

Understanding the Need for Non-Interventional Studies Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility

November 25, 2019

Table of Contents

Background	1
Ethical, Operational, and Resource Barriers to Randomizing, Blinding, and Controlling in the Real World	3
Ethical Considerations.....	3
Operational Considerations	4
Resource Considerations.....	5
Study Design Considerations for Evaluating the Quality and Improving the Credibility of Non-Interventional Studies Using Secondary Data.....	6
1. Study Reporting and Transparency	8
2. Choosing a Valid Study Design to Isolate Treatment Effect.....	10
3. Correctly Identifying Study Population	13
4. Ensuring Comparability Between Treatment Groups.....	14
5. Reducing Subject, Observer, or Analyst Bias.....	15
6. Addressing Measurement Bias in Subjects' Response	16
7. Identifying and Minimizing the Effects of Confounding	17
Summary and Next Steps	20
Appendix A: Workshop Participants	22
Appendix B: The Seven Characteristics of Adequate and Well-Controlled Studies (21 CFR § 314.126)	23
Appendix C: The Spectrum of Real-World Study Designs.....	26
Appendix D. Previous Duke-Margolis Work on RWD and RWE	27
References.....	28

About the Duke-Margolis Center for Health Policy

The Robert J. Margolis, MD, Center for Health Policy at Duke University is directed by Mark McClellan, MD, PhD, and brings together expertise from the Washington, DC, policy community, Duke University, and Duke Health to address the most pressing issues in health policy. The mission of Duke-Margolis is to improve health and the value of health care through practical, innovative, and evidence-based policy solutions. Duke-Margolis catalyzes Duke University's leading capabilities, including interdisciplinary academic research and capacity for education and engagement, to inform policy making and implementation for better health and health care. For more information, visit healthpolicy.duke.edu.

Acknowledgments

The Duke-Margolis Center would like to thank several individuals for their contributions to this white paper. This paper's development was guided by the expertise of the Real-World Evidence Collaborative Methods Working Group listed on [page iv](#). The Working Group's extensive subject matter expertise and insightful comments were integral to the development of this paper. We are also appreciative of the broader Real-World Evidence Collaborative Advisory Group who provided their guidance throughout development of this white paper. We would also like to thank those who participated in the Duke-Margolis private workshop, "Improving RWE Study Credibility and its Role in Totality of Evidence," held in June 2019 (a list of attendees can be found in [Appendix A](#)). In addition, we are grateful to Patricia Green and Sarah Supsiri from the Center for their communications and events guidance. Finally, we acknowledge Kerry Stenke from the Duke Clinical Research Institute for her support in developing graphics for this white paper, and her colleague, Liz Wing, for editorial assistance. Any opinions expressed in this paper are solely those of the authors and do not represent the views or policies of any other organizations external to Duke-Margolis. Funding for this work is made possible through the generosity of the Margolis Family Foundation, which provides core resources for the Center, as well as a combination of financial and in-kind contributions from Real-World Evidence Collaborative members, including AbbVie; Eli Lilly and Company; Genentech, a member of the Roche Group; GlaxoSmithKline; Johnson & Johnson; Merck & Co.; Novartis; Teva; and UCB. For more information on the Real-World Evidence Collaborative, visit <https://healthpolicy.duke.edu/real-world-evidence-collaborative>.

Disclosures

Mark B. McClellan, MD, PhD, is an independent board member on the boards of Johnson & Johnson, Cigna, Alignment Healthcare, and Seer; co-chairs the Accountable Care Learning Collaborative and the Guiding Committee for the Health Care Payment Learning and Action Network; and receives fees for serving as an advisor for Cota and MITRE.

Gregory Daniel, PhD, MPH, receives consulting fees from Genentech, Inc., the Reagan-Udall Foundation, and AbbVie Inc.

Rachel E. Sherman, MD, MPH, is an independent board member for Aptinyx and has served as an advisor to Medidata.

WHITE PAPER

Understanding the Need for Non-Interventional Studies Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility

DUKE MARGOLIS AUTHORS

Nirosha Mahendraratnam

Joy Eckert

Kerra Mercon

Katherine Frank

Morgan Romine

Adam Kroetsch

Rachel Sherman

Gregory Daniel

Mark McClellan

WORKING GROUP

Jean Marie Arduino
Merck & Co.

Shrujal Baxi
Flatiron Health

Marc Berger
International Society for
Pharmacoeconomics and
Outcomes Research

David Blaser
PatientsLikeMe

Stephanie Chiuve
AbbVie

Jennifer Christian
IQVIA

Susan Colilla
Teva Pharmaceuticals

William Crown
OptumLabs

Shannon Ferrante
GlaxoSmithKline

Kevin Haynes
HealthCore

Stacy Holdsworth
Eli Lilly and Company

Solomon Iyasu
Merck & Co.

Linda Kalilani
UCB

Ryan Kilpatrick
AbbVie

Carol Koro
Merck & Co.

Martin Ladouceur
Evidera

Grazyna Lieberman
Genentech

Nicole Mahoney
Flatiron Health

Panagiotis Mavros
Janssen

Emily O'Brien
Duke Clinical Research Institute

Sally Okun
PatientsLikeMe

Christina Parrinello
Flatiron Health

Amy Rudolph
Novartis

Patricia Saddier
Merck & Co.

Meghna Samant
Flatiron Health

Khaled Sarsour
Genentech

Kristin Sheffield
Eli Lilly and Company

Caroline Tai
Evidation

David Thompson
Syneos Health

Eileen Mack Thorley
PatientsLikeMe

Kathleen Villa
Evidation

Yiting Wang
Janssen Research and
Development

Vince Willey
HealthCore

Richard Willke
International Society for
Pharmacoeconomics and
Outcomes Research

David Wormser
F. Hoffmann-La Roche

Keele Wurst
GlaxoSmithKline

ADVISORY GROUP

Ryan Kilpatrick

AbbVie

Carlos Garner

Eli Lilly and Company

Bray Patrick-Lake

Evidation Health

Debra Schaumberg

Evidera

Nicole Mahoney

Flatiron Health

Jacqueline Law

Genentech

John Graham

GlaxoSmithKline

Richard Platt

Harvard Medical School

Marcus Wilson

HealthCore

Marc Berger

International Society for
Pharmacoeconomics and
Outcomes Research

Richard Willke

International Society for
Pharmacoeconomics and
Outcomes Research

Nancy Dreyer

IQVIA

Joanne Waldstreicher

Johnson & Johnson

Solomon Iyasu

Merck & Co.

Barbara Bierer

The Multi-Regional Clinical
Trials Center of Brigham and
Women's Hospital and Harvard

Eleanor Perfetto

National Health Council

Amy Rudolph

Novartis

Paul Bleicher

OptumLabs

Sally Okun

PatientsLikeMe

David Thompson

Syneos Health

Riad Dirani

Teva Pharmaceuticals

David Miller

UCB

Stephanie Reisinger

Veradigm Health

EXECUTIVE SUMMARY

Randomized, blinded, and controlled trials are the gold standard for establishing product effectiveness. However, stakeholders throughout the health care environment (i.e., sponsors, regulators, payers, providers, and patients) are eager to leverage real-world data (RWD) and real-world evidence (RWE) in their decision making. Essential to advancing RWE in health care decision making related to medical product effectiveness is evaluating its quality and credibility. While stakeholders desire the inclusion of randomized controlled trial (RCT) characteristics (i.e., randomization, blinding, and controlling) in the design of real-world clinical studies for evaluating treatment effects, many of the challenges to conducting RCTs are also applicable to interventional studies in the real world. Furthermore, due to significant advancements in both generating fit-for-use RWD and developing robust analysis methods, other types of real-world study designs, including non-interventional study designs, may also be able to support valid causal inference.

This paper explores the feasibility challenges to conducting interventional studies in the real world to provide guideposts for when non-interventional studies could inform regulatory decisions related to effectiveness. Specifically, ethical, operational, and resource barriers that may hinder the ability to randomize, blind, and control are discussed. The paper then highlights key considerations for how to demonstrate the credibility of a non-interventional study using secondary data that is intended to support an evidence package submitted for regulatory decision making. This includes a discussion of various methods for non-interventional studies using secondary data, the biases they intend to address, and how they can be mapped to the characteristics of an adequate and well-controlled study as outlined in the Code of Federal Regulations.

While there are different types of real-world study designs, this paper focuses on non-interventional studies using secondary data because their potential applications in support of regulatory decisions regarding effectiveness are not as well characterized as interventional RWE approaches that include randomization (e.g., pragmatic or large simple trials) or hybrid designs that utilize RWD as an external control. Non-interventional studies using secondary data can better reflect the broader patient populations, settings, and drug uses that are typical of clinical practice and, in some cases, can be conducted more efficiently than traditional RCTs. These studies should be viewed as a complementary source of evidence to RCTs that add unique and valuable information regarding the performance of medical products.

How This Paper Was Developed

This paper is informed by a literature review, a full-day private workshop on “Improving RWE Study Credibility and its Role in Totality of Evidence” (June 20, 2019), and the expert opinion of the Duke-Margolis RWE Collaborative Methods Working Group. During the workshop, stakeholder experts representing sponsors, academic research groups, data vendors, providers, and patient networks provided input for key considerations for real-world study designs and methods. This paper focuses on the quality and credibility of individual studies. The forthcoming companion methods paper focuses on the role of RWE in an evidence package, using a totality of evidence approach. This work builds on prior Duke-Margolis work, including the white papers [Determining Real-World Data’s Fitness for Use and the Role of Reliability \(2019\)](#), [Characterizing RWD Quality and Relevancy for Regulatory Purposes \(2018\)](#), and [A Framework for Regulatory Use of Real-World Evidence \(2017\)](#).

Background

Stakeholders are eager to increase the use of real-world data (RWD), or “data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources,” throughout the life-cycle of drug development, approval, and access.¹ In particular, they are interested in analyzing RWD to generate real-world evidence (RWE) about the use and potential benefits or risks of medical products that is actionable by a wide array of health care decision makers.¹ The FDA is exploring potential use of these data and evidence within regulatory decision making in keeping with Congressional mandates in the 21st Century Cures Act and 6th Prescription Drug User Fee Act (PDUFA). FDA activity in this area includes the December 2018 Framework for its Real-World Evidence Program that outlines core considerations for using RWD and RWE for regulatory decisions related to effectiveness for marketed drugs and biologics. The Framework includes a three-pronged approach that considers whether: “1) RWD are fit for use; 2) studies that use RWD can provide adequate scientific evidence to answer regulatory questions; and 3) study conduct meets regulatory requirements.”¹

In its RWE Framework, the FDA identifies three categories of potential study designs that use RWD to generate RWE (RWE studies) to support effectiveness: 1) randomized designs using RWD; 2) non-randomized, single-arm trials with external RWD controls; and 3) observational studies.¹ While it would be ideal to conduct interventional studies in the real world where one can randomize, blind, and control to support an effectiveness labeling change decision, it is not always feasible, appropriate, or necessary.

