



Advancing the role of real-world evidence in comparative effectiveness research

Monica Daigl^{1*}, Seye Abogunrin¹, Felipe Castro², Sarah F McGough³, Rachele Hendricks Sturup⁴, Cornelis Boersma^{5,6,7} & Keith R Abrams^{8,9}

¹Global Access F. Hoffmann-La Roche Ltd., 4070 Basel, Switzerland

²Data Science, F. Hoffmann-La Roche Ltd., 4070 Basel, Switzerland

³Computational Sciences, Genentech Inc., South San Francisco, CA 94080, USA

⁴Duke-Robert J. Margolis Institute for Health Policy, Washington, DC 20004, USA

⁵Health-Ecore, 3704 HE Zeist, The Netherlands

⁶Department of Health Sciences, University Medical Center Groningen, Groningen, 9700 AB, The Netherlands

⁷Department of Management Sciences, Open University, Heerlen, 6419 AT, The Netherlands

⁸Department of Statistics & Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK

⁹Centre for Health Economics, University of York, York, YO10 5DD, UK

*Author for correspondence: monica.daigl@roche.com

Journal of **Comparative Effectiveness Research**

Aim: Comparative effectiveness research (CER) is essential for making informed decisions about drug access. It provides insights into the effectiveness and safety of new drugs compared with existing treatments, thereby guiding better healthcare decisions and ensuring that new therapies meet the real-world needs of patients and healthcare systems. **Objective:** To provide a tool that assists analysts and decision-makers in identifying the most suitable analytical approach for answering a CER question, given specific data availability contexts. **Methods:** A systematic literature review of the scientific literature was performed and existing regulatory and health technology assessment (HTA) guidance were evaluated to identify and compare recommendations and best practices. Based on this review a methods flowchart that synthesizes current practices and requirements was proposed. **Results:** The review did not find any papers that clearly identified the most appropriate analytical approach for answering CER questions under various conditions. Therefore, a methods flowchart was designed to inform analyst and decision makers choices starting from a well-defined scientific question. **Conclusion:** The proposed methods flowchart offers clear guidance on CER methodologies across a range of settings and research needs. It begins with a well-defined research question and considers multiple feasibility aspects related to CER. This tool aims to standardize methods, ensure rigorous and consistent research quality and promote a culture of evidence-based decision-making in healthcare.

Plain language summary: What is this article about?: This article discusses how researchers and healthcare decision-makers can determine the best way to compare the effectiveness of different drugs. This type of research, known as comparative effectiveness research (CER), helps to make better healthcare decisions by providing information on how new drugs perform compared with existing treatments. The article aims to offer a tool that guides analysts in choosing the right method for their CER based on the data they have.

What were the results?: The study found that there are no existing papers that clearly explain which method to use for different CER questions under various conditions. To address this gap, the authors created a tool - a methods' flowchart - that will facilitate a transparent way of choosing which method should be used for a specific CER question. This tool helps researchers start with a specific question and then choose the best method to answer it, rather than forcing a one-size-fits-all approach.

What do the results of the study mean?: The results mean that researchers and decision-makers now have a clear guide to help them choose the most appropriate methods for their CER questions. This new tool aims to make CER more standardized and consistent, which can lead to higher quality research and better, evidence-based decisions in healthcare. Ultimately, this can improve patient care by ensuring that new therapies meet real-world needs.

Shareable abstract: This study introduces a methods flowchart to guide comparative effectiveness research (CER). It helps identify the best analytical approach, ensuring standardized, high-quality, evidence-based healthcare decisions. #CER #decision-making #comparative-effectiveness, #evidence-based

First draft submitted: 2 July 2024; Accepted for publication: 23 September 2024; Published online: 11 October 2024

Keywords: comparative effectiveness research • methods • real-world evidence

What this paper aims to achieve

Comparative Effectiveness Research (CER) is defined as “*the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers and policy makers to make informed decisions that will improve healthcare at both the individual and population level*” [1]. As new therapeutic interventions are developed, healthcare systems worldwide grapple with evaluation including decisions on marketing authorization and funding based on the available clinical evidence. CER currently plays a critical role for crafting clinical guidelines and reimbursement policies that facilitate informed decision-making to enhance healthcare outcomes [2,3].

Through our extensive experience with reimbursement dossier submissions and evaluations, we’ve encountered a dichotomy in the existing guidelines. These can be broadly categorized into two schools of thought: the ‘traditional’ school, which emphasizes the importance of experimental evidence in particular the importance of randomized designs and; the ‘evolving’ school, which considers alternative evidence when randomized studies are not feasible, unethical, impractical or underpowered.

Our stance is that use of randomized designs and alternative evidence, while seemingly at odds, can be synergistic rather than mutually exclusive. Scarcity of data often hampers efforts to comprehensively assess and quantify the added value of new interventions compared with evolving standards of care. Accordingly, the two approaches may each serve a purpose depending on the research context and the availability of the data at the time of decision-making and can complement each other.

Since the development of the ‘evidence pyramid’ which details a hierarchy of evidence informing medical decision-making prioritizing randomized studies over observational ones, there has been scant guidance on integrating diverse methodological approaches to evaluate evidence effectively [4]. This paper proposes a framework to navigate the challenges posed by limited resources – such as time, data and funding – while addressing pressing research questions.

Multiplicity versus parsimony

William of Ockham, also known as Occam, was a 14th century English philosopher and theologian whose razor theory has greatly influenced the scientific community with the concept of parsimony in research. Occam’s razor, suggests that when faced with multiple competing hypotheses, the simplest explanation is often the most likely to be true. This principle encourages scientists to prefer simpler and more replicable explanations over complex ones.

Occam’s razor has been instrumental in shaping scientific inquiry. It serves as a guiding principle in hypothesis formulation and model building, pushing researchers to prioritize simplicity in their work. By doing so, Occam’s razor has led to more efficient and effective scientific progress, as it encourages scientists to avoid unnecessary complexity that can lead to confusion and inefficiency.

CER aims at answering the question “Does it work?” and it compares the benefits and harms of different medical interventions both via evidence generation and evidence synthesis [5]. The goal of CER is to provide evidence-based information to assist patients, doctors and healthcare systems identifying the best therapeutic choices. When (multiple) research questions regarding a new intervention are formulated, the number of experiments that can be conducted and leveraged is limited. Randomized controlled trials (RCTs) comparing multiple alternative options require time and resources and may face challenges in their implementation. Despite the RCT being the gold standard, it cannot answer all questions that arise when a new intervention becomes available, moreover generalizability to the specific populations might not be straightforward.

