


Comparative Effectiveness Research Using Randomized Trials and Observational Studies: Validity and Feasibility Considerations

Behnood Bikdeli^{1,2,3} Joseph S. Ross^{3,4} Syed Bukhari⁵ Molly M. Jeffery^{6,7,8} Gregory Y.H. Lip^{9,10} 
Seng Chan You¹¹ David J. Cohen^{12,13} James L. Januzzi Jr^{14,15} Joshua D. Wallach¹⁶

¹Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States

²Thrombosis Research Group, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States

³YNHH/Yale Center for Outcomes Research and Evaluation (CORE), New Haven, Connecticut, United States

⁴Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, United States

⁵Department of Cardiology, Johns Hopkins University, Baltimore, Maryland, United States

⁶Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota, United States

⁷Division of Health Care Delivery Research, Mayo Clinic, Rochester, Minnesota, United States

⁸Department of Emergency Medicine, Mayo Clinic, Rochester, Minnesota, United States

⁹Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart and Chest Hospital, Liverpool, United Kingdom

Address for correspondence Behnood Bikdeli, MD, MS, Cardiovascular Medicine Division, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, United States (e-mail: bbikdeli@bwh.harvard.edu; Behnood.bikdeli@yale.edu).

¹⁰Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

¹¹Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

¹²Cardiovascular Research Foundation, New York, New York, United States

¹³St Francis Hospital and Heart Center, Roslyn, New York, United States

¹⁴Division of Cardiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States

¹⁵Baim Institute for Clinical Research, Boston, Massachusetts, United States

¹⁶Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, United States

Thromb Haemost 2026;126:335–345.

Abstract

Keywords

- ▶ randomized controlled trials
- ▶ comparative effectiveness
- ▶ target trial framework
- ▶ thrombosis
- ▶ validity

In comparative effectiveness research (CER), ensuring internal, construct, and external validity is crucial. Internal validity determines whether observed outcomes are causally linked to an intervention; construct validity assesses whether a study measures what it intends to; and external validity relates to generalizability in routine practice. Double-blind randomized trials optimize internal validity by minimizing bias and confounding, while construct validity is strengthened through pre-specified protocols and standardized data collection. However, controlled conditions limit external validity. Pragmatic RCTs improve generalizability but may compromise internal validity due to open-label designs. Observational CER studies—including observational studies following the target trial emulation framework—offer broader external validity and feasibility in less time and at lower cost. However, due to lack of random assignment, these studies are susceptible to measured and unmeasured confounding. Several techniques help

The review process for this paper was fully handled by Christian Weber, Editor in Chief.

received

May 12, 2025

accepted after revision

July 23, 2025

accepted manuscript online

July 24, 2025

article published online

August 8, 2025

© 2025. Thieme. All rights reserved.

Georg Thieme Verlag KG,
Oswald-Hesse-Straße 50,
70469 Stuttgart, Germany

DOI <https://doi.org/10.1055/a-2664-7887>.
ISSN 0340-6245.

mitigate these concerns, including a detailed pre-specified protocol, tools such as propensity score matching to balance measured confounders, falsification endpoint testing for assessing the presence of unmeasured confounders, and quasi-experimental designs (including instrumental variable analysis), which may be able to address both. Pre-specified sensitivity analyses and triangulation with complementary data sources further enhance robustness. Construct validity in observational CER depends on accurate patient profiling and validated computational phenotypes for identifying patients, exposures, and outcomes. Thoughtful study design and analytic rigor are essential for balancing these validity considerations. This brief review highlights these issues with examples from thrombosis research.

Introduction: Design Options for Comparative Effectiveness Research

Continued evaluation of the effectiveness and safety of health interventions, as well as the assessment of potential off-label use of therapies, requires rigorous comparative effectiveness research (CER) studies. Such evidence can be generated using a range of study designs, including blinded clinical trials; pragmatic, often open-label clinical trials that may be registry-based or linked to electronic health record (EHR) data;¹ and observational comparative effectiveness and safety studies designed to emulate hypothetical or existing target trials. However, key considerations must be addressed, including the feasibility and resource requirements for designing and conducting these studies, as well as ensuring their validity to make the findings relevant and impactful for clinical care. This brief review summarizes the utility and challenges of various types of CER studies for generating clinical evidence. It highlights when each design may be appropriate and reasonable, discusses key validity considerations for appropriate interpretation to inform patient care, and provides illustrative examples in the domain of thrombosis research.

Randomized Trials for Comparative Effectiveness Research: Internal, Construct, and External Validity Considerations

Randomized controlled trials (RCTs) are often considered the gold standard regarding CER. Random assignment of interventions helps balance both measured and unmeasured factors across trial groups, while blinding of the intervention minimizes post-assignment biases in treatment administration, testing procedures, and outcome reporting. These two factors strengthen the internal validity, which refers to the extent to which the observed outcomes are causally attributable to the intervention (exposure) rather than other factors.

Preparing a pre-specified protocol and monitoring adherence to it, prospectively capturing and auditing data using pre-defined elements and instruments, and adjudicating outcomes can all improve construct validity. This type of

validity refers to the extent to which the study population, the interventions and adherence to them, and clinical outcome definitions and execution measure what they are intended to measure.

Double-blind RCTs often lack external validity, which refers to the generalizability of the findings to related populations and settings outside the exact conditions of the study. In particular, RCTs often have stringent eligibility criteria and detailed protocols for follow-up, which are costly, time-consuming, and challenging to generalize to patients in routine care settings who typically have complex comorbidities and limited resources for treatment and follow-up. In addition, RCTs may underrepresent certain racial or ethnic groups, which is important given known variations in clinical outcomes of thrombotic diseases between populations, such as White patients versus those of Asian descent.² Moreover, the delivery of interventions in RCTs may or may not reflect routine clinical practice. For example, the mAFA-II trial³ implemented holistic atrial fibrillation management using a mobile health platform based on the Atrial Fibrillation Better Care (ABC) pathway. The feasibility of this approach in routine care settings was later supported by the MIRACLE-AF trial, which successfully delivered the ABC pathway through village doctors supported by telehealth in rural communities.⁴ In other settings, however, complex interventions or those that require extensive resources may not be actionable in resource-limited environments.

