June 18, 2024

Real-World Evidence to Support Causal Inference: Methodological Considerations for Non-Interventional Studies

healthpolicy@duke.edu

Duke | MARGOLIS INSTITUTE *for* Health Policy

DUKE-MARGOLIS AUTHORS

Nora Emmott
Maryam Nafie
Dure Kim
Rachele Hendricks-Sturrup

Acknowledgements

Duke-Margolis would like to thank several individuals for their contributions to this white paper. The paper would not have been possible without the months-long collaboration of the Causal Inference Workstream, which was comprised of representatives from the Duke-Margolis Real-World Evidence Collaborative. Their expert perspectives, open discussion, and thoughtful feedback on working drafts were indispensable, and we are grateful for their engagement. We also thank the participants in the expert workshop Duke-Margolis convened in December 2023 and Mark McClellan for the valuable insight they provided. Finally, the authors wish to acknowledge Patricia Green within the Institute for communications support and guidance; Laura Hughes, Emma Kikerkov, and Patrick Rodriguez for support with copyediting, graphics, and design; and Jieyu Zhang for support with summarizing relevant peer-reviewed literature.

Any opinions expressed in this paper are solely those of the authors, and do not represent the views or policies of any other organization 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 Institute, as well as a combination of financial and in-kind contributions from Real-World Evidence Collaborative members, including Abbvie, Amgen, Bayer, Boehringer Ingelheim, Genentech, GSK, Eli Lilly and Company, Merck, Novartis, and Teva.

About the Duke-Margolis Institute for Health Policy

The Robert J. Margolis, MD, Institute 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, health equity 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 policymaking and implementation for better health and health care. For more information, visit healthpolicy.duke.edu.

Disclosures

Rachele Hendricks-Sturrup served as an independent executive with the National Alliance Against Disparities in Patient Health and is presently an independent director on the board of Public Responsibility in Medicine & Research (PRIM&R).

Mark B. McClellan, MD, PhD, is an independent director on the boards of Johnson & Johnson, Cigna, Alignment Healthcare, and PrognomIQ; co-chairs the Guiding Committee for the Health Care Payment Learning and Action Network; and receives fees for serving as an advisor for Arsenal Capital Partners, Blackstone Life Sciences, and MITRE.

Duke | MARGOLIS INSTITUTE for Health Policy

Executive Summary

The aim of non-interventional studies, a type of study in which patients receive the marketed drug of interest during routine medical practice and are not assigned to an intervention according to a protocol,[1] is to uncover insights that may be inaccessible through controlled trials. Non-interventional or observational study designs can play a crucial role in assessing treatment effects (i.e., causality) beyond the confines of traditional randomized controlled trials (RCTs). Within these study designs, routine clinical care outcomes are observed among real-world populations, as opposed to research participants of RCTs selected according to narrow inclusion/exclusion criteria. Real-world data (RWD), derived from sources such as electronic health records, claims data, and registries, offers a less constrained environment that better reflects the complexity and diversity of clinical practice. Additionally, real-world studies typically have much larger sample sizes, facilitating subgroup analyses often infeasible in RCTs. Subgroups, in this case, describes an analysis unit of a subset of participants within a given study population.[2] This nuanced understanding can inform health care decision-making by capturing real-world outcomes, patient variability, and long-term effects of interventions observed as part of regular clinical care.

Real-world evidence (RWE) complements RCTs by providing timely insights into effectiveness across diverse populations beyond traditional clinical trials. Regulatory initiatives, such as the U.S. Food and Drug Administration's (FDA) Advancing Real-World Evidence Program,[3] acknowledge the value of RWE, aiming to modernize evidence generation and incorporate patient perspectives. However, ensuring the credibility of RWE for causal inference requires clear design, fit-for-purpose RWD, communication, and rigorous statistical analysis. Promoting RWE's capacity for causal inference is essential for advancing evidence-based health care. Regulators recognize that certain limitations accompany the use of RWD to determine or measure causality. Proposed approaches might involve established concepts like target trial emulation and/or other causal frameworks to address confounding and other types of bias and schemas to describe overall study designs. Integrating RWE's strengths with traditional research methods like RCTs can present a more comprehensive understanding of health care interventions and their real-world impacts.

Promoting RWE's capacity for causal inference is essential for advancing evidence-based health care.

How This Paper Was Developed

This whitepaper was informed by: a December 12th, 2023 private workshop entitled, "Generating and Leveraging RWE for Causal Inference," hosted by the Duke-Margolis Real-World Evidence Collaborative; several regular working group and stakeholder calls with members of the Collaborative; and literature cited throughout this white paper. Workshop participants included industry representatives, sponsors, research groups, and data curators and vendors who each provided insight, examples, and expert perspectives on leveraging RWD in causal inference studies. A glossary of terms is provided in Appendix D.

Background

RCTs are explanatory studies that involve rigorous study designs aimed at establishing causality between interventions and outcomes under discernable or isolated conditions. Pragmatic trials, while explanatory, differ from RCTs in that they investigate outcomes following an intervention(s) in usual care versus isolated settings.[4] In both designs, and more so in RCTs, randomization in the assignment of an intervention(s) or observation of outcomes following standard treatment(s), and use of control or reference groups, serves to mitigate, where possible, the influence of potential bias and confounding, thereby enhancing the validity of the study.

However, both designs are accompanied by important considerations and limitations. Tightly controlled settings within RCTs can lead to potentially inaccurate portrayals of treatment outcomes across diverse patients, health care settings, and other environments in the real-world. Yet, in pragmatic trials, issues such as selection bias, differential loss to follow-up, gaps caused by limited questions—which could be otherwise answerable using RCTs, as well as other sources of bias can further limit the generalizability of pragmatic trial findings. In certain cases, or in instances where patients can easily exhaust all available standard of care (SOC) options (e.g. rare disease), randomization to a negative control arm may not be ethical or might warrant the need for a self-controlled study design—such as one of multiple non-interventional, observational study designs—that involves the use of natural history RWD.