Inspired by the principle of parsimony and driven by a pragmatic approach, published guidance on this topic was revisited and a methodological framework to guide the choice of the methodological approach that is best suited to answer each of the research questions being formulated around a new intervention is proposed.

Interventional data & real-world data can provide supportive evidence for assessing the effectiveness of new therapies

As previously noted, RCTs are considered the gold standard to assess a new drug intervention. Yet, RCTs are limited in their potential to address questions that arise when the new intervention becomes available within the health market for a specific indication and is compared in terms of safety and efficacy to accepted standard of care interventions [6]. One way to address this is to collect, compare and evaluate interventional data from clinical trials (e.g., network meta-analysis) and RWD, both of which can provide supportive real-world evidence (RWE) to assess the effectiveness of new therapies against a standard of care treatment. This is especially true in cases of rare disease where a prospective and blinded RCT to assess medical product safety and efficacy for a new treatment may not be feasible within a specific and/or reasonable time-frame and/or may be unethical. Nonetheless, in both rare and common disease use cases, patient outcomes and/or end points observed in single-arm premarket trials may generate data that are comparable to observational data reflecting a standard of care approach in real-world settings.

For example, a study published in 2021 assessing treatment outcomes among patients with *ROS1*+ non-small-cell lung cancer (NSCLC), a rare and biomarker-defined lung cancer, leveraged electronic health record data reflecting outcomes from patients treated with crizotinib (n = 65, USA only), a then-current standard of care treatment [7]. The outcomes were compared with clinical trial data (phase I and one phase II single-arm study data) reflecting outcomes among *ROS1*+ patients treated with entrectinib (n = 94, Europe, Australia, Asia and the USA), a drug that received US Food and Drug Administration (FDA) approval in August 2019 for the treatment of NSCLC. Time-to-treatment discontinuation (TTD) was evaluated as the primary end point in this study, based on two key factors noted by the authors: the availability of common variables between the clinical trial and RWD cohorts and TTD was previously identified as a pragmatic end point for use in retrospective RWE studies [7,8]. Such factors, among potential others across various disease use cases, are useful and insightful for the consideration of CER study designs that leverage RWD to generate supportive evidence of effectiveness that is comparable to clinical trial evidence for newly marketed treatments. Thus, in the next section, the timeliness of these considerations to support health technology assessment (HTA) and regulatory authority decision-making is discussed.

Non-randomized evidence in health authority decision-making

Pharmacoepidemiological studies utilizing non-randomized RWE have been widely used to gather data and evidence supporting the evaluation of the post-marketing safety of approved medicinal products after they have been brought to market. The global use and scope of these studies for regulatory decision-making have grown significantly, leading to the development of various guidelines and best practices documents by regulatory bodies and professional societies [9].

In 2016, the US introduced the 21st Century Cures Act (Cures Act) [10], aimed at expediting the development of medical products and ensuring that innovative treatments reach patients more quickly and efficiently. As a result, the FDA established a RWE Program to assess the potential use of RWD in regulatory decision-making for drugs [11]. Subsequently, the FDA has released a series of regulatory guidance documents pertaining to the use of RWD/E for various purposes such as evaluating product safety and effectiveness, assessing medical devices, utilizing electronic health records, establishing data standards and conducting externally controlled trials [9,11,12].

In parallel, the European Medicines Agency (EMA) has also been actively engaged in this area. They initiated the Adaptive Pathways Pilot in 2014 [13] and have since published guidance documents like the Operational, Technical and Methodological (OPTIMAL) framework for leveraging RWE in regulatory decision-making [14]. More recently, the EMA, in collaboration with the Heads of Medicines Agencies (HMA), established the Big Data Task Force to enhance standardization of RWD across Europe [15]. Other countries such as Canada, China and Japan have also published guidelines outlining general principles for planning and designing RWD studies to support regulatory approvals [16,17]. Additionally, payer and HTA organizations in Europe have been influenced by the growing interest in RWE. The RWE4Decisions initiative was implemented to define the roles of various stakeholders involved in the process and propose practical actions for payers, HTA bodies and other stakeholders to incorporate RWE, particularly in relation to innovative technologies [18].

Non-randomized evidence for the assessment of relative treatment effects is increasingly relevant for HTA, however use of RWE is inconsistent [19]. And while CER utilizing RWE has primarily been evaluated in HTA settings, both regulatory agencies and payers/HTAs now demonstrate greater receptiveness toward the use of RWE. However, they agree that RWD should be considered complementary to randomized evidence, with its acceptability dependent on the quality, validity and methodological rigor of the data collection process and analytical approaches to control for bias and confounding. Therefore, the application of CER to support regulatory or payer/HTA discussions, especially using RWD, should adhere to a rigorous process that meets the expectations of these agencies.

Lack of common standards leads to inconsistency in methods used & lower reliability of/trust in CER

Guidelines outline the evidence that health technology developers need to produce to inform regulatory and HTA decision-making, respectively. While multiple guidelines exist regarding the use of RCT or RWD, a clear standard encompassing both the use of interventional and observational data, as well as methodologies for analyzing these data, is notably absent. Furthermore, guidelines across various jurisdictions often present conflicting advice. For example, the practical guideline on comparisons recently issued by the EU member state coordination group on HTA in March 2024 indicates explicitly that it does not make any recommendations about whether a submitted direct and indirect treatment comparison should be accepted by the Member States [20]. Such ambiguity can lead to methodological discrepancies when researchers tackle similar research questions, complicating the comparison and synthesis of the results.

These methodological variations and inconsistent use of methods can introduce bias and uncertainty into CER studies, undermining their reliability and hereby leading to arbitrary decision-making. This can result in confusion, inequities, or inefficiencies in patient access to care and uncertainty for healthcare professionals and patients when making treatment decisions. Ultimately, this lack of standardization may erode public and stakeholder confidence in the overall validity and utility of CER findings. Addressing this issue is critical to enhance the consistency and credibility of research outcomes, which is foundational for evidence-based practice and policy-making.

Overview of paper structure

The paper is structured as follows: the section "Review of CER methods", along with Appendix A–E, summarize published guidance around Comparative Effectiveness Research; the section "Proposed CER methods flowchart" introduces the CER methods flowchart; the following section "Methods flowchart in practice: application to an example study", provides examples on the use of the methods flowchart in practice; and the section "Future directions & challenges" proposes future direction and discusses challenges for researchers and policy makers.