Simple pragmatic RCTs may more accurately reflect patients in routine practice. Registry-based trials are often less costly, can be pragmatic, and have the capacity to test representativeness in routine practice through access to the parent registry. However, they are often open-label and may have slightly lower internal validity due to less stringent control measures. Achieving adequate representation of patients in RCTs is a complex issue that may require solutions beyond trial design alone. For example, female representation is low in many clinical trials and some studies suggest that having a diverse trial leadership team helps improve diverse patient representation for a variety of reasons.⁵

Beyond limitations in representation, inclusiveness, and retention of participants across demographic subgroups,

RCTs may be unfeasible in many scenarios due to the rarity of the condition, ethical challenges, or prohibitive costs and resource requirements. Furthermore, evidence from RCTs is often not durable over time.^{6,7} For instance, the DAPT trial established evidence for prolonged dual antiplatelet therapy after drug-eluting stent implantation.⁸ However, subsequent analyses showed different results when reweighted to represent contemporary patient populations and use of second-generation drug-eluting stent technology.⁹ In addition to potential changes in clinical characteristics over time, improved implantation techniques, facilitated by use of intravascular imaging have reduced the risk of adverse events such as stent thrombosis and thus reduced the potential benefit from prolonged dual antiplatelet therapy.^{10,11}

CER Using Observational Data

The above-mentioned limitations of traditional RCTs, combined with the increasing availability and granularity of data sources based on routine practice (real world), including EHR data and administrative claims, and advances in design and analytic methodology, have stimulated interest in observational methods to assess the effectiveness and safety of healthcare interventions. The availability of a broader range of data elements, such as national claims data linked with prescriptions and death registries, patient-level registries, and rich multicenter EHR data, allows for the inclusion of variables not typically available in most claims databases, such as reports of imaging studies. This can enhance patient profiling and improve both internal validity—by adjusting for confounders—and construct validity—by providing more accurate profiling of patients and outcomes.

Off-label use represents another area where observational CER studies can provide valuable insights when RCT evidence is lacking. For example, for decades, metformin was contraindicated in patients with renal impairment due to concerns about lactic acidosis, limiting its use in diabetic patients who often have concurrent kidney disease. However, multiple observational studies demonstrated the safety of metformin in mild to moderate chronic kidney disease, and the FDA revised labeling for metformin to permit its use in patients with mild–moderate kidney impairment.¹² In contrast, studies of direct oral anticoagulants (DOACs) in patients with rheumatic heart diseases highlight potential limitations of observational data. Despite promising initial observational findings,¹³ RCTs such as INVICTUS indicated harm,^{14,15} highlighting the need for cautious interpretation of observational evidence. Particularly, in this example, the observational CER study lacked robust foundations for construct validity in patient selection, exposure definition, and outcome measurement, and it did not account for unmeasured confounders.¹³ With stronger methods, the results may have more closely approximated those of an RCT.

Observational Studies Following the Target Trial Emulation Framework: Strengths, Limitations, and Validity Considerations

The target trial emulation framework^{16,17} has been increasingly used to guide observational CER studies. The framework, which provides a structured process to mimic the design and analyses of RCTs, is intended to reduce biases when conducting research using observational data by drafting a pre-specified protocol that establishes the eligibility criteria, exposure variables (substitutes for interventions in RCTs), outcome variables, follow-up period, and analytical approaches and to make it available to reviewers and readers. These elements can help mitigate the impact of confounders to improve internal validity, particularly compared with other forms of observational CER studies, and improve transparency related to adherence to pre-specified methods.

Observational CER studies that follow the target trial emulation framework can allow for inference in areas where RCTs are nonexistent due to difficulties with trial design or conduct, can complement RCTs by generating effectiveness and safety data among patients who may have been routinely excluded from RCTs,^{18–24} and can be conducted at a fraction of the time and cost of a traditional RCT. However, observational CER studies, including those that follow the target trial emulation framework, have inherent limitations that are necessary to consider during study design, analysis, and reporting, and interpretation at bedside.

With respect to internal validity, the exposure variables (interventions) in observational CER studies that follow the target trial framework are not randomly assigned, and baseline characteristics are less likely to be balanced. Several methods, including multivariable regression and propensity-score weighting or matching, can help adjust for measured confounders. The list of variables should not be selected based on statistical criteria only. In particular, it is necessary to use an informed approach based on theoretical models (e.g., directed acyclic graphs) that consider the relationships between exposures and outcome. These models should be informed by clinical content experts who have good appreciation of measured and unmeasured factors that may impact the outcomes. Sensitivity analyses and subgroup analyses can help strengthen the robustness of the results, and falsification endpoints might partially alleviate the concern for unmeasured confounders. However, residual and unmeasured confounding remain possibilities. For instance, administrative claims data are often unable to account for important patient and prescriber behaviors, such as choices between various therapies or no treatment, which could influence the observed findings. Instrumental variable analysis is a technique used to address both measured and unmeasured confounders in observational CER studies. It relies on identifying a valid

instrument—a variable that is associated with the exposure variable but influences the outcome variable only through its effect on the exposure. However, the challenge with this approach lies in finding an instrument that meets the necessary assumptions for validity.