This white paper begins by exploring challenges to conducting interventional studies in the real world, including ethical, operational, and resource barriers, to provide guideposts for when non-interventional studies may be applicable for the regulatory question at hand. The paper highlights key considerations for evaluating study quality and improving the individual credibility of non-interventional studies that use secondary data that are intended to support a regulatory decision. This section includes a discussion of various methods for conducting non-interventional studies using secondary data,* the biases they intend to address, and how they can be mapped to the characteristics of adequate and

The Value of Real-World Data and Real-World Evidence

RWE studies can complement evidence from randomized controlled trials (RCTs) and contribute to a robust evidence package to support regulatory decision making. There is a well-established history of the FDA using RWE to support labeling changes related to safety; however, RWE studies might also be useful in labeling changes related to effectiveness.²

RWD is often collected by providers as part of clinical practice throughout the health system. Therefore, RWD can support analyses that better represent the broader impact of a medical product, including routine clinical care and self-care. RWD can also continuously capture the evolving standard of care, whereas RCTs capture information during a specified timeline. Drawing from RWD, RWE studies often have broader inclusion criteria than traditional RCTs, which might provide insight into the impact of a drug on patients who were not represented in the RCT. RWE studies might also capture outcomes that are more relevant to prescribers and patients. RWE might be generated more efficiently and with fewer resources, increasing the availability of information that might not otherwise be collected.

* While we focus on non-interventional studies using secondary data, the FDA RWE Framework also identifies randomized designs using RWD, non-randomized, single-arm trials with external RWD control, and other non-interventional studies (e.g., registries) as potential study designs to support effectiveness.¹

well-controlled studies as outlined in the Code of Federal Regulations (21 CFR § 314.126*). This white paper is intended to be a resource for stakeholders thinking through the suitability of non-interventional studies using secondary data for regulatory decision making, including sponsors designing non-interventional studies or regulators considering guidance in this space.

We specifically focus the mapping exercise on methods for *non-interventional studies using secondary data*[†] because their potential applications in support of regulatory decisions regarding effectiveness are not as well characterized as interventional RWE approaches that include randomization (e.g., pragmatic or large simple trials) or hybrid designs that utilize RWD as an external control. This distinction is especially important as the lack of randomization in non-interventional studies often gives rise to concern over the validity, reliability, or suitability of such observational data for regulatory purposes.

Still, the motivating factors for potentially utilizing non-interventional studies using secondary data for effectiveness labeling changes are increasingly clear.⁴ Such studies can help fill critical evidence gaps because of their ability to reflect routine medical practice, so long as there are appropriate study designs and analytic approaches to minimize the potential for bias and confounding. These gaps include the ability to generate evidence on a broader range of outcomes that may be more reflective of patients' perspectives or over a longer follow-up time compared to traditional RCTs. In some cases, non-interventional studies using secondary data may be able to answer certain research questions faster and more cost-efficiently.

FDA previously identified this area as a priority in the RWE Framework with the program item to “issue guidance about observational study designs using RWD, including whether and how these studies might provide RWE to support product effectiveness in regulatory decision-making.”¹ The forthcoming companion methods white paper focuses on the role of RWE in an evidence package, using a totality of evidence approach.

Types of Study Designs

Interventional Study: A type of clinical study in which participants are assigned to groups that receive one or more interventions/treatments (or no intervention) so that researchers can evaluate the effects of the intervention on biomedical or health-related outcomes. The assignments are determined by the study's protocol. Participants may receive diagnostic, therapeutic, or other types of interventions.

Observational Study: A type of clinical study in which participants are identified as belonging to study groups and are assessed for biomedical or health outcomes. Participants may receive diagnostic, therapeutic, or other types of interventions, but the investigator does not assign participants to specific interventions/treatments.

These definitions are from the ClinicalTrials.gov Glossary of Common Site Terms.³

* The full text of 21 CFR § 314.126 is available in [Appendix B](#).

† Secondary data are any data that were collected for another purpose but that are used to answer additional research questions.

Ethical, Operational, and Resource Barriers to Randomizing, Blinding, and Controlling in the Real World

Randomized, blinded, and controlled clinical trials are the gold standard for generating evidence about the efficacy of a medical product. Similarly, interventional RWE studies that are characterized with RCT features (e.g., randomization) are likely to be the most trusted form of RWE. **However, conducting interventional RWE studies may not always be feasible, appropriate, or necessary.**

In the next section, we will discuss feasibility challenges, including ethical, operational, and resource barriers, to randomizing, blinding, and controlling in the real world. It is important to note that many of the obstacles to conducting RCTs are applicable to interventional studies in the real world. These challenges highlight instances when non-interventional study designs that use RWD* may be best suited for generating evidence for a given research question.

Ethical Considerations

First and foremost, clinical study investigators are charged with protecting participants and ensuring that they are not exposed to unnecessary risk. When designing a study, sponsors are legally obligated to adhere to the ethical principles in the Belmont Report to conduct human research in the United States.⁶ Since interventional RWE studies share characteristics of both clinical research and clinical care, there are unique ethical questions to consider when designing a study.

Ethical Barriers — Randomized Treatment Assignment

For any prospective study design, it is unethical to implement randomization without equipoise between treatments being compared.⁷ This means if the intervention of interest is already available on the market and is known to be a better treatment option,

Why RCTs Are Considered the Gold Standard

Randomized, blinded, and controlled trials are the gold standard for establishing product effectiveness. In clinical studies, using chance to determine patients' treatment assignments is called randomization. The purpose of **randomization** is to balance the characteristics of treatment groups of a study population. Randomly assigning patients to treatment intends to reduce both observed and unobserved biases, and differences between treatment groups.

There are several ways to implement **blinding** to treatment assignment to limit the potential impact of biases (e.g., blinding the subject, treatment, provider, evaluator, and/or analyst). Blinding participants to the treatment they receive can reduce differential bias in measured outcomes, such as a placebo effect. Alternatively, blinding those conducting the study to participants' treatment can reduce the potential for ascertainment bias.⁵

Studies with a **control** arm have a group of participants who do not receive the treatment of interest (i.e., a control/comparator group) to help isolate the treatment effect of an intervention from other factors. Often a control/comparator group receives either the placebo or accepted standard of care, if appropriate. Use of a control/comparator arm allows investigators to measure and isolate the effects of an intervention by comparing event rates between two groups of participants with similar baseline characteristics who received a similar level of care except for the intervention.

* A more detailed discussion of the spectrum of real-world study designs can be found in [Appendix C](#).

then study administrators cannot randomize patients to a less effective treatment. The existence of equipoise may be undermined by existing RWD that indicate there may be a clinical benefit for a given use of a product. A challenge in the real world is determining whether equipoise exists, especially in cases with limited data.

Another important ethical consideration in both clinical research and clinical care is acquiring informed consent. For prospective RWE studies, providers may not be able to implement randomization because once a product is on the market, patients may be less willing to consent to receiving a randomized treatment. Patients may be resistant to participating in clinical care that diminishes their autonomy and decision-making power.⁸

Ethical Barriers — Blinding

There are many ethical barriers to implementing blinding in routine health care, especially in situations of trauma or life-threatening or life-changing diseases with limited treatment options. For example, patients do not have the ability to give informed consent to blinding if they are unconscious or in a desperate situation. Patients may also be unwilling to agree to receive care knowing that they or their clinician are blinded to their treatment plan. Relatedly, blinding providers to the intervention could impact the quality of care. When clinicians are blinded to the treatment that patients receive, they may be unable to manage known potential drug interactions or adverse symptoms that arise in patients.⁹

Ethical Barriers — Identifying a Control Arm

Often in rare or life-threatening diseases, it is not feasible or ethical to recruit patients to a control arm due to small patient populations with the disease or due to the lack of availability of alternative treatment. Consequently, it can be challenging to recruit an adequate number of patients to sufficiently power a study that can detect a treatment effect (e.g., patients are too healthy or too sick to join a study, or patients are unable to travel to study sites).¹⁰

For long-term extensions of RCTs, it is not ethical to keep patients in placebo control arms, so most extension trials are single-arm. Moreover, patients with diseases or conditions with limited or no known treatment available may not be willing to enter a trial or study that includes a control arm.

In some cases, the understanding of the disease course provides an implicit control for the single-arm treatment, and single-arm trials can provide substantial evidence of effectiveness — for example, in late-stage cancer in which metastases regress with treatment, or in a progressive neurodegenerative disease that stabilizes with treatment.¹¹

Operational Considerations

Medical practice in traditional clinical trials is different from routine care, reflecting

When Time is a Barrier

Time can be a barrier for recruiting enough patients to adequately power an interventional study. For example, when measuring long-term or rare outcomes, the time that it would take to recruit a population large enough to conduct a robust study is too long to add relevant, meaningful information to the body of evidence for a product. Additionally, researchers are often concerned about patients dropping out of a study or being lost to follow-up in a long-term study. Use of a control arm can extend the amount of time it takes to recruit a sufficient study population and thus can extend the time it takes to make a drug available for a given use. Any delay in availability of effective treatment options impacts patients with rare or life-threatening diseases with limited treatment options in particular. To maximize the amount of data that can be gathered in a study, sponsors may choose to conduct a single-arm study and use RWD for an external control arm.

each setting's unique goals. Typically, clinical researchers administer interventions in RCTs based on predetermined protocols to evaluate treatment benefit, while medical interventions in routine care are administered to treat how a patient feels, functions, or survives based on clinical opinion. Furthermore, most existing health care system infrastructure was not designed to conduct research, but rather to aid in patient care, facilitate workflows, and support billing. For example, electronic health records (EHRs) are designed to collect and display patient information for clinician decision making and billing, rather than assign treatment. These differences in infrastructure between RCTs and clinical care represent important operational considerations for researchers conducting real-world interventional studies.

Operational Barriers — Randomized Treatment Assignment

There are many barriers to randomization in both inpatient and outpatient care settings due to how medical products are prescribed, administered, and dispensed. Practitioners are encouraged to make clinical decisions based on the best available evidence, and random assignment of treatment runs counter to conventional medical practice. A mechanism for assigning treatment to patients based on chance does not exist in most clinical care settings. Furthermore, incorporating a process that includes a discussion of participation in research and informed consent as part of typical clinical work flow may be disruptive.