Review of CER methods

Is there a consensus about which method is better suited?

Using PRISMA reporting guidelines [21], a systematic review of the literature identified from electronic indexed databases and select HTA sources was undertaken to identify publications and methodological guidelines that offer suggestions and best practices for conducting CER. The pre-defined selection criteria utilized a modified SPIDER framework [22], which is described in Appendix A. The publications included in the review were those that featured a comparison of at least two approaches for conducting CER, along with specific recommendations or guidance on how to proceed with CER. Publications that applied CER methods to a specific disease and guidelines limited to one single method, or topics such as policy, data quality, education and stakeholder engagement were not considered. Additionally, to be included in the review, publications needed to have either full-text or abstracts available, and be written in English. However, inclusion of guidelines was not restricted by language.

The database search was conducted on the 30 May 2023. Keywords synonymous with “comparative effectiveness”, “practice guidelines”, “methodology” and “recommendations” were combined to identify records that could potentially provide information on the recommendations and best practices for decision-making when conducting a CER. These searches were supplemented by a search of key HTA and regulatory websites for methodological guidelines published up to April 2024.

A total of 952 unique records were identified by the searches of EMBASE and MEDLINE (Appendix B). These records were screened based on their title and abstract using the Abstrackr tool (<http://abstrackr.cebm.brown.edu>) by each two independent reviewers who carefully assessed the citations (records review was performed by FC, MD, SA and SM). Of these title and abstract records, full-texts of 39 of them were selected and screened further using a

spreadsheet (Google Sheets), again by two independent reviewers, to determine whether they satisfied the criteria for inclusion, and if data should be extracted from them. In cases where there were disagreements at both levels of screening, discussions took place to arrive at a decision to either include or exclude the record and when excluded at the full-text level, a reason was documented for exclusion. Following this screening process, five publications were deemed relevant [23–27]. In addition, the search of HTA and regulatory websites identified 22 [20,28–48] and 2 [49,50] relevant methodological guideline documents. As a result, 29 publications and methodological guidelines were included in the review and subsequently, underwent data extraction (Appendix C).

Data extraction elements

It is sometimes unclear for researchers involved in CERs, especially those unfamiliar with the methods, to know how to decide what types of methodological approach to use for the analyses or which data is best suited to be analyzed in a CER. As such, guidance related to certain key decision points considered important during the preparation for a CER was extracted in a predesigned data extraction form (using Google Sheets). The reporting of these data are presented in Appendix D. See Appendix E for questions that CER researchers should aim to answer when preparing for such research and how these relate to reporting of CER guidance in evidence identified by the review.

Summary of findings of the review

Most of the guidance documents identified from peer-reviewed literature [23,24] and HTA [20,28,29,32–48] and regulatory [49,50] websites provide a very comprehensive description on specific methods and their application, however, no single scientific publication or methodological guideline document was found to provide a comprehensive overview to inform the choice of the methodological approaches that should be used in different situations when preparing for CERs.

The peer-reviewed documents [23–27] identified from the literature provided sparse guidance with the most information being reported on the appropriateness of using a RWE study, representativeness of a RWE study, and identification of confounders in RWE studies used for CER.

HTA documents [28,29,34–48] reported more specific methods guidance when considering the use of data from randomized studies: availability of clinical data from comparator trials, connectivity of a network based on data from RCTs and considerations around effect modifiers was deemed important. Guidance on the use of RWE for CERs was focused mostly on representativeness of the sample included in such studies, and identification and measurement of confounders.

The Regulatory documents reviewed [49,50] provided mainly guidance for how to use RWE for CER to support benefit-risk evaluations, with a focus on the following topics: appropriateness of designing a RWE study, feasibility of a prospective or retrospective study, availability of data from observational studies or single arm trials, representativeness of a RWE study, identification of confounders in RWE studies used for CER, and availability of individual patient data (IPD) from RWE studies.

Proposed CER methods flowchart

In this section of the paper a methods flowchart is introduced to help decide when randomized studies or observational studies and single arm trial (SAT) are appropriate to perform comparative effectiveness/safety research and to help select the right methodological approach.

Specifying a research question

Evidence Based medicine starts with formulating a clear research question. The Population, Intervention, Comparator, Outcome (PICO) framework helps with this task as it guides researchers into defining the patient group for which we want to make inferences (P), the intervention being considered (I), the relevant comparison (C) and the clinical outcome of interest (O) [51].

CER methods flowchart

After a research question has been formulated, the proposed decision tree/flowchart (Figure 1) can be adopted to select the best data and methodological approach. As a first step the feasibility of conducting a prospective randomized head-to-head trial against the comparator of interest in the target population is evaluated. If such a trial is feasible, then consideration should be given to a prospectively planned RCT or a randomized pragmatic trial



Figure 1. The comparative effectiveness research methods flowchart allows the selection of appropriate comparative effectiveness research methodologies given the target population, research question and data availability and suitability.

IPPD: Individual patient data; ITC: Indirect treatment comparison; MAIC: Matching-adjusted indirect comparison; NC: Negative controls; NMA: Network meta-analysis; PSA: Propensity score analysis; RCT: Randomized controlled trial; SAT: Single arm trial; SLR: Systematic literature review; STC: Simulated treatment comparison.

(RPT). The RCT, introduced in 1946 with the MRC Streptomycin in Tuberculosis trial in the UK, has quickly become a widely accepted model for design and implementation in clinical research [52]. Its strength lies in its ability to minimize potential systematic errors or biases. However, concerns about applicability of trial results to everyday practice were raised in subsequent years. As a result, pragmatic trials have gained interest in the scientific community more recently [53–55].

Despite the intrinsic values of randomized experimental designs, not every research question can be answered using such designs. When these are not feasible, ethical or timely, it is advisable to document the challenges and evaluate alternative approaches for generating evidence. The next step in this process would be to conduct a systematic literature review (SLR) to identify published clinical studies involving the comparators of interest. The purpose of a SLR is to provide a concise and unbiased summary of the available evidence [56]. Many HTA bodies require a SLR as a starting point for the evidence generation process. Through the SLR, published studies involving the comparators of interest can be identified.

If clinical studies have been published and the data are representative of the target population of interest for the comparison, availability of results from RCT or RPT can be verified, as these studies provide the best unbiased estimates of efficacy and safety. When direct evidence from RCT or RPT is available, it is recommended to perform meta-analyses to summarize the available evidence [21,57].