Beyond these issues, true *intention-to-treat* analyses, where participants are analyzed according to their assigned treatment rather than actual treatment to provide conservative unbiased estimates, are not feasible in observational CER studies. An initiators (as-started) analysis, though not based on random assignment, provides a partial approximation,^{16,25} and a sustained treatment effect analysis is proposed as an approximation of as-treated analysis of RCTs.¹⁶ Finding a suitable *no treatment* comparator group is also a major challenge in observational CER studies that follow the target trial framework due to concerns for confounding and selection bias. Biases impacting internal validity can also arise in subsequent care (performance bias), subsequent assessment (detection bias), or follow-up duration (attrition bias) due to lack of blinding. Finally, in RCTs, and particularly observational CER studies that follow the target trial framework, clustering of observations (e.g., within sites) can be addressed to avoid biases assessment of the treatment effect of the interventions on outcomes of interest.

To ensure construct validity in observational CER studies, it is crucial to develop computable phenotypes using data from administrative claims, such as International Classification of Diseases (ICD) codes and Current Procedural Terminology (CPT) codes, in ways that have been validated against “ground truth” (which can be approximated from clinical information in medical charts).^{26,27} This approach helps accurately identify interventions and outcomes, minimizing the risk of misidentifying patients or producing false positive or negative results. Furthermore, there may be patient and clinical characteristics ascertained in clinical trials that cannot be matched exactly in claims or EHR data, including cancer staging or imaging parameters.²⁸ As for outpatient drugs, claims data can tell us if a prescription was filled but will not inform us whether the patient actually took the drug. In RCTs, this is typically complemented by drug diary and drug-level testing.

A key strength of observational CER studies that follow the target trial framework, especially those conducted using routine practice data, is their generalizability (external validity). These studies can capture diverse patient populations with respect to age, sex, race, ethnicity, location, and treatment settings. However, some data sources may still have limited external validity, such as Department of Veteran Affairs data, which overrepresent male individuals. As an increasing number of large, compiled datasets are also being curated, ascertaining the provenance of these data sources and understanding which patients and data elements they represent would be essential for utilizing them for CER analyses.

Practical Considerations for Clinical Investigators Conducting CER Studies

To effectively leverage RCTs or observational CER studies to address clinical questions, it is crucial to understand their inherent characteristics and to enhance the quality of modifiable components that influence internal validity (►Fig. 1). At the design phase, the tradeoffs between internal and external validity for different forms of RCTs and observational CER studies should be considered to select the appropriate design (►Table 1). Approximating the standards of RCTs using the target trial framework in observational CER studies can strengthen the design and internal validity. As for construct validity, standard procedures applied in many RCTs including prospective screening, core laboratories, and blinded event adjudication help maximize the construct validity for various data elements. In observational CER studies using claims data, efforts should be made to use validated algorithms to ascertain disease conditions, exposure variables, and outcomes (►Tables 2 and 3). In studies that use EHR data, including both registry-based RCTs and observational CER studies, incorporation of validated natural language processing tools and large language models may help automate the extraction of detailed complex data elements to improve construct validity and efficiency.^{26,29}

At the level of reporting the results in scientific manuscripts, RCTs are familiar to many readers. However, regarding observational CER studies that follow the target trial framework, the use of the term *trial* and other RCT-related terminology may mislead some readers into assuming the study design involves actual randomization. Therefore, it would be beneficial for manuscripts reporting results from CER studies based on the target trial emulation framework to clearly specify the observational design used in these studies.^{30,31}

Heterogeneity of Treatment Effect in CER Studies

Heterogeneity of treatment effect (HTE) refers to the variation in the magnitude or direction of a treatment’s effects across different subgroups or individuals. If there is a high degree of HTE among certain subgroups in a clinical trial, the external validity of the results may be at risk. Failing to consider HTE can also mask a true treatment effect that is restricted to specific subgroups. For example, if a stroke trial includes a wide range of patients, most of whom are enrolled within 24 hours of stroke onset, the overall treatment effect of fibrinolytic therapy may appear limited. However, subgroup-specific analyses are likely to reveal a distinct treatment effect in patients who present within the first 3 hours of symptom onset. Similar principles can apply to observational CER studies.