When intentionally designed, non-interventional or observational study designs that involve fit-for-purpose data can help researchers assess treatment effectiveness and measure causality in real-world settings. Within these study designs, routine care outcomes are observed and measured across all patient populations in their usual care or treatment settings. Certain non-interventional studies have also been described as hypothesis-evaluating studies,[5] or hypothesis evaluating treatment effect (HETE) studies. Registration, data traceability, and clear reporting are key measures to instill confidence in the research process and mitigate the risk of study biases. Bias assessments and sensitivity analyses post-study are also essential to verify the robustness and validity of the study's findings.

While some nonrandomized studies and RCTs may yield similar outcomes,[6] discrepancies and reproducibility issues can arise due to differences in study populations, methodologies, and settings. Non-interventional studies can be a mechanism to generate insights that may be otherwise inaccessible via RCTs–one example of this is stronger subgroup analyses since RWE studies typically comprise larger sample sizes. Therefore, results from non-interventional studies can be useful to inform health care decisions by capturing real-world outcomes, patient variability, and long-term effects of interventions.

Regulatory Overview

Regulatory initiatives, such as the United States FDA's Advancing Real-World Evidence Program, acknowledge the value of RWE, aiming to modernize evidence generation and incorporate patient perspectives.[8] However, ensuring the credibility of RWE for causal inference purposes requires clear design, fit-for-purpose RWD, communication, and rigorous statistical analysis.

Regulators are aligned that certain limitations accompany the use of RWD to determine or measure causality. For instance, the FDA has stated that some RWD sources are inherently biased (e.g., claims and electronic health record [EHR] data are driven by health care utilization and payment). Likewise, the Taiwan FDA (TFDA) has noted that "due to methodological limitations, observational studies

are more difficult to establish causality... the results of observational studies are not suitable to be used alone as confirmatory evidence of drug effectiveness but real-world evidence from well-defined, well-designed, and well-executed observational studies may serve as supporting evidence of efficacy."[9] Regulators from several countries recognize that observational study designs that use rigorous methods, make appropriate study design and data analysis choices, and use fit-for-purpose data to mitigate bias with reliable statistical methods, have the potential to measure causality. Moreover, the FDA and Health Canada/CADTH have recognized that causality diagrams could be helpful in providing a rationale for study design and analysis choices.[10, 11] Both regulators, therefore, encourage sponsors to describe how their confounder (see glossary in Appendix D) variables were selected and whether or not they were informed by causal diagrams, and also provide information on the proposed approach to support causal inference.

Single-world intervention graphs (SWIGs) provide a visual and analytical framework to identify causal effects, clarify adjustment criteria and help identify steps needed to control biases in observational data. SWIGs allow researchers to explicitly represent counterfactual scenarios and better understand the causal relationships between variables, making them a valuable tool for mitigating biases and improving the validity of causal inferences, thus remedying the limitation of Directed Acyclic Graphs (DAGs) of not allowing counterfactual outcomes to be depicted. The FDA has described SWIGs and DAGs as useful schema to describe an overall study design and provide a causal diagram to specify theorized causal relationships.[12]

While the FDA has mentioned that they do not endorse the use of one causal framework over another, they encourage sponsors to describe their proposed approach to support causal inference and mitigate bias and confounding.[13] In this white paper, we describe causal inference frameworks that can be informative for researchers seeking to implement study designs that involve the integration of RWD/RWE and are useful within regulatory contexts that support drug labeling and/or labeling expansion.

Overview of Causal Inference

Causal inference using RWE is a methodology that seeks to establish causal relationships between exposures or interventions and outcomes using observational data collected from real-world settings, such as electronic health records, claims databases, patient registries, and other sources. Unlike RCTs, which are designed to assess causal effects under controlled conditions, RWE focuses on understanding causality in the context of everyday clinical practice and population-level data. Causal inference from observational data relies on research design and methodology that includes appropriate data selection and measurements and efforts to minimize bias.

RWE study approaches that integrate causal frameworks offer a systematic approach to tackle complex problems by integrating theory with practice, translating theoretical concepts into analytical steps to ensure accurate data analysis reflection. Such frameworks clarify and focus research questions into clear pathways, improving transparency around assumptions and statistical methods tailored to research needs and data constraints. They also enhance communication of research, clarifying methodology, assumptions, and findings as well as include sensitivity analyses that allow for the assessment of the consequences of violating causal assumptions.

RWE study approaches that integrate causal frameworks offer a systematic approach to tackle complex problems by integrating theory with practice, translating theoretical concepts into analytical steps to ensure accurate data analysis reflection.

To estimate the estimand[14] reliably from observational data and avoid design-related mistakes, multiple frameworks exist to assess causality using RWD: the estimand framework, target trial framework, and causal roadmap framework. We discuss these frameworks in detail below, bearing in mind that these three causal inference frameworks complement one another, are highly interrelated, and are aligned with foundational assumptions. By combining rigorous causal inference frameworks and methodologies with rich and diverse real-world data sources, researchers can generate valuable insights into the causal effects of health care interventions, improve patient outcomes, and drive evidence-based decision-making in health care and public health domains.