Operational Barriers — Blinding

A potential operational barrier to blinding is the ability for the patient to access the drug. In some cases, payers are not willing to reimburse, or require prior authorization for interventions that are not included in their formulary or that are prescribed off-label. Additionally, patient blinding may not be possible in circumstances where there is a cost difference between the treatment of interest and the standard of care option. Logistically, it is also difficult to blind the patient, prescriber, and dispenser to an outpatient intervention in the real world.

Resource Considerations

While there are several sources of information that regulators, patients, clinicians, and payers can use to learn about a product's safety and efficacy, such sources are of varying levels of quality and accessibility (e.g., many journal articles are behind a pay wall). FDA product labels are considered a gold standard for informing treatment decisions. They confer significant trust not only because they are publicly accessible, but also because of the rigorous process to incorporate evidence into a label.^{12,13} Typically efficacy/effectiveness information in a product label is informed by evidence generated in clinical trials. However, rigorously designed non-interventional studies using secondary data may offer complementary information that would otherwise not be accessible to decision makers.

Researchers who develop drugs (including sponsors) have finite resources including time, human capital, and financial capital. The investment of these resources has been cited as a barrier for conducting RCTs, with many also applicable to conducting interventional RWE studies. Because of this investment, the effectiveness outcomes traditionally studied in RCTs and subsequently included in the product label are those that facilitate product approval. However, there is value in studying non-traditional outcomes as well. In conjunction with traditional effectiveness outcomes, non-traditional outcomes (e.g., patient experience data or health care resource use) may aid improved clinical understanding for regulatory review as well as for provider and patient decision making. While these non-traditional outcomes may not be included in an RCT because they are not fundamental to the product's approval, they are more likely to be collected in a real-world setting, and can be studied using non-interventional study designs.

Study Design Considerations for Evaluating the Quality and Improving the Credibility of Non-Interventional Studies Using Secondary Data

Non-interventional studies using secondary data can generate valuable evidence that broadens the understanding of a marketed medical product and is highly valuable and relevant to patient and provider decision making. Significant interest exists in understanding how non-interventional studies using secondary data can inform regulatory decision making, including product labeling as it is the authoritative source of drug information.^{2,4} Yet concerns exist about the ability of these RWE studies to establish valid causal inferences in relation to a product's effectiveness because they are susceptible to systematic biases and threats to internal and external validity, all of which can diminish a study's credibility. While some of these biases may be unavoidable, there are a variety of design and statistical approaches researchers can utilize to quantify their impact on study findings.¹⁴

This section explores the second and third questions by mapping the seven characteristics of an “adequate and well-controlled study” as defined in Regulation 21 CFR § 314.126 (Table 1)* to the bias (e.g., misclassification, selection, information, and confounding) and threats to validity each characteristic attempts to address.† We also discuss non-interventional study design considerations rooted in pharmacoepidemiologic best practices that can be used to ameliorate the bias or threat to validity. Although these regulations were written for the

Critical Questions About Observational Studies in the FDA RWE Framework

The FDA RWE Framework identifies three critical questions for determining if RWE is appropriate for supporting effectiveness determinations for observational studies (or non-interventional studies using secondary data sources):

- “1. *What are the characteristics of the data (e.g., contain data on a relevant endpoint, consistency in documentation, lack of missing data) that improve the chance of a valid result?*
2. *What are the characteristics of the study design and analysis that improve the chance of a valid result?*
 - a. *Can an active comparator improve the chance of a valid result?*
 - b. *Given potential unmeasured confounders in non-randomized RWD studies, as well as potential measurement variability in RWD, is there a role for non-inferiority designs?*
3. *What sensitivity analyses and statistical diagnostics should be prespecified for observational studies using RWD to generate RWE for effectiveness?”*

Framework for FDA's Real-World Evidence Program¹

* The statutorily defined evidentiary bar for an effectiveness claim to be supported is “substantial evidence.” According to the Federal Food, Drug, and Cosmetic Act, substantial evidence is defined as “evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof.” In existing Guidance for Industry, FDA interpreted this to mean “that Congress generally intended to require at least two adequate and well-controlled studies, each convincing on its own, to establish effectiveness Nevertheless, FDA has been flexible within the limits imposed by the congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug were convincing.”

† Prior Duke-Margolis work addresses the first question FDA outlines in the critical questions about observational studies in its Framework on the role of data characterization and generating valid results ([Appendix D](#)).

context of traditional interventional studies, the core principles and methods discussed can be applied to non-interventional studies using secondary data.¹⁵ Interpretations of the characteristics and the corresponding biases can vary, particularly among different academic disciplines, but all are related to improving the ability to make a valid causal inference.

Table 1. The Seven Characteristics of Adequate and Well-Controlled Studies as Defined by 21 CFR § 314.126¹⁵

SUMMARY OF ADEQUATE AND WELL CONTROLLED CHARACTERISTICS (21 CFR § 314.126)*	
1	There is a clear statement of the objectives of the investigation and a summary of the proposed or actual methods of analysis in the protocol for the study and in the report of its results.
2	The study uses a design that permits a valid comparison with a control to provide a quantitative assessment of drug effect.
3	The method of selection of subjects provides adequate assurance that they have the disease or condition being studied, or evidence of susceptibility and exposure to the condition against which prophylaxis is directed.
4	The method of assigning patients to treatment and control groups minimizes bias and is intended to assure comparability of the groups with respect to pertinent variables. ^{†‡}
5	Adequate measures are taken to minimize bias on the part of the subjects, observers, and analysts of the data (e.g., blinding).
6	The methods of assessment of subjects' response are well-defined and reliable. The protocol for the study and the report of results should explain the variables measured, the methods of observation, and criteria used to assess response.
7	There is an analysis of the results of the study adequate to assess the effects of the drug. The report of the study should describe the results and the analytic methods used to evaluate them, including any appropriate statistical methods. [§]

The threats to valid causal inference and the study methods discussed in this section are **not comprehensive** and are meant to serve as examples. Other resources exist that seek to describe best practices for designing a credible pharmacoepidemiologic study^{14,16-24}; however, we chose to use the characteristics of adequate and well-controlled studies as a framework for discussing considerations related to the credibility of non-interventional studies using secondary data. These considerations are reflective of a review of the literature and consultation with experts from the Duke-Margolis [RWE Collaborative](#) to identify common biases and methods in non-interventional studies using secondary data. In practice, researchers are responsible for selecting data that are fit for use and study methods that are appropriate for the study questions. It is important to note that these seven characteristics are inherently connected and impacted by similar study design features and analysis techniques.

* The full text of 21 CFR § 314.126 is available in [Appendix B](#).

† Age, sex, severity of disease, duration of disease, and use of drugs or therapy other than the test drug.

‡ Ordinarily, in a concurrently controlled study, assignment is by randomization, with or without stratification.

§ The analysis should assess, among other things, the comparability of test and control groups with respect to pertinent variables, and the effects of any interim data analyses performed.

1. Study Reporting and Transparency

Due to the abundance and heterogeneity of some secondary data (e.g., claims data), several concerns exist related to the transparency and implementation of non-interventional studies that use secondary RWD. In particular, data dredging (or data fishing) is a practice of conducting several analyses of the same dataset in order to find relationships between variables that yield a desired outcome, versus conducting a hypothesis-driven study.²⁵ Data dredging may involve altering any part of the research question or study protocol, after the analysis is conducted, to support a specific result. A related issue is selectively reporting or publishing results that support the outcome of interest (i.e., cherry picking). While one permutation of all the potential methods for a given set of data yields significant results, there is always a risk that study results are due to random variation or that the results are overstated. Transparent reporting and correction for multiple testing can help mitigate these concerns.

An essential part of building study credibility is prospectively identifying the research questions that one aims to answer and specifying associated elements in a study protocol/statistical analysis plan (SAP) including study design, study population, key variables, and analysis methods.^{17,26} However, given the plethora of RWD sources and uses of RWD studies, it is important to note that not all studies using RWD need to be prespecified, but only those that are based on a hypothesis evaluating treatment effectiveness (HETE). Furthermore, while protocols and SAPs should be prespecified, it does not mean they must have a singular approach or are inflexible. For example, planned, scientifically-driven sensitivity analyses are valid approaches that provide flexibility when the prior evidence is lacking (e.g., subgroup differences). Such sensitivity analyses should be justified by and rooted in available literature and prior evidence whenever possible.

Adequate and Well-Controlled Characteristic

1

*“There is a **clear statement** of the objectives of the investigation and a summary of the proposed or actual methods of analysis in the protocol for the study and in the report of its results. In addition, the protocol should contain a description of the proposed methods of analysis, and the study report should contain a **description of the methods of analysis** ultimately used. If the protocol does not contain a description of the proposed methods of analysis, the study report should describe how the methods used were selected.”*

21 CFR § 314.126

Categories of RWD Studies on Treatment Effectiveness

Exploratory Treatment Effectiveness Studies

- Usually do not hypothesize the presence of a specific treatment effect and/or its magnitude
- Can serve as a first step to learn about possible treatment effectiveness (hypothesis generating)
- Are typically less preplanned to allow for adjustments as investigators gain understanding of the data

Hypothesis Evaluating Treatment Effectiveness (HETE) Studies

- Test a specific hypothesis in a specific population
- Evaluate the presence or absence of a prespecified effect and/or its magnitude
- When evaluated in conjunction with other evidence, the results may provide insights that may inform treatment recommendations (e.g., *whether a treatment effect in a study that takes place in the real world where adherence is lower is the same as in a RCT*)

Adapted from Berger et al.¹⁸

To facilitate an interactive and efficient application process, FDA reviews study protocols as well as SAPs for new drug applications **prior to study initiation**.^{*} Without the submission of protocols/SAPs to FDA, they cannot be certain that these documents were pre-specified and unchanged during data selection and analyses. A similar level of prespecification and review would apply to non-interventional studies using secondary data that are intended to support a labeling change related to effectiveness.

Lack of public transparency can present a barrier to RWE study credibility and reproducibility.^{26,28} Researchers conducting any type of human health or behavioral study are able to publicly register with [ClinicalTrials.gov](https://clinicaltrials.gov), which is publicly searchable but designed for RCT registration.²⁹ For some studies, such as RCTs, sponsors are required to register and report summary trial results on the website, but there is no requirement for investigators to register non-interventional studies using secondary data.³⁰ The Real-World Evidence Transparency Partnership, led by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), the International Society for Pharmacoepidemiology (ISPE), Duke-Margolis, and the National Pharmaceutical Council (NPC), focuses on the need for a common public registration site for these real-world studies that could be used by researchers and viewed as a credible registry by regulatory, payer/HTA, and clinical decision makers.³¹ To encourage non-interventional studies using secondary data registration, it could be a requirement for peer-reviewed publication, as it frequently is for RCTs.