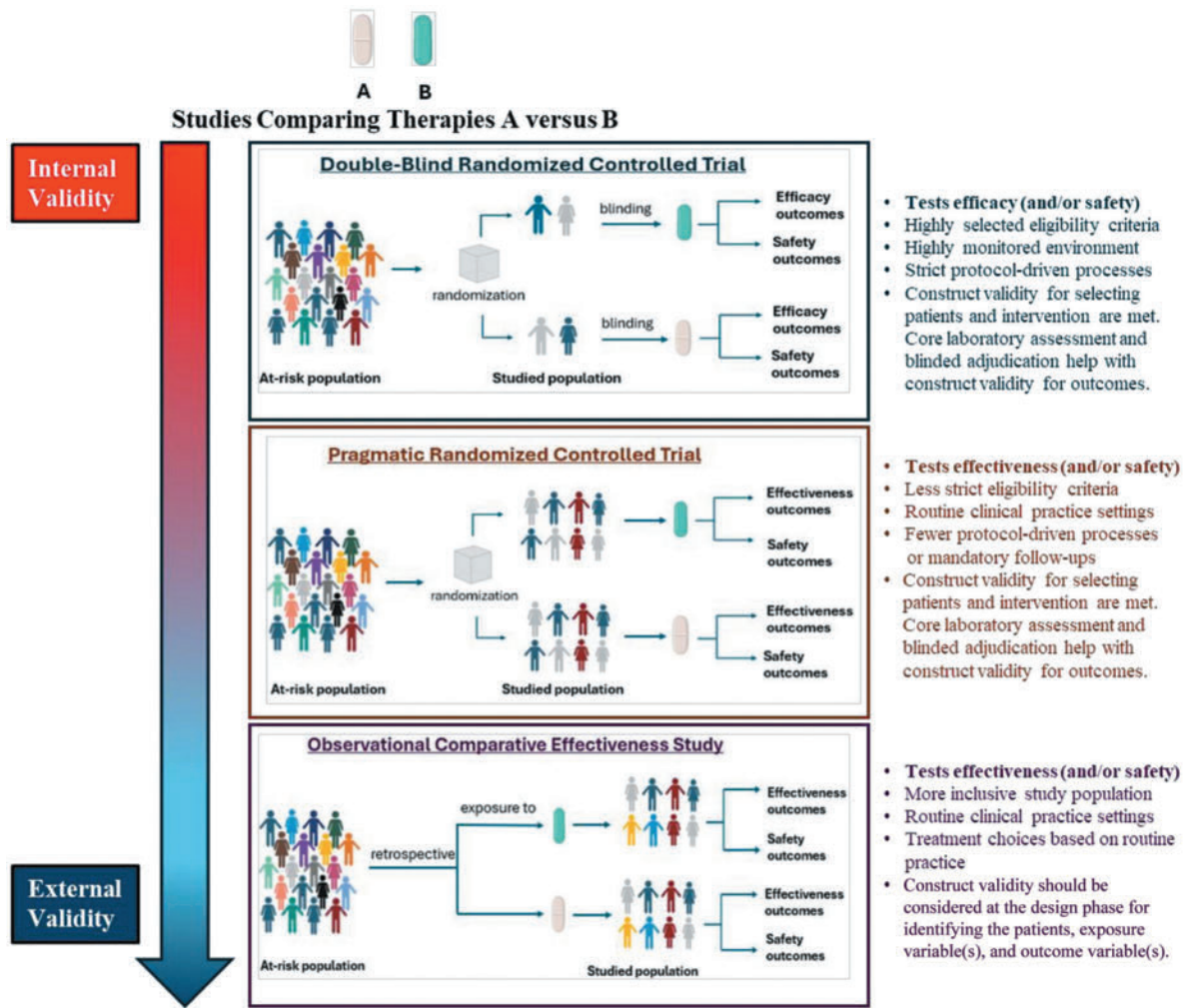


Fig. 1 Comparison of randomized clinical trials (RCTs), pragmatic clinical trials (PCTs), and observational comparative effectiveness research (CER). This figure provides a conceptual comparison of RCTs, PCTs, and observational CER across key dimensions, including study design, clinical setting, patient eligibility, treatment allocation and follow-up, and the emphasis on internal versus external validity. Although these categories reflect general methodological principles, individual studies may vary. For example, observational CER can incorporate rigorous inclusion criteria or leverage prospective registry data, and some explanatory RCTs may adopt more flexible protocols depending on study objectives. In addition, although pharmacologic interventions are depicted here for illustrative purposes, these principles broadly apply to other types of health interventions evaluated in clinical research. Many elements of construct validity are already fulfilled in RCTs since the patients are selected prospectively and the intervention is being randomly assigned prospectively. For outcome ascertainment, core laboratories (for blood, imaging, or other biomarkers) and blinded adjudication (for clinical outcomes) help improve the construct validity. Observational CER studies need detailed pre-specified plans for construct-validated approaches for selecting patients, exposure variables, and outcome variables. Created in BioRender. Hays, A. (2025) <https://BioRender.com/wdvcf1j>.

Practical Considerations for Implementation of Results in Practice

Both RCTs and observational CER studies can vary in methodological and reporting quality. Therefore, it would be important for the readers to familiarize themselves with essential components of successful CER studies. Although not all clinicians are research methods experts, they should review the CER studies to understand that there was a clear central question, that the question was pre-specified, that appropriate tools were implemented to identify patients and to assess outcomes (construct validity), that appropriate measures were taken to minimize bias and confounding in providing a causal inference for the intervention toward the outcome of interest (internal

validity), and that the findings are reasonably applicable to patients whom they see in clinical practice (external validity). ▶ **Tables 2 and 3** share illustrative summaries in thrombosis research both for observational CER studies and RCTs.

Conclusion

Both RCTs and observational CER studies play vital roles in advancing our understanding of the effectiveness and safety of health interventions. As the field evolves, besides considerations for cost and feasibility, careful attention to internal, construct, and external validity will be essential for the design, conduct, and interpretation of high-quality CER studies.

Table 1 Summary of design, execution, and validity trade-offs in comparative effectiveness research

Feature	Double-blind randomized controlled trial	Pragmatic randomized controlled trial (e.g., registry-based)	Observational emulation of target trial
Internal validity	++++	+++	+++
Construct validity	++++	+++	+++
External validity	++	++++	++++
Cost	++++	+++	++
Resource intensity	++++	+++	++
Expertise required ^a	++++	++++	++++
Ethical complexity ^b	+++	++	+
Ability to provide detailed phenotyping	++++	+++	+++

Notes: Plus signs (+) represent a consensus-based semiquantitative scale indicating the relative magnitude or intensity of each feature (e.g., cost, validity, expertise) across study types. A higher score does not necessarily indicate greater desirability—for example, higher cost (“++++”) is less desirable than lower cost (“+”).

“+” = very low/minimal; “++” = low/modest; “+++” = moderate; “++++” = high; “+++++” = very high.

^aAll comparative effectiveness study designs need sufficient expertise. For clinical trials, the domains of expertise include trial design, clinical research methods, biostatistics, and leadership skills. For observational comparative effectiveness studies, in addition to clinical research methods and biostatistics, expertise is often required in data science and causal inference using observational data.

^b*Ethical complexity* refers to the degree of ethical complexity inherent to each study design. For randomized controlled trials, especially placebo-controlled or double-blind trials, ethical considerations may include withholding potentially beneficial treatments or subjecting participants to investigational therapies. Pragmatic trials often pose fewer ethical challenges as they are conducted in routine clinical care settings with treatments already in clinical use. Observational studies, including comparative effectiveness research based on routine care data, generally involve the least ethical complexity, as there is no experimental intervention or randomization, though issues related to data privacy, biospecimens, and informed consent may still apply.

