes not performed within a hypothesis-testing framework but instead by identifying a sample size of sufficient magnitude to make the confidence interval for the difference in the complication rates between the two versions of the surgical procedure a certain width. For simplicity of presentation, let us assume from now on that any equivalence-trial-type calculations can be reframed into confidence interval format, and thus need not be discussed separately.

Design 4: Continuing to follow patients at high risk of stroke, and continuing to assume that the goal of the registry is to compare two different versions of the surgical procedure, now additionally assume that this comparison will include an adjustment for case mix. Within the logistic regression paradigm, variables used to adjust for case mix are accounted for as covariates (i.e., additional predictors). Alternatively, propensity-scoring methods could be used to adjust for those variables that predict the assignment of patients to particular versions of the procedure. For concreteness, let us focus on logistic regression. In order to perform a sample size calculation for a logistic regression, the analyst must specify the predictive ability of the covariates and the odds ratio associated with the predictor of interest. (For example, version B of the procedure might increase the odds of complications by a factor of 1.5.) Once these inputs are specified, the sample size calculation is straightforward.

Both the logistic-regression and propensity-scoring approaches suffer from the fundamental drawback that they can adjust only for covariates that are observed. In particular, if there are variables that predict outcome that are unmeasured (e.g., a physician's assessment of a patient's likelihood to comply with treatment, or an assessment of "stroke in evolution" not included in the administrative database used as the source of data for the registry), then the comparison between the two versions of the surgical procedure is potentially biased. Accordingly, before proposing to use a registry to compare complication rates (e.g., across different versions of a procedure or a device) or other outcomes, it is critical to determine that the following three conditions do not all hold: (1) a patient, provider, system, or other characteristic affects the complication rate; (2) this characteristic is unmeasured within the registry; and (3) there is a reasonable likelihood that this characteristic might be differentially distributed across the different versions of the procedure or the device. If all three conditions (in epidemiologic terms, the conditions for confounding) hold, use of the registry to compare outcomes is potentially dangerous.

Critical to Designs 1–4 is the assumption that the CE complication rate is stable over time. If this is the case, it is appropriate to use the registry to estimate a single complication rate associated with version A of the procedure, estimate another single complication rate associated with version B of the procedure, and compare the rates. On the other hand, if the technology of CE (e.g., physical materials, surgical technique) is improving, then the registry should continue to monitor the performance of CE over time. Such an ongoing monitoring function seems particularly relevant for medical devices and similar applications.

Even when the associated technology is assumed stable, some registries are intended to provide ongoing assessments of outcomes. For example, in a quality assurance context, CE complication rates might be assessed at individual hospitals on an annual basis (e.g., in order to check for problems that have recently arisen). On the other hand, a registry whose purpose is to assess whether complication rates observed in randomized trials could be achieved in usual practice could be designed with a sunset provision to cease operation once this question is answered. The latter type of registry might, for example, support a coverage decision by the Centers for Medicare & Medicaid Services.

Having an ongoing monitoring function induces additional analytical complications, among others a multiple-comparisons problem. Traditional statistical power calculations are performed under the assumption that the sample size is fixed and that, unless otherwise noted, multiple comparisons are not a major issue. Sequential testing methods associated with randomized trials (where, for example, the type I error of .05 is apportioned into an early test with alpha = .001 and a subsequent test with alpha = .499) are not appropriate for this particular design, since most of these methods assume that the maximum sample size is fixed. Some methods assume that what is fixed is not the number of patients but the number of events, but these methods are also inappropriate for registry applications.

Design 5: Suppose the goal is to estimate the complication rate associated with CE at multiple time points for the foreseeable future. Control chart methodology might reasonably be applied to this class of problems. This methodology, often used in the quality assurance and quality improvement context, was originally developed for industrial applications. In this example, the null hypothesis, under which the system in question is "in control," is that the CE complication rate remains at the desired value of 3 percent throughout the entire followup period. Samples are taken at given points in time (e.g., monthly). As an example, if these monthly samples are of size 100, then the standard error is approximately 1.7 percent. The analyst then creates a control chart by plotting these monthly complication rates over time and forming channels based on the standard error. In this example, the channel extending from the point estimate to 1 standard error above the point estimate is 3 percent to 4.7 percent.