In addition to protocol registration, there are a number of best practices that can help researchers maintain transparency and enable replication of the investigation — all of which increase the credibility of a study. Research bodies including ISPOR, ISPE, and the Agency for Healthcare Research and Quality (AHRQ) have released a series of white papers that outline key principles and guidelines for improving the transparency and replicability of observational studies.^{16,18,27} Recommendations include posting the study protocols and SAPs on a publicly accessible study registration site prior to implementing the study. Researchers are also encouraged to provide design diagrams as well as causal directed acyclic graphs (DAGs). Design diagrams provide a visual depiction of the anticipated timeline and should include benchmarks such as when and how patients will enter the cohort (i.e., temporal anchors), how baseline characteristics are defined, and when follow-up will begin and end.³² DAGs provide “a graphical representation of causal effects between variables to help understand whether bias is potentially reduced or increased when conditioning on covariates.”³³

^{*} The FDA notes that “the fundamental principle of investigational new drug (IND) review is that an interactive process between sponsors and Center for Drug Evaluation and Research (CDER) facilitates efficient and thorough development that increases the likelihood of submission of a complete marketing application, or alternatively, prompts early termination of a development program for an unsafe or ineffective drug.”²⁷

2. Choosing a Valid Study Design to Isolate Treatment Effect

The second characteristic of adequate and well-controlled studies addresses bias that can occur when the effect of interest is masked or incorrectly attributed to the drug. Two important study design features for isolating the effect of the drug of interest from other factors (e.g., natural history of the disease, patient or provider characteristics, or time) for non-interventional studies using secondary data are the inclusion of an appropriate comparison/control group and time/temporal anchors.^{32,34}

In non-interventional studies using secondary data, investigators should choose the most appropriate type of control group to answer their research questions.³⁵ One option for a control group is to compare patients who received the intervention to patients who received either no treatment or a placebo treatment, though placebo treatment is less common in secondary RWD. If researchers use a no-treatment group, it is important to ensure that the patients selected actually have the disease and to consider that patient treatment status may depend on factors such as disease severity, overall health status, or age. In dose-comparison studies, researchers compare the effects of the drug of interest at different doses to understand if a higher or lower dose, or if fewer or more doses, impacts the effectiveness of the intervention. Dose-comparison studies can be useful for refining the dose and administration frequency, but they are susceptible to bias from correlation between dose and disease severity. Some studies compare the effect of the drug to a

Adequate and Well-Controlled Characteristic

2

*“The study uses a **design** that permits a **valid comparison with a control** to provide a quantitative assessment of drug effect. The protocol for the study and report of results should describe the study design precisely; for example, duration of treatment periods, whether treatments are parallel, sequential, or crossover, and whether the sample size is predetermined or based upon some interim analysis.”**

21 CFR § 314.126

* Continuation of above text.

(i) Placebo concurrent control. The test drug is compared with an inactive preparation designed to resemble the test drug as far as possible. A placebo-controlled study may include additional treatment groups, such as an active treatment control or a dose-comparison control, and usually includes randomization and blinding of patients or investigators, or both.

(ii) Dose-comparison concurrent control. At least two doses of the drug are compared. A dose-comparison study may include additional treatment groups, such as placebo control or active control. Dose-comparison trials usually include randomization and blinding of patients or investigators, or both.

(iii) No treatment concurrent control. Where objective measurements of effectiveness are available and placebo effect is negligible, the test drug is compared with no treatment. No treatment concurrent control trials usually include randomization.

(iv) Active treatment concurrent control. The test drug is compared with known effective therapy; for example, where the condition treated is such that administration of placebo or no treatment would be contrary to the interest of the patient. An active treatment study may include additional treatment groups, however, such as a placebo control or a dose-comparison control. Active treatment trials usually include randomization and blinding of patients or investigators, or both. If the intent of the trial is to show similarity of the test and control drugs, the report of the study should assess the ability of the study to have detected a difference between treatments. Similarity of test drug and active control can mean either that both drugs were effective or that neither was effective. The analysis of the study should explain why the drugs should be considered effective in the study, for example, by reference to results in previous placebo-controlled studies of the active control drug.

(v) Historical control. The results of treatment with the test drug are compared with experience historically derived from the adequately documented natural history of the disease or condition, or from the results of active treatment, in comparable patients or populations. Because historical control populations usually cannot be as well assessed with respect to pertinent variables as can concurrent control populations, historical control designs are usually reserved for special circumstances. Examples include studies of diseases with high and predictable mortality (for example, certain malignancies) and studies in which the effect of the drug is self-evident (general anesthetics, drug metabolism).

group of patients who received a different active treatment that is known to be effective (often standard of care). Investigators should use the research question to determine the most appropriate type of control group for non-interventional studies using secondary data, and if appropriate, methods to define a comparison group (further described in the fourth characteristic below). Regardless, the control group should have as close to the same characteristics as the intervention group as possible. Additionally, it is important to calculate the number of patients necessary to adequately power the study, a priori.

The roles of time and temporality are especially important in designing non-interventional studies using secondary data to evaluate a drug effect.³² One consideration is that exposure time of an intervention may occur as a single event, but often administered or taken over an extended or indefinite period of time, and may require a more complex definition of exposure to treatment.³⁶ Treatment effects for a given intervention often differ based on phase or duration of exposure. Some treatments have known adverse effects associated with initiation, and in other cases, patients need to take a drug for a period of time for it to reach steady-state and become effective. Additionally, prescribers often adjust dosages, which may not be captured, or switch treatments altogether based on patients' responses. If patients are switching treatments, washout periods may be used to allow the effects of a previous treatment to be eliminated from their systems.³⁷ The appropriate window of observation for a study will depend on existing evidence and the specific study question. Researchers are responsible for defining the dose and duration of treatment for study inclusion and for considering the effects of time-varying exposures and potential confounders.³⁶

Non-interventional studies using secondary data are also susceptible to reverse causality, which can occur when researchers unintentionally reverse the exposure-disease pathway. In other words, the outcome causes the exposure or risk factor under investigation, which may lead one to draw false conclusions about the true effect of the treatment of interest. This can be avoided by confirming that follow-up for study outcomes begins only after a treatment has been initiated, which ensures the correct temporal ordering of exposure and outcomes is ascertained. Inclusion of a lag period is another technique for minimizing the potential for reverse causation.³⁸

Another time/temporal threat is immortal time bias or a period of time during the cohort's observation in which the outcome could not have occurred, often over the period before exposure to the treatment.³⁹ In some cohort studies, to be classified as exposed, the subject has to remain event-free until the exposure has been initiated, biasing the exposure group because they have survived long enough to receive the treatment.⁴⁰ For example, analyses of post-hospitalization outcomes often classify treatment groups using discharge prescription status, which is only relevant for patients who survive hospitalization. Immortal time bias can be addressed with techniques such as a time-dependent analysis or a matched, nested case-control analysis.^{41,42}

Depletion of susceptibles is yet another time/temporal challenge in non-interventional studies using secondary data that rely on a time-to-event analysis to evaluate treatment efficacy. Selection bias can be introduced into a study when patients who remain on a treatment are those who can tolerate it, while those who are at most risk of experiencing an adverse event discontinue soon after treatment initiation.⁴³ Another manifestation of depletion of susceptibles occurs when studies with patients in one treatment group are established on treatment while patients in the other group are mostly new to treatment.⁴⁴ Some patients who are established on treatment may have experienced outcomes of interest before data were collected. In both cases, there are groups of patients who are less susceptible to the outcome, and this bias can overestimate the benefits and underestimate harms. Depletion of susceptibles can be addressed with the use of incident-treatment analysis rather than

prevalent-treatment to ensure that the intervention preceded the outcome of interest.²¹ Time-dependent propensity scores can be used to adjust hazard ratios as another method to address depletion of susceptibles.⁴⁵

New-user design can mitigate some of the study design issues associated with variable misclassification due to time/temporality. This approach includes only patients from the time of treatment initiation, so that administrators can better assess patients' pretreatment characteristics and capture all events occurring anytime during follow-up. New-user design allows for assessment of time-dependent drug effects and adverse events that are associated with treatment duration. It can be used to detect effects that occur with initiation of a treatment.⁴⁶ In using secondary data, investigators need to apply criteria to ensure they are capturing new users of the medication, so the time window to define prior use is a key consideration. New-user design, when used appropriately, can enable a non-interventional design to more closely approximate an RCT and may improve the credibility of study findings.⁴⁷

Considerations for Non-Inferiority Studies

Active comparator RCTs evaluating **efficacy** are usually categorized as superiority, equivalence, or non-inferiority trials.* The main reason for using non-inferiority trials for drug research is ethics. To maintain equipoise in clinical studies, researchers often are not able to compare investigative treatments to placebo and when an active comparator is available. If treatment effects are expected to be similar to the active comparator or there may be potential benefits beyond treatment benefits, researchers may use a non-inferiority trial design. In guidance, FDA notes that "the goal [of non-inferiority trials] is to demonstrate that the test drug has an effect by showing that its effect is sufficiently close to the effect of an active control. [... Specifically,] that the effect of the test drug is not inferior to the effect of the active control by a specified amount, called the non-inferiority margin."⁴⁸ In practice, non-inferiority trials have been conducted for regulatory purposes while comparative effectiveness studies using RWD have been conducted for non-regulatory decision makers (e.g., prescribers and payers). Studies in the real world that "compare the **effectiveness** of two or more interventions or approaches to health care, examining their risk and benefits" are called comparative effectiveness research.⁴⁹

Some of the challenges of implementing a valid, high-quality non-inferiority RCT apply to

non-inferiority, non-interventional studies using secondary data (e.g., difficulties determining appropriate margin sizes and no single analysis approach).⁵⁰ However, there are unique considerations. For example, the need for a large sample size is often cited as a barrier to conducting non-inferiority RCTs, which is less of a challenge in non-interventional studies using secondary data. However, those data will lack randomization and have added heterogeneity (e.g., patient population, levels of treatment exposure, outcome assessment). Consequently, the data may be biased and inadvertently support the non-inferiority claim when in fact the new treatment is inferior. (Having a larger sample size will not address potential bias and only provide a more precise, biased effect.) Also, real-world studies may not be well-suited for reliably detecting small, but clinically meaningful treatment effects because of data heterogeneity. However, non-inferiority designs could be useful when expected treatment effects are small compared to existing treatment. Like in non-inferiority RCTs, it is critical for researchers to validate non-inferiority margins, to account for variability in measurements, and potential unmeasured confounding to generate credible evidence about non-inferiority from non-interventional studies using secondary data.⁵⁰

* Superiority trials test the hypothesis that one treatment is superior to another.⁵¹ Equivalence trials test the hypothesis that the effects of two treatments are the same (within a prespecified range). Non-inferiority trials test the hypothesis that one treatment is no worse than another treatment.^{51,52}

3. Correctly Identifying Study Population

The third characteristic of adequate and well-controlled studies focuses on reducing misclassification that can occur when the study population is incorrectly defined or identified. Fundamental to identifying the appropriate study population (and other study variables including exposure and covariables) is using a fit-for-use dataset that is both relevant and reliable. Because secondary RWD were collected for a different purpose, the information necessary to identify patients may not be available or captured at the ideal level of granularity, a challenge for the fit-for-use assessment. Subsequently, using proxies and algorithms based on the available data for classifying patients' disease or treatment statuses may be necessary.