Conflict of Interest

Outside the submitted work, B.B. was supported by a Career Development Award from the American Heart Association and VIVA Physicians (#938814). B.B. was supported by the Scott Schoen and Nancy Adams IGNITE Award and is supported by the Mary Ann Tynan Research Scientist award from the Mary Horrigan Connors Center for Women’s Health and Gender Biology at Brigham and Women’s Hospital, and was supported by the Heart and Vascular Center Junior Faculty Award from Brigham and Women’s Hospital. B.B. reports that he was a consulting expert, on behalf of the plaintiff, for litigation related to two specific brand models of IVC filters. B.B. has not been involved in the litigation from 2022 to 2025 nor has he received any compensation from 2022 to 2025. B.B. reports that he is a member of the Medical Advisory Board for the VasculLearn Network, and serves in the Data Safety and Monitoring Board of the NAIL-IT trial funded by the National Heart, Lung, and Blood Institute, and Translational Sciences. B.B. is a collaborating consultant with the International Consulting Associates and the US Food and Drug Administration in a study to generate knowledge about utilization, predictors, retrieval, and safety of IVC filters. B.B. receives compensation as an Associated Editor for the *New England Journal of Medicine*, *Journal Watch Cardiology*, as an Associate Editor for *Thrombosis Research*, and as an Executive Associate Editor for *JACC*, and is a Section Editor for *Thrombosis and Haemostasis* (no compensation). S.C.Y. reports grants

from Daiichi Sankyo. He is a coinventor of granted Korea Patent DP-2023-1223 and DP-2023-0920, and pending Patent Applications DP-2024-0909, DP-2024-0908, DP-2022-1658, DP-2022-1478, DP-2022-1365, PATENT-2025-0039190, PATENT-2025-0039191, PATENT-2025-0039192, PATENT-2025-0039193, and PATENT-2025-0039194 unrelated to current work. S.C.Y. is a chief executive officer of PHI Digital Healthcare. D.J.C. reports institutional research grant support from Edwards Lifesciences, Boston Scientific, Abbott, Medtronic, Corvia, Cathworks, Philips, Zoll Medical, I-Rhythm, JenaValve, and ANCORA as well as consulting income from Medtronic, Edwards Lifesciences, Boston Scientific, Abbott, Zoll Medical, and Elixir Medical. J.L.J. is supported in part by the Adolph Hutter Professorship at Harvard Medical School. J.L.L. reports a board position with Imbria Pharma and equity in Jana Care. He is a Deputy Editor at *JACC* and receives current/recent grant support from Abbott, AstraZeneca, BMS, HeartFlow, and Novartis Pharmaceuticals, consulting income from Abbott Diagnostics, AstraZeneca, Beckman-Coulter, Boehringer Ingelheim, Eli Lilly, Janssen, Novartis, Prevencio, Quidel, and Roche Diagnostics, and serves on clinical endpoint committees/data safety monitoring boards for Abbott, AbbVie, Amgen, CVRx, Medtronic, Pfizer, and Roche Diagnostics. J.S.R. currently receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Food and Drug Administration for the Yale-Mayo Clinic Center for

Table 2 Illustrative examples of thrombosis-related observational comparative effectiveness study and its validity considerations

Studies	Clinical scenario	Pre-specified available protocol	Construct-validated approach for patient selection	Construct-validated approach for exposure variable selection (proxy for intervention)	Construct-validated approach for outcome ascertainment	Detailed and reasonable analytical plan to address measured confounders and P-hacking	Reasonable plan to address unmeasured confounders	Appropriate disclosure of strengths and limitations	External validity considerations	Conclusion about the study
You et al ²²	In patients with acute coronary syndrome treated with percutaneous coronary intervention, does ticagrelor, compared with clopidogrel, result in different rates of ischemic and hemorrhagic events?	Yes Pre-specified available protocol publicly shared as a supplement	Yes ICD-9 and ICD-10 codes validated in a sub-sample of the study used for identifying ACS	Yes Adopted a previously validated prescription record method, although prescription is not necessarily accounting for adherence	Yes Restricted outcome to primary diagnoses; used "blanking period" rule to exclude duplicate diagnoses	Yes Incorporated multiple propensity score methods; prespecified a statistical analysis protocol to address p-hacking	Yes Multiple falsification endpoint tests performed	Yes	Yes Use of diverse, multinational databases to show the consistency of the results with triangulation	Strong study design with a robust methodology, without major methodological flaws
REAL-PE study ³²	In patients with pulmonary embolism, how do safety outcomes, including major bleeding and intracranial hemorrhage, compare between ultrasound-assisted catheter-directed thrombolysis and mechanical thrombectomy?	No	No Used EHR-based codes without employing validated methods for the PE diagnosis or procedure exposure	No Relied solely on device identifiers mapped from routine coding without a construct-validation process	No Used ICD-10, SNOMED, CPT codes, and laboratory data to define outcomes without incorporating explicit validation steps	No No specific strategies to mitigate P-hacking or adjust for multiple comparisons	No No falsification endpoint testing, instrumental variable analysis, or quantitative bias analysis performed	No Inclusion of thrombectomy patients from a period predating the device's commercial availability was not appropriately disclosed as a major limitation	No Use of a single U.S.-based EHR platform without further validation in other data sources	Clinically interesting question and design but facing several methodological limitations
Cohen et al ³³	In patients with venous thromboembolism and chronic kidney disease, how do the effectiveness and safety outcomes compare between apixaban and warfarin?	No	No Referred to previous literature validating ICD codes for VTE and CKD	No Used pharmacy claims data without listing specific National Drug Codes or validating the selection of the codes	No Used ICD-9-CM and ICD-10-CM codes for outcomes without incorporating explicit validation steps	No Used stabilized inverse probability of treatment weighting to address measured confounders, but lacked a prespecified analysis protocol to address p-hacking	No No falsification endpoint testing, instrumental variable analysis, or quantitative bias analysis performed	Yes Yes	No Use of CMS Medicare and Medicaid databases excludes Medicaid and uninsured population, and limits generalizability	Clinically interesting design but facing several methodological limitations
Chowdhury et al ³⁴	In patients with non-valvular AF and diabetes, how do the effectiveness and safety outcomes compare between apixaban and rivaroxaban?	Yes Pre-specified available protocol shared as a supplement	No Study neither validated their coding approach nor explicitly referred to prior validation work	No Used prescriptions dispensed by practitioner without validation approach	No Used ICD-10 codes for outcomes without explicitly validating the codes	Yes Incorporated propensity score methods, and a prespecified statistical analysis protocol	No No falsification endpoint testing, instrumental variable analysis, or quantitative bias analysis performed	Yes	No Use of single country (UK) database that lacks detailed racial/ethnic identifiers	Clinically interesting design but facing some methodological limitations