To reliably estimate the estimand from observational data and avoid design-related mistakes, there are multiple frameworks to assess causality using RWD: the estimand framework, target trial framework, and causal roadmap framework

Target Trial Framework

The target trial framework is typically used in observational studies to emulate an RCT to estimate the effect of a treatment on an outcome. The target trial framework can be useful for mitigating certain biases, such as immortal time bias and selection bias (see glossary in Appendix D).[15] This approach involves specifying a hypothetical RCT's protocol (defining eligibility, treatment/treatment regimen, follow-up periods, outcomes, etc.) and mimicking these components using observational data, adjusting for confounders to align with the RCT to mitigate biases. With the target trial approach, researchers compare and assess the differences between the target trial and the emulation using RWD, typically using a table. Trial emulations using RWD need to be interpreted with caution because of the likelihood of unmeasured confounding. The target trial approach is one method to help ensure the reliability and applicability of RWE by addressing inherent biases and facilitating more accurate assessments of causality and drug effectiveness.[16]

The literature offers key points and guidance that are relevant to the importance of assessing how well causal analyses of observational data can replicate the specific target trial of interest, especially when making decisions among various strategies. For example, Hernán and Robins discussed the core components of the target trial protocol while also acknowledging challenges that are inherent to target trial emulation using observational data.[17, 18] The target trial approach, being grounded in counterfactual theory, offers a practical and intuitive framework for causal inference and serves as a unifying principle for various causal inference methods, and offers a structured process for evaluating observational RWD. The approach is also useful to assess the effects of sustained treatment strategies and can help researchers circumvent frequently observed methodological challenges, such as data bias and quality assurance.

The target trial approach is accompanied by limitations that should be noted. For example, Wallach, et al., conducted an observational study that aimed to emulate an ongoing randomized controlled trial comparing cardiovascular outcomes for prostate cancer patients treated with degarelix versus leuprolide (PRONOUNCE study).[19] Using retrospective claims data to identify patients who met the study's eligibility criteria and compare the risk of major adverse cardiovascular events between treatment

groups, the target trial framework was used to replicate PRONOUNCE as closely as possible. The study was limited, however, by incomplete data needed to replicate PRONOUNCE inclusion/exclusion criteria and endpoints, as well as potential residual confounding.

Nonetheless, growing consensus exists that the target trial approach increases transparency, validity, and interpretability of causal inferences established using observational data. This potential is particularly useful to overcome instances in which randomized clinical trials are infeasible, yet a demonstration of causality is either desired or required and accomplishable using RWD. The risk of perpetuating bias due to information censoring in target trials using external comparator data can also be overcome or corrected by applying inverse probability of censoring weighting.[20]

Some observational studies have shown conflicting results between target trial emulation and RCTs. For example, Martinez-Ales, et al., reported such conflicting results when assessing the mortality of hospitalized COVID-19 patients who received a flexible thromboprophylaxis therapy compared to a standard low dose strategy.[21] In previous RCTs for critical COVID-19 patients, however, there were no significant mortality differences across the different treatments.

Estimand Framework

The estimand (see glossary in Appendix D) framework[22] is a structured approach to clarify study objectives and address uncertainties, particularly in the presence of deviations—including intercurrent events (see glossary in Appendix D), such as treatment discontinuation or emergency medication use. Emphasizing the precise definition of the treatment effect to be measured, the estimand framework aims to reflect the trial's true intent and ensure the accurate interpretation of results. This framework consists of five key attributes: treatment, population, outcome variable, population-level summary, and handling intercurrent events. These components help to specify the trial's focus clearly, including the treatments compared, the patient group studied, the endpoint to be assessed, how group comparisons are summarized, and how deviations from the intended treatment are managed. The framework consequently provides a clearer picture of the intended treatment effect and enhancing the trial's relevance and applicability.

The literature also offers key points and guidance that are relevant to the importance of aligning study design, data collection, and estimation with the estimand. For example, Kahan, et al.,[23] described the role of study methods in a researcher's ability to estimate a desired estimand. The estimand framework aims to create guidelines and resolve issues by increasing transparency on the treatment effect of interest through researchers outlining each quantifiable attribute.

The estimand framework aims to reduce ambiguity in both the research questions and conclusions. Yet, limitations to the identifiability of the estimand exist and are important to understand. Observational data can be incomplete and contain measurement errors and biases.[24] These data limitations can be addressed through a series of analyses conducted with the intent to explore the robustness of inferences from the main estimator to deviations from its underlying modeling assumptions and limitations in the data. By ensuring research questions are clearly described, the estimand framework can help to assess treatment effects and help researchers avoid methodological challenges or shortcomings.

Combined Target Trial and Estimand Framework

The estimand and the target trial framework can be usefully combined to determine causality and reach similar conclusions to RCTs. For example, one study used the ELARA phase II data source with both the target trial and the estimand frameworks to examine the causal effect of tisagenlecleucel treatment strategy compared to SOC.[25] This study found that using target trial and the estimand framework in tandem leads to early internal alignment on study objectives, common understanding of potential sources of bias, and an early assessment of the quality and relevance of external controls. This combined approach also can help researchers clarify the target trial design, improve the transparency of assumptions needed to emulate the target trial, and help facilitate choices around the best estimand.

The estimand and target trial frameworks are interrelated and have similar purposes in answering the scientific question, however a limited number of studies exist that use this approach. For example, one study combined the estimand and target trial frameworks to compare long-term survival outcomes of a pooled set of three previously reported randomized phase 3 trials studying patients with metastatic, non-small cell lung cancer receiving front-line chemotherapy and similar patients treated with front-line chemotherapy as part of routine clinical care.[26] The researchers described their methods to combine both approaches: first, they defined the hypothetical target trial structured according to the estimand framework; then the study that attempted to emulate it, thus leveraging elements from both frameworks.

Causal Roadmap

The causal roadmap[27] is a practical guide on the implementation of RWE studies that provides a practical, unified structure for designing and analyzing these studies, ensuring the inclusion of all relevant information. It is an explicit, itemized, and iterative process that guides investigators to prespecify study design and analysis plans and addresses a wide range of guidance within a single framework. By supporting the transparent evaluation of causal assumptions and facilitating objective comparisons of design and analysis choices based on prespecified criteria, the framework helps investigators evaluate the quality of evidence that a given study is likely to produce, specify a study to generate high-quality RWE, and communicate effectively with regulatory agencies and other stakeholders.