Adequate and Well-Controlled Characteristic

3

*"The **method of selection of subjects** provides adequate assurance that they have the disease or condition being studied, or evidence of **susceptibility and exposure to the condition** against which prophylaxis is directed."*

21 CFR § 314.126

Researchers are responsible for identifying patients with a given condition who received the treatment(s) of interest. The literature should be reviewed to evaluate whether existing validated algorithms can be used to identify patients' disease and treatment and can be appropriately applied to the data that will be used.⁵³ It is critical that proxies and algorithms account for how the RWD were collected and the concept it represents. RWD are often entered or collected by a multitude of end users, and there is a potential for patient diagnoses to be incorrect or vary by clinicians, facilities, or health systems. For example, different clinicians evaluating the same patient may enter different primary diagnoses. If no appropriate algorithm exists, a custom-developed algorithm should be validated against a gold standard to ensure accuracy, sensitivity/specificity, and positive predictive value.

It is important to note, in non-interventional studies using secondary data, information about patients' diagnoses/disease status and treatment status may come from multiple data sources (e.g., a combination of registry, EHR, or claims data). Regardless of source, researchers must have reasonable assurance that patients included in a study have the disease or condition and received the treatment(s) of interest.

4. Ensuring Comparability Between Treatment Groups

The fourth characteristic is related to how patients are assigned to treatment versus control/comparator group and minimizing selection bias. As discussed in the second characteristic, it is necessary to ensure balanced covariates (both known and unknown) across both groups to help isolate the effect of the drug. While this is ideally done through randomization, there are methods to reduce the selection bias that can occur when treatment cannot be assigned as in non-interventional studies using secondary data.¹⁴

Instead, in non-interventional studies using secondary data, **patients are grouped into treatment categories based on the study design and rules that are applied.** As discussed in the third characteristic, data fitness-for-use is essential for identifying all relevant variables of interest including treatment exposure. Additionally, the clinical process of treatment selection and patient use is not always clear. For example, medications with several or broad indications or with known off-label uses may not be feasibly matched to the disease or condition they were intended to treat without additional diagnostic data. Furthermore, it can be challenging to know whether patients actually took a drug that was prescribed and dispensed to them based purely on EHR data. Using claims data, researchers can examine if a patient filled a prescription, but additional data sources may be required to assess whether a patient took the drug regularly or at all. It is at the discretion of researchers to determine the most appropriate inclusion criteria for a given study question.

Selection bias can be introduced in a non-interventional study using secondary data if the baseline characteristics of the group of patients who received the intervention are significantly different from the comparator group, and both groups do not represent the target population. This can occur for several reasons, including but not limited to:

- Providers may prescribe interventions differently based on the overall condition or health status of patients (i.e., level of disease severity or progression or comorbid conditions)
- Providers may be hesitant to prescribe newer interventions or may only be willing to prescribe certain treatments as a second- or third-line option
- Patient access to newer treatments may be limited by the expense of a treatment and/or health insurance status, leading to channeling of higher socioeconomic status receiving therapy

Researchers should be aware of potential selection bias during the design and analysis phases of studies, and should generate well-formed hypotheses and develop plausible causal diagrams (e.g., directed acyclic graphics) during the design phase.³³ If patients have different known baseline characteristics, data analysts can adjust for these differences using regression. However, prior to the analysis phase, consideration may be given to restriction based on prescribing patterns and differences in characteristics.

Adequate and Well-Controlled Characteristic

4

“The method of assigning patients to treatment and control groups minimizes bias and is intended to assure comparability of the groups with respect to pertinent variables such as age, sex, severity of disease, duration of disease, and use of drugs or therapy other than the test drug. The protocol for the study and the report of its results should describe how subjects were assigned to groups. Ordinarily, in a concurrently controlled study, assignment is by randomization, with or without stratification.”

21 CFR § 314.126

It is important for data collectors and aggregators to capture information about subjects' baseline characteristics whenever possible to detect and adjust for these differences.

Some methods that adjust for baseline differences also address confounding, which is discussed in the section below on the seventh characteristic of adequate and well-controlled studies. One method to control for unmeasured confounding is instrumental variable analysis, which aims to mimic randomization.⁵⁴ A valid instrumental variable predicts the treatment but is not associated with the outcome, though identifying a valid instrument can be challenging.

5. Reducing Subject, Observer, or Analyst Bias

The fifth component of adequate and well-controlled studies aims to mitigate information bias that can occur when study participants, observers, or data analysts change their behavior, either consciously or unconsciously, based on knowledge of a subject's treatment status or another factor. For example, more intensive follow-up for the new treatment of interest versus the known standard of care (ascertainment bias). Often, blinding and standardized methods of data collection are used in RCTs to reduce some types of information bias.

Adequate and Well-Controlled Characteristic

5

"Adequate measures are taken to minimize bias on the part of the subjects, observers, and analysts of the data. The protocol and report of the study should describe the procedures used to accomplish this, such as blinding."

21 CFR § 314.126

Blinding participants in non-interventional studies using secondary data is often not feasible. In retrospective studies using secondary data, blinding of patients and providers is outside the control of study investigators, since patients and prescribers make treatment decisions together. One advantage of retrospective studies is that participants do not know that they are being studied for a particular research question when data are collected, so there is reduced potential for responder bias. However, study analysts can be masked to patient status for some non-interventional studies.

Prespecification of the study protocol and statistical analysis plan and blinding of data analysts are important techniques for minimizing information bias on the part of data analysts. Blinding data analysts can reduce potential for analyst bias, but it also restricts the ability of analysts to perform unplanned tests or study modifications based on preliminary results. Data aggregators and preparers should use detailed, prespecified methods for selecting and assessing eligible patients. Another technique for minimizing information bias is for investigators to choose endpoints that are less subject to different interpretations. For mitigating measurement bias, blinding is unnecessary for some objective endpoints, such as death, recurrent myocardial infarction, or hospitalized infection.

6. Addressing Measurement Bias in Subjects' Response

Accurate and reliable methods of assessment are part of the sixth characteristic of adequate and well-controlled studies. In RCTs, study administrators strive to collect measurements that reflect the clinical concept of interest from subjects with as little variability as possible to reduce potential measurement bias. Because RWD are collected in routine care rather than in a controlled, experimental environment, there is added potential for measurement bias that could impact fitness-for-use of data. Similar to concerns raised in the third characteristic about misclassification of a patient's disease or exposure status, a patient's outcome status can be measured incorrectly or misclassified in secondary data, depending on who is collecting the information.

Adequate and Well-Controlled Characteristic

6

"The methods of assessment of subjects' response are well-defined and reliable.

The protocol for the study and the report of results should explain the variables measured, the methods of observation, and criteria used to assess response."

21 CFR § 314.126

In the real world, encounters with the health care system are not protocolized, so there are often gaps in clinical data over time. Equipment and practices in hospitals and other clinical settings vary widely, and it is not always feasible for data aggregators or study sponsors to control or adjust for these differences in RWD. Diagnosis or disease classification often depends on the interpretation of the provider reading results or examining a patient. For example, in oncology, when providers are determining disease progression or a patient's response to treatment, providers may reach different conclusions depending on the method of measurement (e.g., radiologic measures or pathologic measures). Differential variation in measurements can introduce measurement bias into study results. As mentioned in several of the earlier characteristics of adequate and well-controlled studies, it is critical for researchers to use data that are rigorously collected, aggregated, and validated as fit for use. It is important for researchers to specify how outcomes will be defined and what algorithms will be used.

If investigators have a series of data points for each study subject, analyses of within-subject differences for a given measurement can help analysts quantify the level of measurement bias. Analysts can also examine differences in measurements between clinical sites. If data analysts can describe the size and direction of measurement bias, it may be possible to adjust for variability in measurements. Structural equation modeling (SEM) is one option for estimating measurement bias.⁵⁵ In some cases, when little is known about variability in measurements, it may be appropriate for investigators to cluster patients by clinical care site, assuming that measurements are conducted similarly within clinical care sites. This approach allows measurement variations across clinical sites to be quantified and accounted for in effect size estimation using methods such as hierarchical modeling. Additionally, investigators can conduct sensitivity analyses that offer a range of estimates based on potential size of misclassification bias and the impact it could have on overall estimates of effect, thereby offering an evaluation of the robustness of estimates.

7. Identifying and Minimizing the Effects of Confounding

The seventh characteristic of adequate and well-controlled studies addresses measured and unmeasured confounding in a study. In a clinical study, confounding occurs when a factor outside the causal pathway is associated with the exposure and is a risk factor for the outcome variable.⁵⁶ The presence of confounding can make it appear as though an association exists when it does not, or mask the true association between a treatment and an outcome.

An important distinction is whether confounding variables are assumed to be observed or whether some are unobserved. When all confounders are observed, researchers can use analytical approaches that are similar to those that control for baseline differences between treatment groups. When some are unobserved, approaches must rely on relationships between the unmeasured confounders and observed variables, on designs or analyses that proxy randomization, on external information, or on other assumptions about the nature of the confounders.⁵⁷ However, by the very nature of the problem, it is usually difficult to be certain about whether unmeasured confounders are well controlled.