(Continued)

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 2 (Continued)

Studies	Clinical scenario	Pre-specified available protocol	Construct-validated approach for patient selection	Construct-validated approach for exposure variable selection (proxy for intervention)	Construct-validated approach for outcome ascertainment	Detailed and reasonable analytical plan to address measured confounders and P-hacking	Reasonable plan to address unmeasured confounders	Appropriate disclosure of strengths and limitations	External validity considerations	Conclusion about the study
Weycker et al ³⁵	In outpatient treatment of pulmonary embolism, how do the effectiveness and safety outcomes compare between apixaban and warfarin?	Yes Pre-specified available protocol shared as a supplement	No Study neither validated their coding approach nor explicitly referred to prior validation work	No Exposure was determined based on National Drug Codes, but did not validate this approach nor referred to validated definitions for drug exposure in administrative data	No Outcomes identified using ICD-9 and ICD-10 codes without validating these codes or citing prior validation studies	Yes Incorporated 1:1 propensity score matching, and had a prespecified statistical analysis protocol	No No falsification endpoint testing, instrumental variable analysis, or quantitative bias analysis performed	Yes	No Use of commercial insurance databases that excludes uninsured population and lacks racial/ethnic identifiers, stringent exclusion criteria	Clinically interesting design but facing some methodological limitations
Butala et al ³⁶	In adults undergoing transfemoral TAVR, does the use of an embolic protection device compared to no device reduce the risk of in-hospital stroke?	Yes Prespecified outcomes and analytic methods defined and applied	No Patient selection from Transcatheter Valve Therapy registry with clinical information available	No The exposure variable, embolic protection device, was an existing data element in the registry. Although details not specified in the manuscript, the registry is known to have regular audits.	No While the outcomes were based on established definitions, the outcomes were not explicitly validated for the study	Yes Incorporated propensity score overlap weighting, instrumental variable analysis, and a prespecified statistical analysis protocol	Yes Instrumental variable analysis performed to address potential unmeasured confounders	Yes	Yes Used routine practice data from a large, diverse cohort of patients across multiple U.S. hospitals, reflecting typical clinical practice for TAVR	The results of the study using instrumental variable analysis are consistent with the large-scale randomized clinical trials that were completed after this study ³⁷ was published

Abbreviations: AF, atrial fibrillation; CKD, chronic kidney disease; CPT, Current Procedural Terminology; EHR, electronic health record; ICD, International Classification of Diseases; PICO, patient, intervention, comparator, outcome; SNOWMED, Systematized Nomenclature of Medicine; VTE, venous thromboembolism.

Table 3 Illustrative examples of double-blind and pragmatic thrombosis-related randomized controlled trials and their critical evaluation

Trial	PICO	Well-defined patient selection criteria	Blinding	Allocation sequence concealment	Blinded adjudication of the outcomes	Well-defined pre-specified outcomes	Intervention and ancillary procedures feasible in routine clinical practice	Study designed to assess effectiveness of treatment in clinical practice	Adequate representation of diverse groups by age, sex, and race-ethnicity	Robust methods to assess and report treatment adherence	Intention-to-treat analytic approach
Thrombosis-related double-blind randomized controlled trials											
ARISTOTLE ²⁸	In patients with AF not associated with severe MS, is apixaban noninferior to warfarin in preventing stroke or systemic embolism?	Yes	Yes	Yes	Yes	Yes	No Attaining high time in therapeutic target for INR is challenging in routine practice	No Stringent eligibility criteria limits generalizability to routine practice settings (including severe kidney disease); strict adherence plan in place in the trial	No Only 30% of study population was females, despite estimates suggesting roughly 50% of patients with AF being female in general ²⁹ ; racial/ethnicity data not reported	Yes Employed blinded INR monitoring, dose adjustment algorithms, tracking of time in therapeutic range	Yes
PLATO ⁴⁰	In patients hospitalized with acute coronary syndrome, does treatment with ticagrelor, compared to clopidogrel, reduce the risk of death from vascular causes, myocardial infarction, or stroke without increasing the risk of major bleeding?	Yes	Yes	Yes	Yes ³	Yes	Yes Both medications are approved, widely available, and commonly used for the management of acute coronary syndrome; follow-ups mirror typical cardiology follow-up schedules	No Highly selected and narrow inclusion criteria and strict protocol-driven processes	No Females represented only 28% of the study population, which was predominantly White (92%)	No While the study employed standard mechanisms (e.g., blinding, frequent visits) to support adherence, it did not report adherence using robust, detailed quantitative methods	Yes
Thrombosis-related pragmatic trials											
ADAPTABLE ⁴¹	In patients with established atherosclerotic cardiovascular disease, does daily aspirin 81 mg, compared with 325 mg, result in different rates of death, myocardial infarction, stroke, or major bleeding?	Yes	No The study was open-label and there was a substantial proportion of patients who crossed over with respect to aspirin dose	Yes	Yes	No Outcome adjudication did not follow traditional metrics of other trials in which central adjudication of all events occur	Yes Aspirin widely available, limited direct patient interaction during the study period, medications distributed via mail, participants were largely responsible for self-management	Yes Patient population recruited from standard clinical practice, with less stringent selection criteria, resulting in enhanced generalizability	No Study comprised only 30% females, 9% Black and 3% Hispanics	No Relied on self-reporting and self-purchase, with no pill counting or pharmacy refill data	Yes
VALIDATE-SWE-DEHEART ^{b,42}	In patients with MI undergoing PCI, does bivalirudin, compared with unfractionated heparin, reduce the	Yes	No	Yes	Yes	Yes Some of the source data were less	Yes Study involves standard treatments/	Yes The trial used a registry-based design with patients	No Less than 30% of the study population comprised	No Treatment was short term and in the	Yes