In many cases, head-to-head comparisons between the treatment options of interest may not have been conducted. In such situations, the next step is to determine whether a network of connected randomized studies exists. This occurs when, for example, a health technology developer has conducted a study comparing their drug against a placebo, and a competitor study has also compared their drug against placebo. In this scenario, the placebo arm can be used as a bridge to indirectly compare the two treatment options [58].

Indirect comparisons, such as network meta analyses (NMA) and indirect treatment comparisons (ITC), are valuable tools for synthesizing evidence when head-to-head comparisons are not available or feasible [59]. However, it is important to acknowledge the potential limitations and biases associated with these methods, particularly in the presence of differences in study populations. When important effect modifiers are present the standard ITC/NMA methods may not adequately handle these scenarios. These methods assume consistency of treatment effects across different patient populations, which may not hold true. In the presence of effect modifiers, caution should be exercised when interpreting the results of indirect comparisons. Consider for example two oncology drugs being compared, one of which can pass the blood–brain barrier whereas the other doesn't: if the first drug were evaluated in a study population with higher prevalence of patients with brain metastases, whereas the second in a less prevalent population there would not be equipoise in the indirect comparison. Additional approaches, such as subgroup or sensitivity analyses may be necessary to explore the impact of effect modifiers and assess the robustness of the findings. Transparency in reporting and clear communication of the strengths and limitations of the analysis will contribute to the overall scientific rigor of the study. By acknowledging and addressing these challenges, researchers can ensure a more accurate interpretation of the findings from ITC/NMA and facilitate informed decision-making in clinical practice.

When effect modifiers are well-balanced across study populations, standard ITC and NMA methods can be safely employed. However, in cases where differences exist between the study populations, alternative methods may be more appropriate. These include IPD NMA, matching adjusted indirect comparisons (MAIC), simulated treatment comparisons (STC) and multi level network meta-regressions [60–62]. In certain rare cases, when the populations have identical distributions of prognostic factors and effect modifiers, unadjusted comparisons might be preferable as they are more straightforward to comprehend.

When published clinical data is either unavailable or does not meet the discussed criteria, use of observational IPD is a good alternative. Observational data, oftentimes sourced from large databases (registry, electronic health records, claim databases) have the advantage that observations reflect the reality of the real world, where patients are treated according to their local standard of care.

The starting point before engaging in the comparison with observational IPD is to perform a database or data source landscape review to understand which data sources are available. A data dictionary for assessing availability of data in identified data sources can be prepared. The preferred data source for the population of interest should consider: data availability (can the data be accessed?), representativeness (is the amount of data sufficient and is it representative of the target population?), confidentiality (are conditions associated with access to the data?) and the most appropriate analytical approach.

Priority shall be given to prospectively collected data, if such data is available and the researcher can have access to these data with full attention to fitness for purpose, provenance and governance [63]. Additional considerations include the representativeness of the data to ensure external validity and availability of measurements to control for potential sources of bias. As observational data is prone to selection bias and confounding, the availability of data on confounding variables should be considered.

In cases where IPD is available for both the health technology developer's clinical trial and observational data containing the comparator population, external control comparisons (e.g., propensity score analyses [PSA], instrumental variable/negative controls [NC]) are among the appropriate approaches to deal with measured and unmeasured confounding [64,65], whereas population averaged approaches detailed above are the method of choice in case that IPD is not available.

In cases where the study population is not representative, or confounders are not appropriately measured, the research question cannot be answered and the data may be alternatively used to formulate new research questions or hypotheses.

Validity assessments

The methods flowchart in [Figure 1](#) presents a framework for the selection of appropriate CER methodologies given the target population, research question and data availability and suitability. As such, it provides a starting point for analysis. However, to ensure that the subsequent results are reliable and useful for decision-making, it is crucial to perform validity assessments. Validity assessments can be broadly categorized into internal validity and external validity. Each type of validity requires different methods of investigation, presented briefly below.

Internal validity refers to the degree to which the results of the analysis accurately reflect the reality within the study population, unaffected by external or confounding factors. This is especially important in cases of CER study designs without randomization, and is needed to ensure confidence in the findings. To conduct internal validity assessments:

Verify statistical assumptions

Internal validity assessments should start with a verification of the model specification and statistical assumptions of the chosen model. For meta-analyses, testing of the statistical assumptions can include an assessment of heterogeneity through the I^2 statistic and Q-test to determine whether the analysis requires a fixed- or random-effects model or subgroup analysis [66]. For meta-regression, this can include verifying assumptions of linearity, independence and homoscedasticity of residuals and involve diagnostics such as inspecting residual plots or calculating variance inflation factor (VIF) for multicollinearity [67]. For propensity score methodologies, this can include verifying the specification and assumptions of the underlying propensity score model and assessing its sensitivity to the modeling of different covariates or different functional forms [68].

Check balance diagnostics

For propensity score-based approaches, it is important to check for balance in covariates between comparator groups after adjustment (such as reweighting or matching). Standardized mean differences can be used for this purpose, with cut-off thresholds of 0.1 or 0.25 commonly cited in the literature [68].

Conduct qualitative & quantitative bias analysis

By examining causal diagrams [69] researchers can identify potential confounding variables in their models that need to be measured and adjusted for in order to estimate the true causal effect accurately. Quantitative bias analysis (QBA) involves quantifying the extent to which multiple biases (like selection bias, information bias and confounding) could have affected the results. In the context of comparative effectiveness analyses, the concern over unmeasured confounding is particularly significant. A useful metric is the E-value, or 'evidence for causality', which quantifies the extent to which unmeasured confounding could explain away the observed results [70]. It requires no prior knowledge of or assumptions about potential unmeasured confounders, and is simple to compute. Other QBA methods include sensitivity analyses varying key model parameters and probabilistic bias analysis to assess uncertainty due to bias [71].

Evaluate & report limitations

This includes limitations in data quality, sample size, or data availability to adjust for sources of bias. Acknowledging limitations helps decision-makers better interpret the results.

External validity concerns the extent to which the findings of a study can be generalized to other settings, populations, or times. External validity is vital for the practical application of CER. Here, transportability methods can be employed, which involve the methods detailed in [Figure 1](#) applied to a new target population [72–74]. Standard assumptions for internal validity are again required as outlined above.

Methods flowchart in practice: application to an example study

To comprehend the practical implementation of the proposed methods flowchart, here we retrospectively apply the flowchart to published CER analyses between entrectinib and crizotinib in ROS1-mutated NSCLC.