The literature[28] also offers considerations relevant to the key elements of a study design and analysis plan using the roadmap steps. Researchers outline the seven steps to help investigators prespecify design and analysis plans for studies that utilize RWD: 1) specify the causal question, estimand, and model; 2) define the observed data that will be or have been collected; 3) assess identifiability of the causal estimand from the observed data; 4) define the statistical estimand; 5) specify the statistical model, estimator, and method of confidence interval construction; 6) specify the sensitivity analyses; and 7) compare feasible study designs (Steps 1–6) using outcome-blind simulations. Limitations to the causal roadmap are similar to the data limitations of other frameworks using observational data (e.g., selection bias, bias due to baseline confounding, and the ability to correctly define the index date for comparison). By using the causal roadmap, researchers aim to produce high-quality estimates of causal effects using RWD and to evaluate whether the proposed methods are adequate for drawing causal inferences.

Causality in RWE to Satisfy Substantial Evidence Criterion Frameworks

To measure causality using non-interventional study designs, the FDA[29] has encouraged sponsors to clearly describe: the research question (study objective) or outcome of interest, hypothesis, and choice of study design and data sources (including rationale); the proposed approach to support causal inference (target trial emulation or other conceptual approach); the common data model(s) used; and the plans to address confounding and other types of bias (e.g., prevalent user bias, selection bias, immortal time bias, etc.),[30] especially when studies lack randomization or blinding.[31] The frameworks described above are procedural steps toward meeting such recommendations to potentially satisfy regulators' substantial evidence criterion using RWE. The European Medicines Agency's (EMA) draft reflection paper on non-interventional studies, for example, describes the target trial framework as useful, structured, and coherent enough to design non-interventional studies with causal objectives.[32]

A few notable examples of studies inferring causality in RWE to satisfy substantial evidence criterion exist. For example, a randomized, double-blinded, placebo-controlled phase 3 study examined the risk of bone fractures among individuals receiving a cancer treatment called radium-223 (RA-223).[33] The risk was measured based on an independent data monitoring committee's recommendation following relatively more observations of fractures and deaths within the experimental arm.[34] This observation led to a decision made by the EMA's Pharmacovigilance Risk Assessment Committee to change the labeling indication for RA-223.[35] They recommended specific conditions for the market authorization of Xofigo (RA-223) and a non-interventional, post-authorization safety study (PASS) to further characterize the safety and efficacy of RA-223. The subsequent PASS study to evaluate the real-world risk of bone fractures further integrated observational data (Swedish Prostate Cancer Registry) to emulate the target trial. The study successfully met the EMA requirement and confirmed the higher risk of bone fractures consistent with earlier trials. The study concluded that differences in design and potential biases in observational settings compared to randomized trials can lead to variation in outcomes, underscoring the importance of carefully adjusting for confounding factors in trial emulations.[36]

The RCT-DUPLICATE demonstration project successfully replicated RCTs using the hypothetical target trial emulation approach and by integrating claims data.[37] The authors recognize that while it was not possible to achieve a perfect RCT emulation using secondary clinical data, using close observational analogues to design study pairs intentionally that may address similar and clearly defined research questions can create a path to reproducible RCT emulation using observational data. They further note, given their findings, that database studies can provide valuable complementary evidence and address important questions regarding real-world treatment effects that are not answerable by traditional RCTs.

General Data Considerations

Observational studies, while abundant and reflective of real-world conditions, lack the randomization that helps control for confounding variables in RCTs. Fit-for-purpose and high-quality RWD, in combination with a thoughtful, well-specified research question associated with well-designed study and appropriate analytics, can provide meaningful causal inference information.[38] Specifying detailed target trial protocols, including a clear description of eligibility criteria, interventions, outcomes, follow-up periods, causal contrasts, and analysis plans, helps align observational data with RCT standards and ensures the data's robustness for causal inference.[39] Data quality is

always critical; however, given observational and RCT data are collected for different purposes, clearly justifying the use of a specific RWD source(s) and describing it as fit-for-purpose is important.

Uses of observational data in causal inference studies require transparent identification and/or consideration of best practices. Adopting methodologies that meet regulatory standards, such as Good Clinical Practice (GCP), and following guidelines from regulatory bodies are fundamental to uphold data integrity. Protocol templates, such as the HARmonized Protocol Template and others, aim to enhance reproducibility and validity in RWE studies by providing structured templates that detail study designs and analysis plans.[40] Standardized protocol templates can provide transparency about the origins and content of data sources crucial for enhancing the reliability of data. Documenting data sources and characteristics alongside steps taken to produce a final analytical dataset can help uphold the credibility and reproducibility of findings.[41] Clear documentation helps stakeholders, including regulators, understand the context and limitations of the data, facilitating better decision making and enhancing the trustworthiness of the study results.

Detailed database characteristics on outcomes and key confounders can help support study validity. Additionally, the validity of study results should be examined through the lens of comparisons with eligibility criteria, patient characteristics, operational definitions, distributions of confounders and outcomes, study timing, etc. Therefore, consensus-driven operational definitions of data quality and other factors, such as relevance and reliability, pertaining to various RWD sources are ideal to ensure consistency, reliability, and reproducibility of study results and reduce unwanted variation in cohort selection and analytic methods used.[42] Likewise, prospective data collection validation rules and routine quality checks can be used to help mitigate biases arising from poor data quality and ensure data reliability.[43] Altogether, these steps would be essential to reach for high standards of data integrity that are communicable to regulators.