Once the basic control chart—which goes by different names depending on the scale of measurement of the outcome variable—is formed, the plot is checked for various violations of the null hypothesis of constant complication rates. The set of possible violations to be flagged as statistically significant might include (1) any observation more than 3 standard errors from the mean; (2) two of three consecutive observations more than 2 standard errors from the mean; (3) eight observations in a row that increase or decrease; and (4) eight observations in a row on one side of the mean. These rules of thumb implicitly take into account the multiple comparisons problem by requiring noteworthy departures from the null hypothesis in order to be flagged; they are also based on the observed properties of physical machines as they fall out of adjustment: suddenly breaking down and producing an extreme outlier, or gradually heating and thus producing sequentially higher readings. Complication rates of CE might or might not follow the properties of physical machines, but the decision rules from control chart methodology are at least a good place to start.

Appendix B. Copyright Law

Copyright law confers exclusive legal rights on the owner of the copyright.[1] The exclusive rights of copyright may be sold, assigned (transferred), or licensed (limited transfer of rights for use on specific terms or conditions) to others; these rights may also be waived (quit claim). Licensing ordinarily consists of a private agreement governed by contract rather than copyright law.[2]

However, the exclusive rights conferred by copyright to prepare derivative works and distribute copies of a health information registry may be limited by regulatory requirements. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule restrictions may limit data use, reuse, and disclosures or may require additional patient authorizations for subsequent research use. The conditions of institutional review board (IRB) approval under the Common Rule may also limit reuse and further disclosure of registry data. The terms of patient authorization and consent, a data use agreement, or a business associate agreement may modify the scope and nature of rights protected by copyright law. These limitations can be avoided by the use of deidentified health information, as defined by the Privacy Rule, plus information that is not subject to the Common Rule, if they suffice for the scientific or other purposes of the registry. Without resort to copyright protections, State laws may directly restrict access to registry data, as well as the use and disclosure of data from registries developed by public health agencies.

Formal copyright registration[3] with the U.S. Copyright Office is not necessary but may be desirable for registries anticipated to have commercial value. The owner of a copyright is generally the author[4] or author's employer; ownership of the copyright for a compilation is not ownership of the underlying facts or data.[5] Copyright law presumes that an employer owns the copyright in materials created by an employee within the scope of his or her employment as a "work made for hire."[6] Institutional policies and procedures frequently prescribe whether the registry developer, his or her employer, or a funding agency owns the copyright. Employee manuals often contain an employer's position on the intellectual property created by employees. Research institutions frequently reserve the right to the intellectual property produced by their employees. Intellectual property issues are explicitly negotiated in most sponsored research contracts. Authors of a joint work are co-owners of copyright in the work.[4]

Several factors determine whether the use of a registry protected by copyright for scholarship, research, or certain other purposes is within the statutory fair use limitation on copyright.[7] In general, these factors will support subsequent uses of registry data for research, even though it may be protected by copyright. In any given set of circumstances, a specific analysis of the statutory factors is necessary to determine whether use is likely to be viewed within the fair use limitation on copyright.[1]

Copyright law may provide some legal protections for compilations such as health information registries. The extent of this protection depends on the specific characteristics of the registry. In general, the concept of ownership does not comfortably apply to health information, even when limited to copyright. Nevertheless, some registry developers may want to consider adding the

legal protections of copyright to reinforce controls on access to and use of registry data. Registry developers may also encounter copyright protections on health information held by healthcare providers. Use of health information protected by copyright for research purposes may constitute fair use under copyright law.

References for Appendix B

1. Harris RK, S.S. R. Copyright protection for genetic databases. Jurimetrics Journal. 2005;45:225-50.
2. 17 U.S.C. § 106.
3. United States Copyright Office. http://www.copyright.gov. Accessed February 19, 2020.
4. 17 U.S.C. § 201(a).
5. 17 U.S.C. § 102(b). Feist Publications, Inc., v. Rural Telephone Service, Co., Inc., 499 U.S. 340 (1991).
6. 17 U.S.C. § 201(b).
7. 17 U.S.C. § §107.

U.S. Department of
Health and Human Services

Agency for Healthcare Research and Quality
5600 Fishers Lane
Rockville, MD 20857

AHRQ Pub. No. 19(20)-EHC020
August 2020
ISBN: 978-1-58763-028-6