Firstly, it is essential to begin with a well-defined research question that encompasses patient populations, comparators and outcomes. In this particular case, the study research question aimed to generate comparative efficacy evidence for entrectinib versus crizotinib, which was the standard of care when entrectinib was approved, for the ROS1+ NSCLC patient population. The initial step, then, involved assessing the feasibility of conducting a randomized study. However, due to the rarity of ROS1 fusions, an RCT/RPT was not feasible in this scenario.

The next step entails identifying relevant clinical trials evaluating the efficacy of the therapies of interest in the same patient population. This was accomplished by conducting an SLR, which found three studies of entrectinib and one study of crizotinib with similar patient populations [75]. All of the studies were single arm studies and IPD was available for the three entrectinib studies, whereas aggregated level data was available for crizotinib. As all studies were single-arm trials, it was not possible to perform a network of connected RCTs for direct or indirect treatment comparisons. The feasibility assessment considered pooled data from the three entrectinib studies and concluded that it was feasible to perform a MAIC versus the crizotinib single arm trial. Sex, Eastern Cooperative Oncology Group (ECOG) performance status 0 or 1 versus 2, smoking history, age, disease stage at enrollment (stage IIIB versus stage IV non central nervous system [CNS] metastases versus stage IV CNS metastases) and prior treatments (treatment-naïve vs prior treatment) were considered as prognostic factors and effect modifiers. The MAIC suggested improved outcomes with entrectinib versus crizotinib [75].

At the same time data analyses using real world data from a database derived from electronic health records in the USA were conducted [7]. During the conduct of the study, researchers needed to ensure that the selected patient population receiving crizotinib in the EHR system adequately represented the entrectinib patient cohort from three open-label, single-cohort, phase I/II trials. To achieve this, the initial approach was to select a patient population that closely resembles the trial population. Predefined inclusion and exclusion criteria were applied to the EHR data. For example, excluded patients were those who received crizotinib in combination therapies, had an ECOG status >2 and carried other potential mutations such as ALK, BRAF, EGFR and KRAS. Through the identification of disparities in the patient populations, the authors could assess whether the study populations were adequately balanced in terms of confounders. As imbalances existed, the authors leveraged the IPD from both data sources to implement methods to adjust for measured confounding in observational research and utilized the inverse probability of treatment weighting (IPTW) method to compare TTD, Progression-Free Survival and overall survival between the entrectinib and crizotinib cohorts [7]. The comparative effectiveness analysis suggested improved outcome (TTD) with entrectinib versus crizotinib.

Entrectinib was granted accelerated approval by the FDA and received conditional marketing authorization in Europe [76,77]. From an HTA standpoint, the MAIC analysis could be leveraged and entrectinib received positive reimbursement decisions from various HTA bodies [78,79].

The two studies looking at CER of entrectinib versus crizotinib came to similar conclusions using different approaches. While the first one could be easily implemented by comparing IPD with published aggregated level data, the second one required access to the EHR database. If the flowchart were to be followed, it would have been sufficient to perform the first analysis to support both regulatory and HTA decisions.

Future directions & challenges**Recommendations for adoption of the methods flowchart**

For many decades, the scientific community held broad consensus that RCTs should be the basis for developing clinical guidelines and for decisions about marketing authorization, reimbursement and the development of treatment guidelines [80]. The ITC was developed to satisfy the need of healthcare practitioners facing an increasing

number of treatment options and to deal with situations when randomized trials comparing two innovative treatments with placebo or standard treatment, but not comparing the two treatments directly with one another, were available [81]. Historically, indirect comparisons have not been used by regulators as they are built on direct comparisons, which usually form the basis of decisions regarding regulatory approvals. RWE was gathered following conditional regulatory approval when direct comparisons were not feasible. In recent years there is a growing receptiveness to single arm trials and non-randomized evidence in situations where randomized studies are not feasible or practical [82,83]. Hence, there is a need to develop recommendations on methodological approaches and establish standards to facilitate the practical implementation of CER based on SAT or non-randomized RWE.

A systematic review of the scientific literature was conducted and existing regulatory and HTA guidance documents to identify recommendations and best practices. While multiple documents addressed specific elements – such as methods related to meta-analyses, network meta-analyses, or propensity score analyses – none provided a comprehensive overview or clearly identified which approach is most suitable under various conditions.

This perspective paper represents an initial attempt to address this gap. It was developed based on existing guidance and is not intended to introduce new methodologies, but rather to aid in selecting the most appropriate methods for a well-defined scientific question. The methods are designed to answer a specific question without imposing a hierarchy; therefore, the scientific question, not the methods, serves as the starting point of the proposed methods flowchart.

In research, ensuring the quality and validity of data is of utmost importance. Data quality refers to the accuracy, completeness and reliability of the data collected/used for CER, while data validity pertains to the extent to which the data measures what it is intended to measure. These are crucial as they directly impact the reliability and credibility of the research findings. Without addressing data quality and validity, any inferences or conclusions drawn from the data may lack scientific rigor. For this reason the flowchart starts with the well-defined research question and entails multiple aspects related to feasibility of the CER having this question in mind. Literature regarding feasibility of CER has been published [84].

Implications for regulatory, HTA & clinical decision making

By systematically evaluating the benefits and risks of various interventions in diverse populations, CER offers evidence-based insights that influence regulatory, HTA and clinical practice decision making [85]. As decision-making for health systems gets more complex, there is a need for a more standardized dynamic approach that acknowledges the impracticality of a one-size-fits-all design due to the diversity of patient populations and healthcare settings [86], and that can enhance faster evaluations and decision-making. Thus, harmonizing CER methodologies has profound implications, where evidence-based decision-making is crucial for developing sound healthcare policies [6,87]. Our work highlights a lack of harmonization between payers and regulators in the approaches to choosing CER methods and when to use CER in the decision-making process and proposes an approach to bridging this gap.

The proposed flowchart could support regulators, payers and healthcare professionals in conducting robust CERs and enable faster decision-making with the advent of budget pressure and constraints on healthcare systems to enable quicker decision-making.

Limitations

There are a few limitations of the proposed flowchart that we should highlight: the flowchart was specifically designed to help researchers identify the best methodological approach to answer a well-defined research question about comparative effectiveness when no previous knowledge on CER is available. When multiple pieces of evidence, both based on clinical trial and observational data, are available other approaches that combine cross designs may be explored [66,88–91].