Distinguishing between inherent biases, such as those related to treatment assignments and discontinuations and those that can be controlled through improved data quality measures, is critical. Inherent biases often are related to the nature of observational studies where treatment assignments are not randomized and may be influenced by various patient characteristics and physician or health system preferences, and from discontinuations that occur due to adverse effects or lack of efficacy, which are not evenly distributed across treatment groups. Controllable biases should be addressed through rigorous data collection and statistical methods to adjust for confounding factors, while inherent biases need careful consideration during the study design phase to ensure they are properly accounted for.[44] Bias assessments and sensitivity analyses post-study also are helpful to verify the robustness and validity of study findings.

Current and Future Directions

Machine Learning Approaches

Machine learning (ML) approaches are considered a possible operational mechanism to strengthen causal inference research using RWD. Yet, limited examples or guidelines exist for its use. One such example is within the causal roadmap;[45] other opportunities exist to use ML techniques in the estimation stage, particularly as a method for pre-specified, data-adaptive adjustments for baseline confounders, determinants of intercurrent events, and loss to follow-up. However, some researchers argue that ML approaches should be seen as secondary, with a primary focus being consensus on foundational causal inference frameworks and statistical methods to reduce the likelihood of added complexity in the regulatory review process. Recent advances in statistical methods

for causal inference can be useful, such as doubly robust methods and G estimation) estimate treatment effects based on predicting exposures and outcomes for both the intervention and comparison group.[46]

Today, many strategies exist to estimate causal parameters and ML serves as a building block to estimate intermediate or nuisance parameters, given its power as an estimation tool to perform new methods for robust outcomes prediction and classification (e.g., estimates of average treatment effects). This potential is especially true when super-learner approaches are leveraged to help reduce bias from incorrect functional form by estimating several different exposure and outcome models. The targeted maximum likelihood estimation (TMLE) framework,[47, 48] one of many notable statistical examples, combines causal inference with ML methods. TMLE can be useful when used alongside checklist tools, such as the PALISADE Checklist that helps researchers assess, on a case-by-case basis, whether ML would add value to the research plan.

Instilling Trust in Observational Studies

Regulators will continue to maintain confidence in certain evidentiary standards when presented with observational study data to make a regulatory decision. However, room exists for growth to allow the scientific community to improve upon the quality and validity of their published work. A necessary goal for researchers exploring causality using RWD is to instill greater trust in observational studies. Pre-specification, protocol, and data transparency are important first steps. Researchers must identify strategies to prespecify when using observational data, including pre-registration of study protocols, for studies that are not part of a regulatory package. Such protocols can build analysis options that can be helpful in addressing unexpected data issues.

While pre-specifying key analyses in a protocol is possible, a need still exits for approaches to address the many unanticipated issues and questions that arise throughout a study. One such approach involves staging and clean room[49] (see glossary in Appendix D), constructs to safeguard result integrity and improve confidence in results from comparative analyses using RWD. Likewise, researchers publicly outlining their analysis approach would improve observational study and use case transparency to support current and future causal inference analyses. Initiatives, such as the International Society for Pharmacoeconomics and Outcomes Research's (ISPOR) RWE registry[50] and other principles for good practice,[51] promote transparent reporting of RWE studies and encourage researchers to publicly share their study protocols. This increased transparency can help build greater trust in leveraging RWE for causal inference. Further, some are calling[52] for the raising of the quality of published RWE and in part in response, journals are developing guidance documents for using causal language in observational data studies.

A necessary goal for researchers exploring causality using RWD is to instill greater trust in observational studies.

Conclusion

Integrating RCTs and RWE significantly enhances the capacity for causal inference in clinical research, bridging the gap between controlled experimental conditions and the variable realities of clinical practice. While RCTs provide robust causal insights under controlled settings, their stringent environments often fail to reflect the complexity and diversity of real-world settings. RWE can broaden the scope of causal inquiry, extending it to more varied populations and conditions. This synergy allows for a more comprehensive evaluation of how interventions perform across different environments, thus expanding the applicability and relevance of research findings. Drawing reliable causal inferences from RWE requires the adoption of rigorous methodologies, such as the target trial and estimand frameworks, which are designed to align observational studies with the methodological rigor of RCTs. The successful integration of RWE in causal inference relies heavily on maintaining stringent standards for study design and data analysis, including clear protocols, detailed data traceability, and statistical techniques to effectively control for biases inherent in observational data. Therefore, we recommend that current and future efforts to instill trust in observational studies to satisfy substantial evidence criterion and integrate meaningful RWE to achieve this goal consider these factors moving forward.

Ultimately, clinical data that is enriched with RWE gathered through non-interventional studies would promote RWE's capacity for causal inference, which is essential for advancing evidence-based health care.

Ultimately, clinical data that is enriched with RWE gathered through non-interventional studies would promote RWE's capacity for causal inference, which is essential for advancing evidence-based health care. By harmonizing the strengths of RWE with traditional research methods like RCTs, a more comprehensive understanding of health care interventions and their real-world impacts can be answered.