(Continued)

Table 3 (Continued)

Trial	PICO	Well-defined patient selection criteria	Blinding	Allocation sequence concealment	Blinded adjudication of the outcomes	Well-defined pre-specified outcomes	Intervention and ancillary procedures feasible in routine clinical practice	Study designed to assess effectiveness of treatment in clinical practice	Adequate representation of diverse groups by age, sex, and race-ethnicity	Robust methods to assess and report treatment adherence	Intention-to-treat analytic approach
	composite risk of death, myocardial infarction, or major bleeding?			automated allocation system using the SWEDE-HEART registry		detailed than other traditional trials and partially relied on ICD codes	interventions that are routinely performed in the clinical practice	recruited from standard clinical practice, with less stringent selection criteria	females; racial/ethnicity data not reported	catheterization laboratory without major adherence concerns	

Abbreviations: AF, atrial fibrillation; ICD, International Classification of Disease; INR, international normalized ratio; MS, mitral stenosis; N, no; NSTEMI, non-ST-segment elevation myocardial infarction; PCI, percutaneous coronary intervention; STEMI, ST-segment elevation myocardial infarction; Y, yes.

^aThere have been discussions about the magnitude of reported treatment effect for some outcomes, including mortality, whose discussion is beyond the scope of the current manuscript.

^bVALIDATE-SWEDEHEART had certain clear eligibility criteria; however, features such as being embedded in a registry and adherence to some standards of routine practice aligned with a pragmatic design.

Excellence in Regulatory Science and Innovation (CERSI) program (U01FD005938), from the Agency for Healthcare Research and Quality (R01HS022882), and from Arnold Ventures; formerly received research support from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST) and from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644); and in addition, J.S.R. was an expert witness at the request of Relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc. that was settled September 2022. J.D.W. reported receiving grants from the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health (under award 1K01AA028258), Johnson & Johnson (through the Yale Open Data Access Project), Arnold Ventures, and the FDA, as well as former consulting fees from Hagen Berman Sobol Shapiro LLP and Dugan Law Firm APLC outside the submitted work. M.M.J. has received unrelated funding from National Institute on Drug Abuse (NIDA), the United States Food and Drug Administration (FDA), and the Agency for Healthcare Research and Quality (AHRQ).

References

- Ford I, Norrie J. Pragmatic trials. *N Engl J Med* 2016;375(05):454–463
- Kang DS, Yang PS, Kim D, et al. Racial differences in ischemic and hemorrhagic stroke: an ecological epidemiological study. *Thromb Haemostasis* 2024;124(09):883–892
- Guo Y, Lane DA, Wang L, et al; mAF-App II Trial Investigators. Mobile health technology to improve care for patients with atrial fibrillation. *J Am Coll Cardiol* 2020;75(13):1523–1534
- Chu M, Zhang S, Gong J, et al; MIRACLE-AF Investigators. Telemedicine-based integrated management of atrial fibrillation in village clinics: a cluster randomized trial. *Nat Med* 2025;31(04):1276–1285
- Bikdeli B, Khairani CD, Barns BM, et al. Women's representation in venous thromboembolism randomized trials and registries: the illustrative example of direct oral anticoagulants for acute treatment. *Contemp Clin Trials* 2022;115:106714
- You SC, Krumholz HM. The evolution of evidence-based medicine: when the magic of the randomized clinical trial meets real-world data. *Circulation* 2022;145(02):107–109
- Welsh J, Lu Y, Dhruva SS, et al. Age of data at the time of publication of contemporary clinical trials. *JAMA Netw Open* 2018;1(04):e181065
- Mauri L, Kereiakes DJ, Yeh RW, et al; DAPT Study Investigators. Twelve or 30 months of dual antiplatelet therapy after drug-eluting stents. *N Engl J Med* 2014;371(23):2155–2166
- Butala NM, Faridi KF, Tamez H, et al. Estimation of DAPT study treatment effects in contemporary clinical practice: findings from the EXTEND-DAPT study. *Circulation* 2022;145(02):97–106
- Erdogan E, Bajaj R, Lansky A, Mathur A, Baumbach A, Bourantas CV. Intravascular imaging for guiding in-stent restenosis and stent thrombosis therapy. *J Am Heart Assoc* 2022;11(22):e026492
- Bikdeli B, Ujueta F, Rashedi S, et al. Antiplatelet and anticoagulant therapy in the 2025 ACC/AHA guideline for acute coronary syndromes: key recommendations. *JACC* 2025;85(22). In-Press