There are other aspects that are very important in CER, but were not the focus of this research: The issue of multiplicity is equally important in randomized trials as in CER. Some good guidance on this topic, including the relevance of pre-specification, has been published for applications in clinical trials and their concepts can be adapted for use in CER [92]. Another important aspect is stakeholders' involvement, from the design of the research (e.g., feasibility of study) to the validity of the inferences. Frameworks to address this topic have been proposed elsewhere [93].

Conclusion

The proposed methods flowchart offers a clear guidance on CER methodologies across a range of settings and research needs. It begins with a well-defined research question and considers multiple feasibility aspects related to CER. This tool aims to standardize methods, ensure rigorous and consistent research quality, and promote a culture of evidence-based decision-making in healthcare.

Summary points

- Interventonal data and real-world data can provide supportive evidence for assessing the effectiveness of new therapies.
- Lack of common standards leads to inconsistency in methods used and lower reliability of and trust in comparative effectiveness research.
- None of the guidance documents identified from peer-reviewed literature, health technology assessment and regulatory websites provide a comprehensive overview to inform the choice of the methodological approaches that should be used in different situations when preparing for comparative effectiveness research.
- A methods flowchart is proposed to offer a clear guidance on comparative effectiveness research methodologies across a range of settings and research needs.

Financial disclosure

F Castro, M Daigl and S Abogunrin are employees and shareholders of F Hoffmann-La Roche Ltd., Basel, Switzerland. SF McGough is an employee and shareholder of Genentech Inc., CA, USA. C Boersma is a founder and CEO at Health-Ecore and a professor of sustainable Health and Innovation at The Open University and health economic researcher at the University of Groningen, The Netherlands. KR Abrams is a partner and director at Visible Analytics and a professor of Statistics and Data Science at the University of Warwick and honorary professor at the Centre for Health Economics at the University of York, UK. R Hendricks Sturup is a researcher at the Duke-Margolis Institute for Health Policy in Washington, DC, USA. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

1. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Ann. Intern. Med.* 151(3), 203–205 (2009).
2. Luke JJ. The role of comparative effectiveness research in developing clinical guidelines and reimbursement policies. *Virtual Mentor* 13(1), 52 (2011).
3. Price-Haywood EG. Clinical comparative effectiveness research through the lens of healthcare decisionmakers. *Ochsner J.* 15(2), 154–161 (2015).
4. Tugwell P, Knottnerus JA. Is the 'Evidence-Pyramid' now dead? *J. Clin. Epidemiol.* 68(11), 1247–1250 (2015).
5. Luce BR, Drummond M, Jönsson B *et al.* EBM, HTA, and CER: clearing the confusion. *Milbank Q.* 88(2), 256–276 (2010).
6. Wang SV, Schneeweiss S, Gagne JJ *et al.* Using real-world data to extrapolate evidence from randomized controlled trials. *Clin. Pharmacol. Ther.* 105(5), 1156–1163 (2019).
7. Doebele RC, Perez L, Trinh H *et al.* Comparative effectiveness analysis between entrectinib clinical trial and crizotinib real-world data in ROS1+ NSCLC. *J. Comp. Eff. Res.* 10(17), 1271–1282 (2021).

8. Gong Y, Kehl KL, Oxnard GR, Khozin S, Mishra-Kalyani PS, Blumenthal GM. Time to treatment discontinuation (TTD) as a pragmatic endpoint in metastatic non-small cell lung cancer (mNSCLC): a pooled analysis of 8 trials. *J. Clin. Oncol.* 36(Suppl. 15) DOI: 10.1200/JCO.2018.36.15_suppl.906 (2018).
9. Berger M, Daniel G, Frank K. A Framework for Regulatory Use of Real-World Evidence. Duke Margolis Center for Health Policy White Paper (2017). <https://healthpolicy.duke.edu/sites/default/files/2020-08/rwe.white.paper.2017.09.06.pdf>
10. FDA. Food & Drug administration work plan and proposed funding allocations of FDA innovation account (2017). <https://www.fda.gov/downloads/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCA/21stCenturyCureAct/UCM562852.pdf>
11. FDA. Framework for FDA's Real-World Evidence Program (2018). <https://www.fda.gov/media/120060/download?attachment>
12. FDA. Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry (2023). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug>
13. Nicotera G, Sferazza G, Serafino A, Pierimarchi P. The iterative development of medicines through the European Medicine Agency's Adaptive Pathway Approach. *Front. Med.* 6, 148 (2019).
14. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. *Clin. Pharmacol. Ther.* 106(1), 36–39 (2019).
15. EMA. HMA-EMA Joint Big Data Taskforce—summary report (2019). <https://www.ema.europa.eu/en/documents/minutes/hmaema-joint-task-force-big-data-summary-report-en.pdf>
16. Nishioka K, Makimura T, Ishiguro A, Nonaka T, Yamaguchi M, Uyama Y. Evolving acceptance and use of RWE for regulatory decision making on the benefit/risk assessment of a drug in Japan. *Clin. Pharmacol. Ther.* 111(1), 35–43 (2022).
17. Burns L, Roux NL, Kalesnik-Orszulak R *et al.* Real-world evidence for regulatory decision-making: guidance from around the world. *Clin. Ther.* 44(3), 420–437 (2022).
18. Facey KM, Rannanheimo P, Batchelor L, Borchardt M, de Cock J. Real-world evidence to support Payer/HTA decisions about highly innovative technologies in the EU—actions for stakeholders. *Int. J. Technol. Assess. Heal. Care* 36(4), 459–468 (2020).
19. Cook R, Lovell A, Khanzada A, Barham L. Value of real-world evidence in health technology assessment: lost in translation? *Economist Impact* (2022). <https://impact.economist.com/projects/rwe-in-hta/>
20. EU HTA CG. Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons (2024). https://health.ec.europa.eu/latest-updates/practical-guideline-quantitative-evidence-synthesis-direct-and-indirect-comparisons-2024-03-25_en
21. Page MJ, McKenzie JE, Bossuyt PM *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372, n71 (2021).
22. Cooke A, Smith D, Booth A. Beyond PICO. *Qual. Heal. Res.* 22(10), 1435–1443 (2012).
23. Alemayehu D, Cappelleri JC. Revisiting issues, drawbacks and opportunities with observational studies in comparative effectiveness research. *J. Eval. Clin. Pr.* 19(4), 579–583 (2013).
24. Neely JG, Sharon JD, Graboyes EM *et al.* Practical guide to understanding comparative effectiveness research (CER). *Otolaryngol. Head Neck Surg.* 149(6), 804–812 (2013).
25. Luce BR, Drummond MF, Dubois RW *et al.* Principles for planning and conducting comparative effectiveness research. *J. Comp. Eff. Res.* 1(5), 431–440 (2012).
26. Guise J, Viswanathan M. Overview of best practices in conducting comparative-effectiveness reviews. *Clin. Pharmacol. Ther.* 90(6), 876–882 (2011).
27. Tsertsvadze A, Maglione M, Chou R *et al.* Updating comparative effectiveness reviews: current efforts in AHRQ's Effective Health Care Program. *J. Clin. Epidemiology* 64(11), 1208–1215 (2011).
28. PBAC. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC) Section 2 Clinical evaluation (2016). <https://pbac.pbs.gov.au/section2-clinical-evaluation.html>
29. CADTH. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis (2009). https://www.cadth.ca/sites/default/files/pdf/H0462_itsc_tr_e.pdf
30. EUnetHTA. Levels of Evidence: internal validity of randomised controlled trials (2015). https://www.eunetha.eu/wp-content/uploads/2018/01/16.WP7-SG3-GL-int_val.RCTs_amend2015.pdf
31. EUnetHTA. Internal validity of non-randomised studies (NRS) on interventions (2015). https://www.eunetha.eu/wp-content/uploads/2018/01/Internal-validity-of-non-randomised-studies-NRS-on-interventions_Guideline_Final-Jul-2015.pdf
32. EUnetHTA. Comparators & Comparisons: direct and indirect comparisons (2024). <https://www.eunetha.eu/wp-content/uploads/2018/03/Direct-comparators-comparisons.pdf>
33. EUnetHTA. Levels of Evidence. Applicability of evidence for the context of a relative effectiveness assessment (2015). https://www.eunetha.eu/wp-content/uploads/2018/01/Levels-of-Evidence-Applicability-of-evidence-for-the-context-of-a-relative-effectiveness-assessment_Amended-JA1-Guideline_Final-Nov-2015.pdf