The concept of confounding is defined and addressed in different ways across disciplines (e.g., endogeneity in econometrics).^{58,59} Regardless, the issue is the same in that it is a barrier to making a valid causal inference. Additionally, there are several techniques that are commonly used to address confounding throughout the study design and analysis phases ([Table 2](#)).⁵⁷ Confounding is related to other characteristics of adequate and well-controlled studies and, in some cases, can be avoided or addressed in part using methods previously discussed. The techniques listed in [Table 2](#) are meant to serve as examples rather than an exhaustive list. Additional techniques for addressing confounding are discussed in detail in the National Institute for Health and Care Excellence (NICE) technical support document on observational data.¹⁴

Adequate and Well-Controlled Characteristic

7

*“There is an **analysis of the results of the study adequate to assess the effects of the drug.** The report of the study should describe the results and the analytic methods used to evaluate them, including any appropriate statistical methods. The analysis should assess, among other things, the comparability of test and control groups with respect to pertinent variables, and the effects of any interim data analyses performed.”*

21 CFR § 314.126

Table 2. Methods for Addressing Confounding in the Design and Analysis Phases

	DESIGN TOOLS FOR ADDRESSING CONFOUNDING	ANALYSIS TOOLS FOR ADDRESSING CONFOUNDING
APPROACHES FOR MEASURED CONFOUNDING	<ul style="list-style-type: none"> • Restricting⁶⁰ — Limits study subjects to those with a specific value of the confounding variable • Matching — Groups subjects in different exposure cohorts by the same value of the confounding variable(s) such as age, income, disease severity, or comorbidity 	<ul style="list-style-type: none"> • Stratification⁶¹ — Divides subjects into strata that share a common characteristic or value and analyzes them separately • Multivariate regression models⁶² — Accounts for more than one covariate in a model to adjust for confounders (e.g., logistic and linear regression models, Cox proportional hazards model) • Propensity score analysis⁶³⁻⁶⁵ — Summarizes the information from measured confounding variables by estimating the probability of a subject having the exposure of interest given their clinical status and is able to account for many confounders at once*
APPROACHES FOR UNMEASURED CONFOUNDING	<ul style="list-style-type: none"> • Two-stage sampling⁶⁶ — Exposure and outcome variables are determined for the entire sample, but covariates are measured on a subsample • Crossover design⁶⁷ — Assesses the association of transient exposures and acute outcomes by using subjects as their own control • Active comparison group^{46,68} — Compares the effects of a treatment to a control group that was exposed to another treatment for the disease (based on the assumption that patients who receive either treatment have a similar level of disease severity) • Natural experiment⁶⁹ — Compares two groups where treatment assignment is controlled by an independent factor not related to outcomes, such as a difference in laws between states 	<ul style="list-style-type: none"> • Hierarchical models — Able to model multilevel data (e.g., fixed effects and random effects) • Instrumental variable analysis⁵⁴ — Estimates the causal effect in the presence of unmeasured confounding using an unconfounded proxy of the exposure • Sensitivity analysis⁵⁷ — Quantifies potential unmeasured confounding and its effect on the exposure-outcome association • External adjustment⁵⁷ — Uses parameter estimates for an unmeasured confounder(s) from existing research • Regression discontinuity⁶⁹ — A quasi-experimental pretest-posttest design analysis that elicits the causal effects of interventions by assigning/recognizing a cutoff or threshold above or below which an intervention is assigned

(Table 2 adapted from Schneeweiss^{57,70})

* There are multiple methods for propensity score analysis, such as matching, stratification, and inverse probability of treatment weighting.

Uncertainty from unmeasured confounding is a threat to the credibility of non-interventional studies using secondary data. Common unmeasured confounders in secondary datasets include lifestyle, socioeconomic status, clinical factors, and over-the-counter medications.⁵⁷ When successfully randomized, unknown and unmeasured variables are balanced across the intervention and control/comparator groups. When randomization is not possible, one method for detecting the presence of unmeasured confounding is the use of negative controls, which repeats the analysis under conditions in which a null result is expected to verify that it does in fact produce a null result.⁷¹⁻⁷³ Furthermore, to assess the validity and robustness of results, sensitivity analyses can be used to measure the potential impact of unmeasured confounders.

“[Sensitivity analyses are] the last line of defense against biases after every effort has been made to eliminate, reduce, or control them in study design, data collection, and data analysis.”^{57,74}

Sensitivity analyses enable researchers to parameterize a spectrum of plausible effect estimates by varying assumptions about the underlying clinical practice or biology of the treatment.⁵⁷ Sensitivity parameters can be applied to any aspect of a study design (e.g., patient selection, treatment exposure, known and unknown covariates, and outcome) and should be hypothesis-driven, rely on causal diagrams, and prespecified when possible. Common simple analyses include varying the look-back period to assess covariates and new user status prior to treatment initiation as well as varying the follow-up periods such as changing the lag and latency periods to evaluate the impact of treatment initiation and discontinuation on findings. There are also several approaches to conducting sensitivity analyses for unmeasured confounders, and some examples of these approaches are discussed in [Table 3](#).^{57,73}

Table 3. Approaches to Sensitivity Analyses

SENSITIVITY ANALYSIS / EXTERNAL ADJUSTMENT APPROACH	DESCRIPTION
Array approach (basic)	Analyses exploring how observed associations change by varying 1) the strength of the confounder-disease association and 2) the balance of confounders across treatment groups
Rule-out approach	Analysis to determine the strength of unmeasured confounding needed to fully explain the results of a study with the hope that the number of possible unmeasured confounders cannot explain the observed association (e.g., E-value ^{75,76})
External adjustment based on a single binary confounder	Analysis that incorporates prevalence and parameter estimates for an unmeasured confounder from existing research to estimate the impact of confounding
External adjustment based on multiple confounders with various distributions	Analysis that uses propensity score calibration to account for multiple unmeasured confounders rather than outcome information
Simulation-based sensitivity analysis	Computerized analysis technique that provides a distribution of bias-corrected estimates (e.g., Monte Carlo)

(Table 3 adapted from Schneeweiss⁵⁷)

Summary and Next Steps

Non-interventional studies using secondary data can better reflect broader patient populations, care settings, and drug uses found in routine clinical practice. Additionally, in some cases, they can be conducted more efficiently than interventional studies. While non-interventional studies using secondary data offer unique and valuable evidence that complements RCTs and randomized real-world studies, questions remain about the circumstances that merit their use and their ability to support valid causal inferences. Building on a body of work on using non-randomized real-world studies for regulatory decision making and the pharmacoepidemiologic literature on study design and analysis, we explored some of these remaining questions.^{4,14,17,22}

First, we discussed feasibility challenges to conducting interventional studies in the real world to illustrate when non-interventional studies could inform regulatory decisions related to effectiveness. Specifically, we explored the ethical, operational, and resource barriers to randomization, blinding, and controlling studies in the real world. Second, we discussed considerations for designing and analyzing high-quality non-interventional studies using secondary data that can make valid causal inferences to support regulatory decision making. We mapped rigorous pharmacoepidemiologic methods for non-interventional studies using secondary data and the biases they intend to address to the regulatory standard of adequate and well-controlled studies.

Significant progress in the field of pharmacoepidemiology has been made in the past 20 years, and it continues to evolve rapidly.^{4,77} There is ongoing work in the area of developing and refining advanced statistical methods to improve researchers' ability to make a valid causal inference using RWD, such as G-methods and techniques for harmonizing observational and RCT protocols.^{22,78-81} Over time, it is imperative that an understanding of these methods and their potential limitations is continuously harnessed by reviewers to adequately evaluate new evidence.

Demonstrating consistency in the ability of these pharmacoepidemiologic techniques to make causal inference would enable FDA to incorporate this valuable source of information into its decision making.^{4,14} One approach to building this confidence is through prospective replication/duplication studies that demonstrate that the results of studies using RWD match those from RCTs, such as RCT Duplicate and OPERAND.^{4,82,83} Although not all RCTs can be replicated using RWD, RWE can offer new information that complements evidence from RCTs.⁸⁴ Other health care decision makers (e.g., payers and providers) have significant experience in using RWD and these analytical techniques to generate RWE to inform their decision making. Their experience may also offer the Agency examples of how to evaluate the quality of evidence these techniques yield.

This paper explores the ability of non-interventional studies using secondary data to make valid causal inferences by mapping study designs and analysis techniques to the characteristics of adequate and well-controlled studies. However, the adequate and well-controlled characteristics is but one framework to use to guide research on the quality of non-interventional studies using secondary data. As there is no single gold standard for its evaluation, it is important for stakeholders to develop consensus around understanding when non-interventional studies using secondary data are sufficiently credible to be used in regulatory decision making. Such consensus is the first step to distinguishing between high-quality and low-quality RWE.

Future topics related to RWE study credibility may include:

- Continuing to demonstrate that RWE studies (i.e., randomized real-world studies, observational studies, and the use of external control arms) can be used to make valid causal inference to support regulatory decision making (e.g., pilot projects);
- Providing guidance on how RWE can effectively be used as part of evidence packages to support decisions related to effectiveness labeling changes; and
- Providing clarity on how RWE can be communicated in the product label.

APPENDIX A: WORKSHOP PARTICIPANTS

Improving RWE Study Credibility and its Role in Totality of Evidence

Duke-Robert J. Margolis, MD, Center for Health Policy
1201 Pennsylvania Ave, NW, Suite 500 • Washington, DC 20004
June 20, 2019

Participant List

Elodie Baumfeld Andre
Pfizer

Stella Chang
Veradigm

Stephanie Chiuve
AbbVie

Jennifer Christian
IQVIA

William Crown
OptumLabs

Ruthie Davi
Acorn AI, a Medidata company

Kevin Haynes
HealthCore

Stacy Holdsworth
Eli Lilly and Company

Albert Kraus
Pfizer

Martin Ladouceur
Evidera

Grazyna Lieberman
Genentech

Nicole Mahoney
Flatiron Health

Panagiotis Mavros
Janssen

Brian Mayhew
Novartis

David Miller
UCB

Marko Mychaskiw
Teva Pharmaceuticals

Emily O'Brien
Duke Clinical Research Institute

Sally Okun
PatientsLikeMe

Christina Parrinello
Flatiron Health

Andrew Robertson
Sanofi

Khaled Sarsour
Genentech

Kristin Sheffield
Eli Lilly and Company

Caroline Tai
Evidation

David Thompson
Syneos Health

Yiting Wang
Janssen Research and
Development

Richard Willke
International Society for
Pharmacoeconomics and
Outcomes Research

David Wormser
F. Hoffmann-La Roche

Keele Wurst
GlaxoSmithKline

*Duke-Robert J. Margolis, MD,
Center for Health Policy*

Gregory Daniel

Adam Kroetsch

Nirosha Mahendraratnam
Lederer

Kerra Mercon

Morgan Romine

Rachel Sherman

APPENDIX B: THE SEVEN CHARACTERISTICS OF ADEQUATE AND WELL CONTROLLED STUDIES (21 CFR § 314.126)

TITLE 21 — FOOD AND DRUGS
CHAPTER I — FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER D — DRUGS FOR HUMAN USE
PART 314 — APPLICATIONS FOR FDA APPROVAL TO MARKET A NEW DRUG

Subpart D — FDA Action on Applications and Abbreviated Applications

Sec. 314.126¹⁵ Adequate and well-controlled studies.