APPENDIX A
Generating and Leveraging RWE for Causal Inference Workstream members

Khaled Abdelgawwad
Bayer

Nicolas Ballaraini
Novartis

Lori Bash
Merck

Marc Berger
Independent Consultant

Elise Berliner
Cerner Enviza

Alina Bogdanov
Veradigm

Amanda Bruno
Bayer

Lauren Cain
Abbvie

Ulka Campbell
Aetion

Gianmario Candore
Bayer

Jennifer Christian
TargetRWE

Bill Crown
Brandeis University

Elvira D'Andrea
Abbvie

Eric Daza
Evidation

Anne Deitz
Merck

Lev Demirdjian
Janssen

Nancy Dreyer
Dreyer Strategies

Laura Fabbri
Chiesi

Vera Frajzyngier
Aetion

Liz Garry
Aetion

Kira Griffiths
Holmusk

Justin Guinney
Tempus

Lisa Hampson
Novartis

Joe Henk
UnitedHealthCare

Kathleen Hurwitz
TargetRWE

Ronald Ferney Herrera Clavijo
Boehringer-Ingelheim

Jack Ishak
Evidera

Soloman Iyasu
Retired

Dena Jaffe
Oracle Health

Linda Kalilani
GlaxoSmithKline

Nancy Lin
IQVIA

Ilya Lipkovich
Eli Lilly

Erlyn Macarayan
PatientsLikeMe

Martin Marciniak
Chiesi

Anne-Marie Meyer
University of North Carolina

Carrie Mills
Veradigm

May Mo
Amgen

Daniela Moga
University of Kentucky, ISPE

Aimee Near
IQVIA

Mayowa Oyseanya
Holmusk

John Page
Amgen

Emily Palmer
Holmusk

Ithan Peltan
Intermountain

Simone Pinheiro
Abbvie

Dave Pritchard
TargetRWE

Raj Punekar
Syneos Health

Lawrence Rasouliyan
OMNY Health

Shoshana Reshef
Teva Pharmaceuticals

Thibaut Sanglier
Genentech

Khaled Sarsour
Janssen

Debra Schaumberg
Evidera

Mohsin Shah
IQVIA

Deepshikha Singh
PatientsLikeMe

Rachel Sobel
ISPE

Ayse Tezcan
N-Power Medicine

David Thompson
Independent Consultant

Haijun Tian
Eli Lily

Darren Toh
Harvard

Brandon Webb
Intermountain

Leonie Williams
Holmusk

Dick Wilke
ISPOR

Zachary Zalman
PatientsLikeMe

Wei Zhou
ISPE

Nicole Zimmerman
GlaxoSmithKline

APPENDIX B
Duke Margolis 2023 Advisory Group Members

Marc Berger
Independent Consultant

Elise Berliner
Cerner Enviza

Barbara Bierer
Harvard University

Mac Bonafede
Veradigm

Brian Bradbury
Amgen

Jeffrey Brown
TriNetX

Adrian Cassidy
Novartis

Stella Chang
OMNY Health

William Crown
Brandeis University

Mark Cziraky
Carelon

EJ Daza
Evidation

Riad Dirani
Teva Pharmaceuticals

Nancy Dreyer
Independent Consultant

Andenet Emiru
University of California

Omar Escontrias
National Health Council

John Graham
GSK

Marni Hall
IQVIA

Morgan Hanger
Clinical Trials Transformation Initiative

Joe Henk
UnitedHealthCare

Stacy Holdsworth
Eli Lilly

Ryan Kilpatrick
Abbvie

Grazyna Lieberman
Regulatory Policy and Strategy Consultant

Lyn Macarayan
PatientsLikeMe

Christina Mack
IQVIA and ISPE

Anne-Marie Meyer
Independent Consultant

Megan O'Brien
Merck

Eleanor Perfetto
University of Maryland

Jeremy Rassen
Aetion

Stephanie Reisinger
Flatiron

Khaled Sarsour
Janssen

Debra Schaumberg
Evidera, part of PPD clinical research business, Thermo Fisher Scientific

Thomas Seck
Boehringer-Ingelheim

Lauren Silvis
Tempus

Montse Soriano Gabarro
Bayer

Michael Taylor
Genentech

David Thompson
Independent Consultant

Darren Toh
Harvard University

Alex Vance
Holmusk

Richard Willke
ISPOR

Bob Zambon
Syneos Health

APPENDIX C
Participants in the December 12th 2023 Workshop on Generating and Leveraging RWE for Causal Inference

Khaled Abdelgawwad

Xabier Garcia de Albeniz Martinez

Justyna Amelio

Nicolas Ballarini

Andrew Bean

Marc Berger

Jeoff Bispham

Susan Boklage

Jaclyn Bosco

Marie Bradley

Kim Brodovicz

Lauren Cain

Ulka Campbell

Gianmario Candore

Jie Chen

Stephanie Chiuve

Jennifer Christian

John Concato

Christopher Craggs

William Crown

Mark Cziarky

Issa Dahabreh

Elvira D'Andrea

Anne Deitz

Nancy Dreyer

Andenet Emiru

Vera Frajzyngier

Cindy Girman

Andres Gomez-Caminero

Lisa Hampson

Nick Honig

Jack Ishak

Ravi Iyer

Dena Jaffe

Madhuri Jerfy

Eeshan Khandekar

Ryan Kilpatrick

Carol Koro

Kajsa Kvist

Stephan Lanes

Catherine Lee

Jenni Li

Nancy Lin

Ilya Lipkovich

Wei Liu

Orsolya Lunacsek

Erlyn Macarayan

Nicole Mahoney

Chris Meister

Anne Marie Meyer

Pallavi S. Mishra-Kalyani

May Mo

Daniela Moga

Aimee Near

Dianne Paraoan

Ithan Peltan

Laura Perez

Silvia Perez-Vilar

Maya Petersen

Simone Pinheiro

Letizia Polito

David Pritchard

Moitur Rahman

Shoshana Reshef

Trejahn Rutlin

Thibault Sanglier

Debra Schaumberg

Sebastian Schneeweiss

Mohsin Shah

Kristin Sheffield

Jaime Smith

Kim Smith

Rachel Sobel

Ayse Tezcan

David Thompson

Haijun Tian

Darren Toh

Huan Wang

Shirley Wang

Lani Wegrzyn

Richard Wilke

Aprielle Wills

Nicole Zimmerman

APPENDIX D
Glossary

Clean room: An approach to increase confidence that the results from analyses of data from nonrandomized studies are valid, that restricts access to data and to preliminary study results, allowing for exploratory analyses to be conducted, while limiting knowledge for how these will affect the results of subsequent comparative safety and effectiveness analyses.[53]

Confounder: A variable whose presence affects the variables being studied so that the results do not reflect the actual relationship.[54]