- 12 Young K. FDA Loosens Restrictions on Metformin Use in Patients with Mild-to-Moderate Kidney Dysfunction. 2016. Accessed April 29, 2025 at: <https://www.jwatch.org/fw111424/2016/04/11/fda-loosens-restrictions-metformin-use-patients-with-mild>
- 13 Kim JY, Kim SH, Myong JP, et al. Outcomes of direct oral anticoagulants in patients with mitral stenosis. *J Am Coll Cardiol* 2019;73(10):1123–1131
- 14 Connolly SJ, Karthikeyan G, Ntsekhe M, et al; INVICTUS Investigators. Rivaroxaban in rheumatic heart disease-associated atrial fibrillation. *N Engl J Med* 2022;387(11):978–988
- 15 Bejjani A, Bikdeli B. Direct oral anticoagulants: quick primer on when to use and when to avoid. *Thromb Haemost* 2025;125(07):611–617
- 16 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183(08):758–764
- 17 Hansford HJ, Cashin AG, Jones MD, et al. Reporting of observational studies explicitly aiming to emulate randomized trials: a systematic review. *JAMA Netw Open* 2023;6(09):e2336023
- 18 Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. *JAMA* 2024;331(21):1845–1853
- 19 Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019;394(10211):1816–1826
- 20 Wallach JD, Deng Y, Polley EC, et al. Assessing the use of observational methods and real-world data to emulate ongoing randomized controlled trials. *Clin Trials* 2023;20(06):689–698
- 21 Wang SV, Schneeweiss S, Franklin JM, et al; RCT-DUPLICATE Initiative. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* 2023;329(16):1376–1385
- 22 You SC, Rho Y, Bikdeli B, et al. Association of ticagrelor vs clopidogrel with net adverse clinical events in patients with acute coronary syndrome undergoing percutaneous coronary intervention. *JAMA* 2020;324(16):1640–1650
- 23 Hanberg JS, Fu X, Wang X, et al. Effectiveness of a fourth dose of COVID-19 mRNA vaccine in patients with systemic autoimmune rheumatic diseases using disease-modifying antirheumatic drugs: an emulated target trial. *Lancet Rheumatol* 2024;6(01):e21–e30
- 24 Xie Y, Bowe B, Xian H, Loux T, McGill JB, Al-Aly Z. Comparative effectiveness of SGLT2 inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and sulfonyleureas on risk of major adverse cardiovascular events: emulation of a randomised target trial using electronic health records. *Lancet Diabetes Endocrinol* 2023;11(09):644–656
- 25 Kutcher SA, Brophy JM, Banack HR, Kaufman JS, Samuel M. Emulating a randomised controlled trial with observational data: an introduction to the target trial framework. *Can J Cardiol* 2021;37(09):1365–1377
- 26 Bikdeli B, Lo YC, Khairani CD, et al. Developing validated tools to identify pulmonary embolism in electronic databases: rationale and design of the PE-EHR+ study. *Thromb Haemost* 2023;123(06):649–662
- 27 Bikdeli B, Khairani CD, Bejjani A, et al; PE-EHR+ Investigators. Validating International Classification of Diseases Code 10th Revision algorithms for accurate identification of pulmonary embolism. *J Thromb Haemost* 2025;23(02):556–564
- 28 Wallach JD, Zhang AD, Skydel JJ, et al. Feasibility of using real-world data to emulate postapproval confirmatory clinical trials of therapeutic agents granted US Food and Drug Administration accelerated approval. *JAMA Netw Open* 2021;4(11):e2133667
- 29 Cunningham JW, Singh P, Reeder C, et al. Natural language processing for adjudication of heart failure in a multicenter clinical trial: a secondary analysis of a randomized clinical trial. *JAMA Cardiol* 2024;9(02):174–181
- 30 Hubbard RA, Gatsonis CA, Hogan JW, Hunter DJ, Normand ST, Troxel AB. “Target trial emulation” for observational studies—potential and pitfalls. *N Engl J Med* 2024;391(21):1975–1977
- 31 Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA* 2022;328(24):2446–2447
- 32 Monteleone P, Ahern R, Banerjee S, et al. Modern treatment of pulmonary embolism (USCDT vs MT): results from a real-world, big data analysis (REAL-PE). *J Soc Cardiovasc Angiogr Interv* 2023;3(01):101192
- 33 Cohen AT, Sah J, Dhamane AD, et al. Effectiveness and safety of apixaban versus warfarin in venous thromboembolism patients with chronic kidney disease. *Thromb Haemost* 2022;122(06):926–938
- 34 Chowdhury KR, Michaud J, Yu OHY, Yin H, Azoulay L, Renoux C. Effectiveness and safety of apixaban versus rivaroxaban in patients with atrial fibrillation and type 2 diabetes mellitus. *Thromb Haemost* 2022;122(10):1794–1803
- 35 Weycker D, Li X, Wygant GD, et al. Effectiveness and safety of apixaban versus warfarin as outpatient treatment of venous thromboembolism in U.S. clinical practice. *Thromb Haemost* 2018;118(11):1951–1961
- 36 Butala NM, Makkar R, Secemsky EA, et al. Cerebral embolic protection and outcomes of transcatheter aortic valve replacement: results from the transcatheter valve therapy registry. *Circulation* 2021;143(23):2229–2240
- 37 Kapadia SR, Kodali S, Makkar R, et al; SENTINEL Trial Investigators. Protection against cerebral embolism during transcatheter aortic valve replacement. *J Am Coll Cardiol* 2017;69(04):367–377
- 38 Granger CB, Alexander JH, McMurray JJ, et al; ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365(11):981–992
- 39 Volgman AS, Benjamin EJ, Curtis AB, et al; American College of Cardiology Committee on Cardiovascular Disease in Women. Women and atrial fibrillation. *J Cardiovasc Electrophysiol* 2021;32(10):2793–2807
- 40 Wallentin L, Becker RC, Budaj A, et al; PLATO Investigators. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 2009;361(11):1045–1057
- 41 Jones WS, Mulder H, Wruck LM, et al; ADAPTABLE Team. Comparative effectiveness of aspirin dosing in cardiovascular disease. *N Engl J Med* 2021;384(21):1981–1990
- 42 Erlinge D, Omerovic E, Fröbert O, et al. Bivalirudin versus heparin monotherapy in myocardial infarction. *N Engl J Med* 2017;377(12):1132–1142