(a) The purpose of conducting clinical investigations of a drug is to distinguish the effect of a drug from other influences, such as spontaneous change in the course of the disease, placebo effect, or biased observation. The characteristics described in paragraph (b) of this section have been developed over a period of years and are recognized by the scientific community as the essentials of an adequate and well-controlled clinical investigation. The Food and Drug Administration considers these characteristics in determining whether an investigation is adequate and well-controlled for purposes of section 505 of the act. Reports of adequate and well-controlled investigations provide the primary basis for determining whether there is “substantial evidence” to support the claims of effectiveness for new drugs. Therefore, the study report should provide sufficient details of study design, conduct, and analysis to allow critical evaluation and a determination of whether the characteristics of an adequate and well-controlled study are present.

(b) An adequate and well-controlled study has the following characteristics:

(1) There is a clear statement of the objectives of the investigation and a summary of the proposed or actual methods of analysis in the protocol for the study and in the report of its results. In addition, the protocol should contain a description of the proposed methods of analysis, and the study report should contain a description of the methods of analysis ultimately used. If the protocol does not contain a description of the proposed methods of analysis, the study report should describe how the methods used were selected.

(2) The study uses a design that permits a valid comparison with a control to provide a quantitative assessment of drug effect. The protocol for the study and report of results should describe the study design precisely; for example, duration of treatment periods, whether treatments are parallel, sequential, or crossover, and whether the sample size is predetermined or based upon some interim analysis. Generally, the following types of control are recognized:

(i) Placebo concurrent control. The test drug is compared with an inactive preparation designed to resemble the test drug as far as possible. A placebo-controlled study may include additional treatment groups, such as an active treatment control or a dose-comparison control, and usually includes randomization and blinding of patients or investigators, or both.

(ii) Dose-comparison concurrent control. At least two doses of the drug are compared. A dose-comparison study may include additional treatment groups, such as placebo control or active

control. Dose-comparison trials usually include randomization and blinding of patients or investigators, or both.

(iii) No treatment concurrent control. Where objective measurements of effectiveness are available and placebo effect is negligible, the test drug is compared with no treatment. No treatment concurrent control trials usually include randomization.

(iv) Active treatment concurrent control. The test drug is compared with known effective therapy; for example, where the condition treated is such that administration of placebo or no treatment would be contrary to the interest of the patient. An active treatment study may include additional treatment groups, however, such as a placebo control or a dose-comparison control. Active treatment trials usually include randomization and blinding of patients or investigators, or both. If the intent of the trial is to show similarity of the test and control drugs, the report of the study should assess the ability of the study to have detected a difference between treatments. Similarity of test drug and active control can mean either that both drugs were effective or that neither was effective. The analysis of the study should explain why the drugs should be considered effective in the study, for example, by reference to results in previous placebo-controlled studies of the active control drug.

(v) Historical control. The results of treatment with the test drug are compared with experience historically derived from the adequately documented natural history of the disease or condition, or from the results of active treatment, in comparable patients or populations. Because historical control populations usually cannot be as well assessed with respect to pertinent variables as can concurrent control populations, historical control designs are usually reserved for special circumstances. Examples include studies of diseases with high and predictable mortality (for example, certain malignancies) and studies in which the effect of the drug is self-evident (general anesthetics, drug metabolism).

(3) The method of selection of subjects provides adequate assurance that they have the disease or condition being studied, or evidence of susceptibility and exposure to the condition against which prophylaxis is directed.

(4) The method of assigning patients to treatment and control groups minimizes bias and is intended to assure comparability of the groups with respect to pertinent variables such as age, sex, severity of disease, duration of disease, and use of drugs or therapy other than the test drug. The protocol for the study and the report of its results should describe how subjects were assigned to groups. Ordinarily, in a concurrently controlled study, assignment is by randomization, with or without stratification.

(5) Adequate measures are taken to minimize bias on the part of the subjects, observers, and analysts of the data. The protocol and report of the study should describe the procedures used to accomplish this, such as blinding.

(6) The methods of assessment of subjects' response are well-defined and reliable. The protocol for the study and the report of results should explain the variables measured, the methods of observation, and criteria used to assess response.

(7) There is an analysis of the results of the study adequate to assess the effects of the drug. The report of the study should describe the results and the analytic methods used to evaluate them, including any appropriate statistical methods. The analysis should assess, among other things, the comparability of test and control groups with respect to pertinent variables, and the effects of any interim data analyses performed.

(c) The Director of the Center for Drug Evaluation and Research may, on the Director's own initiative or on the petition of an interested person, waive in whole or in part any of the criteria in paragraph (b) of this section with respect to a specific clinical investigation, either prior to the investigation or in the evaluation of a completed study. A petition for a waiver is required to set forth clearly and concisely the specific criteria from which waiver is sought, why the criteria are not reasonably applicable to the particular clinical investigation, what alternative procedures, if any, are to be, or have been employed, and what results have been obtained. The petition is also required to state why the clinical investigations so conducted will yield, or have yielded, substantial evidence of effectiveness, notwithstanding nonconformance with the criteria for which waiver is requested.

(d) For an investigation to be considered adequate for approval of a new drug, it is required that the test drug be standardized as to identity, strength, quality, purity, and dosage form to give significance to the results of the investigation.

(e) Uncontrolled studies or partially controlled studies are not acceptable as the sole basis for the approval of claims of effectiveness. Such studies carefully conducted and documented, may provide corroborative support of well-controlled studies regarding efficacy and may yield valuable data regarding safety of the test drug. Such studies will be considered on their merits in the light of the principles listed here, with the exception of the requirement for the comparison of the treated subjects with controls. Isolated case reports, random experience, and reports lacking the details which permit scientific evaluation will not be considered.

[50 FR 7493, Feb. 22, 1985, as amended at 50 FR 21238, May 23, 1985; 55 FR 11580, Mar. 29, 1990; 64 FR 402, Jan. 5, 1999; 67 FR 9586, Mar. 4, 2002]

APPENDIX C: THE SPECTRUM OF REAL WORLD STUDY DESIGNS

RWE studies can either be interventional (e.g., pragmatic clinical trials [PCTs]) or non-interventional (e.g., observational studies using secondary data or open-label). Although RWD are captured during the delivery of health care, studies that take place in a real-world setting can apply some of the same design elements as traditional clinical trials (e.g., randomization).

In interventional RWE studies, treatments are assigned to patients, and randomization can be utilized. There are different types of interventional RWE studies, including PCTs and large simple trials. PCTs are perceived as the closest real-world study design akin to RCTs, but have broader inclusion/exclusion criteria and include real-world or “pragmatic” design elements. They can allow for a better understanding of the effectiveness of a medical product in routine practice and indicated populations.⁸⁵

Hybrid RWE studies incorporate elements of both interventional and non-interventional studies. Single-arm, open-label extension studies can serve as real-world continuations of RCTs depending on how they are specified in the protocols and, in some cases, can allow researchers to build on RCT data to better understand the long-term effects of a product in an environment that reflects the real world. Additionally, an external control arm comprised of RWD may be as a comparator in a single-arm trial.

Non-interventional studies do not assign treatments to patients. They can use both primary data—information that was collected for a specific research goal such as in a registry study—and secondary data—information that was collected for another purpose but that can be used to answer additional research questions, such as in secondary database analyses.⁸⁶ Non-interventional studies include studies such as analyses of observational disease registries, hypothesis-driven analyses of claims and EHR data, cohort studies, and case-control studies.

APPENDIX D. PREVIOUS DUKE-MARGOLIS WORK ON RWD AND RWE

A Framework for Regulatory Use of Real-World Evidence

September 13, 2017

Describes distinct definitions for what constitutes RWD and RWE, the considerations that should guide the development of RWE that is fit for regulatory purposes, and high-priority opportunities to improve such development and use

Duke | MARGOLIS CENTER
for Health Policy

Characterizing RWD Quality and Relevancy for Regulatory Purposes

October 1, 2018

Provides an overview of the data curation process and identifies potential challenges for regulators in assessing the underlying appropriateness of the RWD source for the given regulatory question of interest

Duke | MARGOLIS CENTER
for Health Policy

Determining Real-World Data's Fitness for Use and the Role of Reliability

September 26, 2019

Examines the key components of RWD fitness-for-use considerations and identifies principles for developing a minimum set of reliability checks

Duke | MARGOLIS CENTER
for Health Policy

Understanding the Need for Non-Interventional Studies Using Secondary Data to Generate Real-World Evidence for Regulatory Decision Making, and Demonstrating Their Credibility

November 25, 2019

Provides guidance on when conducting interventional RWE studies may not be possible and how to demonstrate credibility in the causal inference made by non-interventional studies using secondary data (observational studies)