Estimand: A description of the exact treatment effect a study aims to quantify.[55]

- **Causal Estimand:** A mathematical quantity that represents the answer to the causal question.[56]
- **Statistical estimand:** The causal estimand expressed the causal as a function of the observed data distribution.[57]

Immortal time bias: The error in estimating the association between the exposure and the outcome that results from misclassification or exclusion of time intervals.[58]

Intercurrent events: Post-baseline events (post-randomization events in randomized trials) that affect either the interpretation of outcome data (eg, treatment non-adherence or use of rescue treatment) or the existence of outcome data (eg, death if not already used as part of the outcome definition). Missing data or loss to follow-up are not intercurrent events.[59]

Selection bias: A bias that occurs when individuals or groups in a study differ systematically from the population of interest leading to a systematic error in an association or outcome.[60]

Staging: An approach to increase confidence that the results from analyses of data from nonrandomized studies are valid, that is a multi-step process during, which results from preliminary analyses are reviewed.[61]

References

[1] Food and Drug Administration, "Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products Guidance for Industry," March 2024.

[2] Nora Emmott et al., "Improving Patient Subgroup Representation with Real-World Data: Real World Efficacy and Patient Subgroups" (Duke Margolis Institute for Health Policy, September 28, 2023), https://healthpolicy.duke.edu/publications/improving-patient-subgroup-representation-real-world-data.

[3] Center for Drug Evaluation and Research, "Advancing Real-World Evidence Program," FDA (FDA, June 4, 2024), https://www.fda.gov/drugs/development-resources/advancing-real-world-evidence-program.

[4] Nirosha Mahendraratnam et al., "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 Institute for Health Policy, November 25, 2019).

[5] Virginia Acha et al., "Principles for Good Practice in the Conduct of Non-Interventional Studies: The View of Industry Researchers," *Therapeutic Innovation & Regulatory Science* 57, no. 6 (November 1, 2023): 1199–1208, https://doi.org/10.1007/s43441-023-00544-y.

[6] Shirley V. Wang, Sebastian Schneeweiss, and RCT-DUPLICATE Initiative, "Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials," *JAMA* 329, no. 16 (April 25, 2023): 1376–85, https://doi.org/10.1001/jama.2023.4221.

[7] Emmott et al., "Improving Patient Subgroup Representation with Real-World Data: Real World Efficacy and Patient Subgroups."

[8] Research, "Advancing Real-World Evidence Program."

[9] Taiwan Food and Drug Administration, "Taiwan Food and Drug Administration: Basic Considerations of Real-World Evidence in Support of Pharmaceutical Research and Development," July 2020.

[10] Canada's Drug and Health Technology Agency, Health Canada, and Mina Tadrous, "Guidance for Reporting Real-World Evidence," May 2023.

[11] Center for Devices and Radiological Health US Food and Drug Administration, "Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices" (FDA, June 26, 2023), https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices.

[12] Food and Drug Administration, "Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products Guidance for Industry."

[13] Food and Drug Administration.

[14] Brennan C. Kahan et al., "The Estimands Framework: A Primer on the ICH E9(R1) Addendum," *BMJ* 384 (January 23, 2024): e076316, https://doi.org/10.1136/bmj-2023-076316.

[15] Miguel A. Hernán et al., "Specifying a Target Trial Prevents Immortal Time Bias and Other Self-Inflicted Injuries in Observational Analyses," *Journal of Clinical Epidemiology* 79 (November 1, 2016): 70–75, https://doi.org/10.1016/j.jclinepi.2016.04.014.

[16] Paul Arora et al., "The Inflation Reduction Act: An Opportunity to Accelerate Confidence in Real-World Evidence in the United States," *Value in Health,* April 16, 2024, https://doi.org/10.1016/j.jval.2024.04.007.

[17] Miguel A. Hernán and James M. Robins, "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available," American Journal of Epidemiology 183, no. 8 (April 15, 2016): 758–64, https://doi.org/10.1093/aje/kwv254.

[18] Hernán Miguel A., "Methods of Public Health Research — Strengthening Causal Inference from Observational Data," *New England Journal of Medicine 385,* no. 15 (October 6, 2021): 1345–48, https://doi.org/10.1056/NEJMp2113319.

[19] Joshua D. Wallach et al., "Real-World Cardiovascular Outcomes Associated With Degarelix vs Leuprolide for Prostate Cancer Treatment," *JAMA Network Open* 4, no. 10 (October 1, 2021): e2130587, https://doi.org/10.1001/jamanetworkopen.2021.30587.

[20] Kellyn Arnold et al., "Application of the Target Trial Emulation Framework to External Comparator Studies," *Frontiers in Drug Safety and Regulation* 4 (2024), https://doi.org/10.3389/fdsfr.2024.1380568.

[21] Gonzalo Martínez-Alés et al., "Thromboprophylaxis with Standard-Dose vs. Flexible-Dose Heparin for Hospitalized COVID-19 Patients: A Target Trial Emulation," *Journal of Clinical Epidemiology* 151 (November 2022): 96–103, https://doi.org/10.1016/j.jclinepi.2022.08.006.

[22] International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, "ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials," December 2021.

[23] Kahan et al., "The Estimands Framework."

[24] Letizia Polito et al., "Applying the Estimand and Target Trial Frameworks to External Control Analyses Using Observational Data: A Case Study in the Solid Tumor Setting," *Frontiers in Pharmacology* 15 (January 26, 2024): 1223858, https://doi.org/10.3389/fphar.2024.1223858.

[25] Yanni Hao Lisa V. Hampson Jufen Chu, Aiesha Zia, Jie Zhang, Wei-Chun Hsu, Craig S. Parzynski and Evgeny Degtyarev, "Combining |the Target Trial and Estimand Frameworks to Define the Causal Estimand: An Application Using Real-World Data to Contextualize a Single-Arm Trial," *Statistics in Biopharmaceutical Research* 16, no. 1 (2024): 1–10, https://doi.org/10.1080/19466315.2023.2190931.