34. EU HTA CG. Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons (2024). https://health.ec.europa.eu/latest-updates/methodological-guideline-quantitative-evidence-synthesis-direct-and-indirect-comparisons-2024-03-25_en
35. IQWiG. General Methods Version 7.0 (2023). <https://www.iqwig.de/methoden/general-methods.version-7-0.pdf>
36. HAS. Indirect comparisons. Methods and validity (2019). https://www.has-sante.fr/upload/docs/application/pdf/2011-02/summary_report_indirect_comparisons_methods_and_validity_january_2011_2.pdf
37. HAS. Choices in methods for economic evaluation – HAS (2024). https://www.has-sante.fr/upload/docs/application/pdf/2020-11/methodological_guidance_2020_-choices_in_methods_for_economic_evaluation.pdf
38. HAS. Doctrine de la commission de la transparence (CT). Principes d'évaluation de la CT relatifs aux médicaments en vue de leur accès au remboursement (2023). <https://www.has-sante.fr/upload/docs/application/pdf/2021-03/doctrine.ct.pdf>
39. Vanier A, Fernandez J, Kelley S *et al*. Rapid access to innovative medicinal products while ensuring relevant health technology assessment. Position of the French National Authority for Health. *BMJ Évid.-Based Med.* 29(1), 1–5 (2024).
40. Dias S, Welton NJ, Sutton AJ, Ades A. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials (2014). https://www.ncbi.nlm.nih.gov/books/NBK310366/pdf/Bookshelf_NBK310366.pdf
41. Dias S, Welton NJ, Sutton AJ, Ades A. NICE DSU Technical Support Document 1: Introduction To Evidence Synthesis For Decision Making (2012). https://www.ncbi.nlm.nih.gov/books/NBK395870/pdf/Bookshelf_NBK395870.pdf
42. Ades A, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. NICE DSU Technical Support Document 7: Evidence Synthesis Of Treatment Efficacy In Decision Making: A Reviewer's Checklist (2012). https://www.sheffield.ac.uk/sites/default/files/2022-02/TSD7-reviewer-checklist.final_08.05.12.pdf
43. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades A. NICE DSU Technical Support Document: Inconsistency In Networks of Evidence Based On Randomised Controlled Trials (2014). https://www.sheffield.ac.uk/sites/default/files/2022-02/TSD4-Inconsistency.final_15April2014.pdf
44. Faria R, Alava MH, Manca A, Wailoo AJ. NICE DSU Technical Support Document: The Use Of Observational Data To Inform Estimates Of Treatment Effectiveness In Technology Appraisal: Methods For Comparative Individual Patient Data (2015). <https://www.sheffield.ac.uk/nice-dsu/tsds/observational-data>
45. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons In Submissions To NICE (2016). <https://www.sheffield.ac.uk/nice-dsu/tsds/population-adjusted>
46. NICE. CHTE2020 sources and synthesis of evidence (2020). <https://www.sheffield.ac.uk/nice-dsu/methods-development/chte2020-sources-and-synthesis-evidence>
47. NICE. NICE real-world evidence framework (ECD9) (2022). <https://www.nice.org.uk/corporate/ecd9/resources/nice-realworld-evidence-framework-pdf-1124020816837>
48. ICER. A Guide to ICER's Methods for Health Technology Assessment (2020). https://icer.org/wp-content/uploads/2021/01/ICER_HTA_Guide_102720.pdf
49. ENCePP. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCEPP) Guide on Methodological Standards in Pharmacoepidemiology. EMA/95098/2010 Rev.11 (2010). https://encepp.europa.eu/document/download/f6e403a6-8033-4c22-a5ff-195ba3666299_en?filename=01.ENCePPMethodsGuideRev.11.pdf
50. FDA. Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products Guidance for Industry (2023). <https://www.fda.gov/media/164960/download>
51. Richardson WS, Wilson MMC, Nishikawa MJ, Hayward MRSA. The well-built clinical question: a key to evidence-based decisions. *ACP J. Club.* 123(3), A12–A13 (1995).
52. Bhatt A. Evolution of clinical research: a history before and beyond James Lind. *Perspect. Clin. Res.* 3(4), 149–149 (2010).
53. Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin. Neurosci.* 13(2), 217–224 (2011).
54. Propes C, Sheehan S, Hendricks-Sturup R. Point-of-Care Clinical Trials: Integrating Research and Care Delivery (2022). <https://healthpolicy.duke.edu/sites/default/files/2022-05/Point-of-Care%20Clinical%20Trials%20Integrating%20Research%20Care%20Delivery.pdf>
55. EFPIA. An EFPIA position paper on randomised pragmatic trials to generate high-quality real-world evidence for regulatory decisions (2023). <https://www.efpia.eu/media/guokpw0b/an-efpia-position-paper-on-randomised-pragmatic-trials-to-generate-high-quality-real-world-evidence-for-regulatory-decisions.pdf>
56. Poklepović Perićić T, Tanveer S. Why systematic reviews matter. *Elsevier Connect* (2019). <https://www.elsevier.com/connect/why-systematic-reviews-matter>
57. Higgins JPT, Thomas J, Chandler J *et al*. (Eds). *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023)*. Cochrane (2023).

58. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network Meta-Analysis for Decision-Making*. (2018). <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118951651> Print ISBN:9781118647509 |Online ISBN:9781118951651
59. Sutton A, Ades AE, Cooper N, Abrams K. Use of Indirect and Mixed Treatment Comparisons for Technology Assessment. *Pharmacoeconomics* 26(9), 753–767 (2008).
60. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Méd. Decis. Mak.* 38(2), 200–211 (2017).
61. Phillippo DM, Dias S, Ades AE *et al.* Multilevel network meta-regression for population-adjusted treatment comparisons. *J. R. Stat. Soc.: Ser. A (Stat. Soc.)* 183(3), 1189–1210 (2020).
62. Kawahara T, Fukuda M, Oba K, Sakamoto J, Buyse M. Meta-analysis of randomized clinical trials in the era of individual patient data sharing. *Int. J. Clin. Oncol.* 23(3), 403–409 (2018).
63. Ramsey SD, Onar-Thomas A, Wheeler SB. Real-world database studies in oncology: a call for standards. *J. Clin. Oncol.* 42(9), 977–980 (2024).
64. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar. Behav. Res.* 46(3), 399–424 (2011).
65. Hendricks-Sturup R, Tian F, Izurieta H *et al.* Understanding the Use of Negative Controls to Assess the Validity of Non-Interventional Studies of Treatment Using Real-World Evidence (2023). <https://healthpolicy.duke.edu/events/understanding-use-negative-controls-assess-validity-non-interventional-studies-treatment>
66. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 327(7414), 557 (2003).
67. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat. Med.* 21(11), 1559–1573 (2002).
68. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat. Med.* 28(25), 3083–3107 (2009).
69. Pearl J. Causal diagrams for empirical research. *Biometrika* 82(4), 669–688 (1995).
70. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-Value. *Ann. Intern. Med.* 167(4), 268 (2017).
71. Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int. J. Epidemiology* 43(6), 1969–1985 (2014).
72. Degtiar I, Rose S. A review of generalizability and transportability. *Ann. Rev. Stat. App.* 10, 501–524 (2023).
73. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *J. Causal Inference* 1(1), 107–134 (2013).
74. Turner AJ, Sammon C, Latimer N *et al.* Transporting comparative effectiveness evidence between countries: considerations for health technology assessments. *Pharmacoeconomics* 42(2), 165–176 (2024).
75. Chu P, Antoniou M, Bhutani MK, Aziez A, Daigl M. Matching-adjusted indirect comparison: entrectinib versus crizotinib in ROS1 fusion-positive non-small cell lung cancer. *J. Comp. Eff. Res.* 9(15), 861–876 (2020).
76. FDA. FDA approves entrectinib for NTRK solid tumors and ROS-1 NSCLC (2019). <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-entrectinib-ntrk-solid-tumors-and-ros-1-nsclc>
77. EMA. Rozlytrek EPAR (2020). <https://www.ema.europa.eu/en/medicines/human/EPAR/rozlytrek>
78. CADTH. Entrectinib (Rozlytrek) for ROS1-positive Non-Small Cell Lung Cancer (2021). <https://www.cadth.ca/entrectinib-rozlytrek-ros1-positive-non-small-cell-lung-cancer>
79. NICE. Entrectinib for treating ROS1-positive advanced non-small-cell lung cancer (2020). <https://www.nice.org.uk/guidance/ta643/chapter/1-Recommendations>
80. Seale JP, GebSKI VJ, Keech AC. Generalising the results of trials to clinical practice. *Méd. J. Aust.* 181(10), 558–560 (2004).
81. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J. Clin. Epidemiol.* 50(6), 683–691 (1997).
82. Ribeiro TB, Bennett CL, Colunga-Lozano LE, Araujo APV, Hozo I, Djulbegovic B. Increasing FDA-accelerated approval of single-arm trials in oncology (1992 to 2020). *J. Clin. Epidemiol.* 159, 151–158 (2023).
83. Subbiah V. The next generation of evidence-based medicine. *Nat. Med.* 29(1), 49–58 (2023).
84. Walker A, Patrick AR, Lauer MS *et al.* A tool for assessing the feasibility of comparative effectiveness research. *Comp. Eff. Res.* 3, 11–20 (2013).
85. Marko NF, Weil RJ. The role of observational investigations in comparative effectiveness research. *Value Health* 13(8), 989–997 (2010).
86. Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report – Part I. *Value Health* 12(8), 1044–1052 (2009).
87. Andre EB, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: the changing landscape of the regulatory approval process. *Pharmacoevid. Drug Saf.* 29(10), 1201–1212 (2020).

88. Abogunrin S, Ashaye AO, Cappelleri JC *et al.* Safety and effectiveness of classical and alternative sunitinib dosing schedules for metastatic renal cell carcinoma: a meta-analysis. *Futur. Oncol.* 15(18), 2175–2190 (2019).
89. Hamza T, Chalkou K, Pellegrini F *et al.* Synthesizing cross-design evidence and cross-format data using network meta-regression. *Res. Synth. Methods* 14(2), 283–300 (2023).
90. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* 183(8), 758–764 (2016).
91. Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Res. Synth. Methods* 6(1), 45–62 (2015).
92. Li G, Taljaard M, Heuvel ERV den *et al.* An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int. J. Epidemiology* 46(2), 746–755 (2017).
93. PCORI. The Value of Engagement in Research (2024). <https://www.pcori.org/engagement/value-engagement-research#:~:text=The%20meaningful%20involvement%20of%20patients,study%2C%20and%20disseminating%20study%20results>