Duke | MARGOLIS CENTER
for Health Policy

REFERENCES

1. U.S. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. U.S. Department of Health & Human Services; December 2018.
2. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med*. 2016;375(23):2293-2297.
3. ClinicalTrials.gov. Glossary of Common Site Terms. NIH US National Library of Medicine. Published 2019. Available at: <https://clinicaltrials.gov/ct2/about-studies/glossary>. Accessed November 6, 2019.
4. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the Use of Nonrandomized Real-World Data Analyses for Regulatory Decision Making. *Clin Pharmacol Ther*. 2019;105(4):867-877.
5. Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? *Can J Surg*. 2010;53(5):345-348.
6. Protection of human subjects; Belmont Report: notice of report for public comment. Fed Regist. 1979;44(76):23191-23197.
7. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*. 1987;317(3):141-145.
8. Meyer MN, Heck PR, Holtzman GS, et al. Objecting to experiments that compare two unobjectionable policies or treatments. *Proc Natl Acad Sci U S A*. 2019;116(22):10723.
9. Christian JB, Brouwer ES, Girman CJ, Bennett D, Davis KJ, Dreyer NA. Masking in Pragmatic Trials: Who, What, and When to Blind. *Ther Innov Regul Sci*. 2019;2168479019843129.
10. Rowlands C, Rooshenas L, Fairhurst K, Rees J, Gamble C, Blazeby JM. Detailed systematic analysis of recruitment strategies in randomised controlled trials in patients with an unscheduled admission to hospital. *BMJ Open*. 2018;8(2):e018581.
11. Corrigan-Curay J, Sacks L, Woodcock J. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. *JAMA*. 2018;320(9):867-868.
12. Chen DT, Wynia MK, Moloney RM, Alexander GC. U.S. physician knowledge of the FDA-approved indications and evidence base for commonly prescribed drugs: results of a national survey. *Pharmacoepidemiol Drug Saf*. 2009;18(11):1094-1100.
13. Sullivan HW, O'Donoghue AC, Aikin KJ. Primary care physicians' use of FDA-approved prescription drug labels. *J Am Board Fam Med*. 2014;27(5):694-698.
14. Faria R, Hernandez Alava M, Manca A, Wailoo AJ. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Available at: <http://www.nicedsu.org.uk>. Accessed November 6, 2019.
15. 21CFR314.126. Code of Federal Regulations. Available at: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?fr=314.126>. Accessed November 6, 2019.
16. AHRQ Methods for Effective Health Care. In: Gliklich RE, Dreyer NA, Leavy MB, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014.
17. U.S. Food and Drug Administration. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. Silver Spring (MD): 2013.
18. 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):1033-1039.
19. Gliklich R, Dreyer NA, Leavy MB, Velentgas P, Khurana L. Standards in the Conduct of Registry Studies for Patient-Centered Outcomes Research. Washington (DC): Patient-Centered Outcomes Research Institute;2012.
20. Study design and methods. In: European Network of Centres for Pharmacoepidemiology and Pharmacovigilance Guide on Methodological Standards in Pharmacoepidemiology. Vol 7. European Medicines Agencies; 2018.
21. Stürmer T, Brookhart MA. Study Design Considerations. In: Velentgas P, Dreyer NA, Nourjah P, et al., editors. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Jan. Chapter 2. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK126187/>. Accessed November 6, 2019.
22. Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.

23. 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):1107-1113.
24. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296.
25. Smith GD, Ebrahim S. Data dredging, bias, or confounding. *BMJ*. 2002;325(7378):1437-1438.
26. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1018-1032.
27. U.S. Food and Drug Administration. Good Review Practice: Good Review Management Principles and Practices for Effective IND Development and Review. Silver Spring (MD): 2013.
28. Wang SV, Verpillat P, Rassen JA, Patrick A, Garry EM, Bartels DB. Transparency and Reproducibility of Observational Cohort Studies Using Large Healthcare Databases. *Clin Pharmacol Ther*. 2016;99(3):325-332.
29. ClinicalTrials.gov. Available at: <https://clinicaltrials.gov>. Accessed November 6, 2019.
30. FDAAA 801 and the Final Rule. Published 2018. Available at: <https://clinicaltrials.gov/ct2/manage-recs/fdaaa>. Accessed November 6, 2019.
31. Real-World Evidence Transparency Initiative. Available at: <https://www.ispor.org/strategic-initiatives/real-world-evidence/real-world-evidence-transparency-initiative>. Accessed November 6, 2019.
32. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical Depiction of Longitudinal Study Designs in Health Care Databases. *Ann Intern Med*. 2019;170(6):398-406.
33. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008;8:70.
34. Moher B, Brooks J, Clark MA, et al. A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. *Value Health*. 2003;6(2):90-97.
35. Setoguchi S, Gerhard T. Comparator Selection. In: Velentgas P, Dreyer NA, Nourjah P, et al., editors. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013 Jan. Chapter 5. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK126184/>. Accessed November 6, 2019.
36. Pazzagli L, Linder M, Zhang M, et al. Methods for time-varying exposure related problems in pharmacoepidemiology: An overview. *Pharmacoepidemiol Drug Saf*. 2018;27(2):148-160.
37. Roberts AW, Dusetzina SB, Farley JF. Revisiting the washout period in the incident user study design: why 6-12 months may not be sufficient. *J Comp Eff Res*. 2015;4(1):27-35.
38. Pottegård A, Friis S, Stürmer T, Hallas J, Bahmanyar S. Considerations for Pharmacoepidemiological Studies of Drug–Cancer Associations. *Basic Clin Pharmacol Toxicol*. 2018;122(5):451-459.
39. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol*. 2008;167(4):492-499.
40. Edwards JK, Htoo PT, Stürmer T. Keeping the Demons At Bay When Handling Time Varying Exposures: Beyond Avoiding Immortal Person Time. *Am J Epidemiol*. 2019.
41. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ*. 2010;340:b5087.
42. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. *BMJ*. 2018;360:k182.
43. Moride Y, Abenhaim L. Evidence of the depletion of susceptibles effect in non-experimental pharmacoepidemiologic research. *J Clin Epidemiol*. 1994;47(7):731-737.
44. Renoux C, Dell'Aniello S, Brenner B, Suissa S. Bias from depletion of susceptibles: the example of hormone replacement therapy and the risk of venous thromboembolism. *Pharmacoepidemiol Drug Saf*. 2017;26(5):554-560.
45. Wyss R, Gagne JJ, Zhao Y, et al. Use of Time-Dependent Propensity Scores to Adjust Hazard Ratio Estimates in Cohort Studies with Differential Depletion of Susceptibles. *Epidemiology*. 2019.
46. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015;11(7):437-441.
47. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep*. 2015;2(4):221-228.

48. U.S. Food and Drug Administration. Guidance for Industry: Non-Inferiority Clinical Trials to Establish Effectiveness. Silver Spring (MD); 2016.
49. Patient-Centered Outcomes Research Institute. Available at: <https://www.pcori.org/research-results/about-our-research/research-we-support>. Accessed November 6, 2019.
50. Scott IA. Non-inferiority trials: determining whether alternative treatments are good enough. *Med J Aust*. 2009;190(6):326-330.
51. Lesaffre E. Superiority, equivalence, and non-inferiority trials. *Bull NYU Hosp Jt Dis*. 2008;66(2):150-154.
52. Dasgupta A, Lawson KA, Wilson JP. Evaluating equivalence and noninferiority trials. *Am J Health Syst Pharm*. 2010;67(16):1337-1343.
53. Schulman KL, Berenson K, Tina Shih YC, et al. A checklist for ascertaining study cohorts in oncology health services research using secondary data: report of the ISPOR oncology good outcomes research practices working group. *Value Health*. 2013;16(4):655-669.
54. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol*. 2009;62(12):1233-1241.
55. Beran TN, Violato C. Structural equation modeling in medical research: a primer. *BMC Res Notes*. 2010;3:267-267.
56. VanderWeele TJ, Shpitser I. On the definition of a confounder. *Ann Stat*. 2013;41(1):196-220.
57. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. 2006;15(5):291-303.
58. Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: Relating endogeneity and unobserved heterogeneity to confounding. *Ann Epidemiol*. 1997;7(4):251-257.
59. Gunasekara FI, Carter K, Blakely T. Glossary for econometrics and epidemiology. *J Epidemiol Community Health*. 2008;62(10):858.
60. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. *Gastroenterol Hepatol Bed Bench*. 2012;5(2):79-83.
61. Chan CK, Feinstein AR, Jekel JF, Wells CK. The value and hazards of standardization in clinical epidemiologic research. *J Clin Epidemiol*. 1988;41(11):1125-1134.
62. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report, 2014;Part III. *Value Health*. 2009;12(8):1062-1073.
63. Baser O. Choosing propensity score matching over regression adjustment for causal inference: when, why and how it makes sense. *J Med Econ*. 2007;10(4):379-391.
64. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424.
65. Stürmer T, Wyss R, Glynn RJ, Brookhart MA. Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs. *J Intern Med*. 2014;275(6):570-580.
66. Hanley JA, Csizmadia I, Collet JP. Two-stage case-control studies: precision of parameter estimates and considerations in selecting sample size. *Am J Epidemiol*. 2005;162(12):1225-1234.
67. Delaney JA, Suissa S. The case-crossover study design in pharmacoepidemiology. *Stat Methods Med Res*. 2009;18(1):53-65.
68. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions - a primer for the clinician. *Clin Epidemiol*. 2017;9:185-193.
69. Angrist JD, Pischke JS. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press; 2014.
70. Schneeweiss S. Confounding. In: Hartzema A CA, Porta M, Tilson H, ed. *Pharmacoepidemiology*. 4th ed.
71. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388.
72. Arnold BF, Ercumen A, Benjamin-Chung J, Colford JM, Jr. Brief Report: Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies. *Epidemiology*. 2016;27(5):637-641.

73. Zhang X, Faries DE, Li H, Stamey JD, Imbens GW. Addressing unmeasured confounding in comparative observational research. *Pharmacoepidemiol Drug Saf.* 2018;27(4):373-382.
74. West SL, Strom BL, Poole C. Validity of Pharmacoepidemiology Drug and Diagnosis Data. In *Pharmacoepidemiology*. 3rd edn. Strom BL (ed). Wiley: Chichester, 2000; 668.
75. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. *JAMA.* 2019;321(6):602-603.
76. Ioannidis JPA, Tan YJ, Blum MR. Limitations and Misinterpretations of E-Values for Sensitivity Analyses of Observational Studies. *Ann Intern Med.* 2019;170(2):108-111.
77. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther.* 2017;102(6):924-933.
78. Dahabreh IJ, Robertson SE, Hernán MA. On the Relation Between G-formula and Inverse Probability Weighting Estimators for Generalizing Trial Results. *Epidemiology.* 2019;30(6):807-812.
79. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol.* 2016;183(8):758-764.
80. Lodi S, Phillips A, Lundgren J, et al. Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *Am J Epidemiol.* 2019;188(8):1569-1577.
81. Franklin JM, Pawar A, Martin D, et al. Nonrandomized real-world evidence to support regulatory decision-making: Process for a randomized trial replication project. *Clin Pharmacol Ther.* 2019.
82. RCT Duplicate. Available at: <https://www.rctduplicate.org/>. Accessed November 6, 2019.
83. OPERAND. Available at: <https://mrctcenter.org/blog/projects/real-world-evidence/>. Accessed November 6, 2019.
84. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of Using Real-World Data to Replicate Clinical Trial Evidence. *JAMA Netw Open.* 2019;2(10):e1912869.
85. Christian JB, Brouwer ES, Girman CJ, Bennett D, Davis KJ, Dreyer NA. Masking in Pragmatic Trials: Who, What, and When to Blind. *Ther Innov Regul Sci.* 2019.
86. Prada-Ramallal G, Roque F, Herdeiro MT, Takkouche B, Figueiras A. Primary versus secondary source of data in observational studies and heterogeneity in meta-analyses of drug effects: a survey of major medical journals. *BMC Med Res Methodol.* 2018;18(1):97-97.