[26] Polito et al., "Applying the Estimand and Target Trial Frameworks to External Control Analyses Using Observational Data."

[27] Lauren E. Dang et al., "A Causal Roadmap for Generating High-Quality Real-World Evidence," *Journal of Clinical and Translational Science* 7, no. 1 (n.d.): e212, https://doi.org/10.1017/cts.2023.635.

[28] Dang et al.

[29] Food and Drug Administration, "Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products Guidance for Industry." March 2024

[30] European Medicines Agency, "Reflection Paper on the Use of Real-World Data in Non Interventional Studies to Generate Real-World Evidence," April 15, 2024.

[31] European Medicines Agency.

[32] European Medicines Agency.

[33] Matthew Smith et al., "Addition of Radium-223 to Abiraterone Acetate and Prednisone or Prednisolone in Patients with Castration-Resistant Prostate Cancer and Bone Metastases (ERA 223): A Randomised, Double-Blind, Placebo-Controlled, Phase 3 Trial," *The Lancet Oncology* 20, no. 3 (March 1, 2019): 408–19, https://doi.org/10.1016/S1470-2045(18)30860-X.

[34] Smith et al.

[35] Pär Stattin et al., "Real World Outcomes in Patients With Metastatic, Castration-Resistant Prostate Cancer Treated With Radium-223 in Routine Clinical Practice in Sweden," *Clinical Genitourinary Cancer* 21 (September 2022), https://doi.org/10.1016/j.clgc.2022.09.002.

[36] Smith et al., "Addition of Radium-223 to Abiraterone Acetate and Prednisone or Prednisolone in Patients with Castration-Resistant Prostate Cancer and Bone Metastases (ERA 223)."

[37] Wang, Schneeweiss, and RCT-DUPLICATE Initiative, "Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses."

[38] Hernán Miguel A., "Methods of Public Health Research — Strengthening Causal Inference from Observational Data."

[39] Hernán Miguel A.

[40] Shirley V. Wang et al., "HARmonized Protocol Template to Enhance Reproducibility of Hypothesis Evaluating Real-World Evidence Studies on Treatment Effects: A Good Practices Report of a Joint ISPE/ISPOR Task Force," *Value in Health* 25, no. 10 (October 1, 2022): 1663–72, https://doi.org/10.1016/j.jval.2022.09.001.

[41] Martínez-Alés et al., "Thromboprophylaxis with Standard-Dose vs. Flexible-Dose Heparin for Hospitalized COVID-19 Patients."

[42] William Crown et al., "Can Observational Analyses of Routinely Collected Data Emulate Randomized Trials? Design and Feasibility of the Observational Patient Evidence for Regulatory Approval Science and Understanding Disease Project," *Value in Health,* August 13, 2022, https://doi.org/10.1016/j.jval.2022.07.003.

[43] Wang, Schneeweiss, and RCT-DUPLICATE Initiative, "Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses."

[44] Hernán and Robins, "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available."

[45] Dang et al., "A Causal Roadmap for Generating High-Quality Real-World Evidence."

[46] Byeong Yeob Choi, Chen-Pin Wang, and Jonathan Gelfond, "Machine Learning Outcome Regression Improves Doubly Robust Estimation of Average Causal Effects," *Pharmacoepidemiology and Drug Safety* 29, no. 9 (September 2020): 1120–33, https://doi.org/10.1002/pds.5074.

[47] William V. Padula et al., "Machine Learning Methods in Health Economics and Outcomes Research—The PALISADE Checklist: A Good Practices Report of an ISPOR Task Force," *Value in Health* 25, no. 7 (July 1, 2022): 1063–80, https://doi.org/10.1016/j.jval.2022.03.022.

[48] Megan S. Schuler and Sherri Rose, "Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies," *American Journal of Epidemiology* 185, no. 1 (January 1, 2017): 65–73, https://doi.org/10.1093/aje/kww165.

[49] Paul Muntner et al., "Staging and Clean Room: Constructs Designed to Facilitate Transparency and Reduce Bias in Comparative Analyses of Real-World Data," *Pharmacoepidemiology and Drug Safety* 33, no. 3 (2024): e5770, https://doi.org/10.1002/pds.5770.

[50] ISPOR, "Real-World Evidence Registry," accessed May 21, 2024, https://www.ispor.org/strategic-initiatives/real-world-evidence/real-world-evidence-registry.

[51] Acha et al., "Principles for Good Practice in the Conduct of Non-Interventional Studies."

[52] Issa J. Dahabreh and Kirsten Bibbins-Domingo, "Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals," *JAMA,* May 9, 2024, https://doi.org/10.1001/jama.2024.7741.

[53] Muntner et al., "Staging and Clean Room."

[54] Mohamad Amin Pourhoseingholi, Ahmad Reza Baghestani, and Mohsen Vahedi, "How to Control Confounding Effects by Statistical Analysis," *Gastroenterology and Hepatology from Bed to Bench* 5, no. 2 (2012): 79.

[55] Kahan et al., "The Estimands Framework."

[56] Dang et al., "A Causal Roadmap for Generating High-Quality Real-World Evidence."

[57] Dang et al.

[58] Kabir Yadav and Roger J. Lewis, "Immortal Time Bias in Observational Studies," *JAMA* 325, no. 7 (February 16, 2021): 686–87, https://doi.org/10.1001/jama.2020.9151.

[59] Kahan et al., "The Estimands Framework."

[60] D Nunan, C Bankhead, and JK Aronson, "Catalogue of Bias Collaboration: Selection Bias," 2017.

[61] Muntner et al., "Staging and Clean Room."