ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Systematic Literature Review

Analytical Methods for Comparing Uncontrolled Trials With External Controls From Real-World Data: A Systematic Literature Review and Comparison With European Regulatory and Health Technology Assessment Practice

Milou A. Hogervorst, PharmD, PhD, Kanaka V. Soman, MSc, Helga Gardarsdottir, PharmD, PhD, Wim G. Goettsch, PhD, Lourens T. Bloem, PharmD, PhD

ABSTRACT

Objectives: This study aimed to provide an overview of analytical methods in scientific literature for comparing uncontrolled medicine trials with external controls from individual patient data real-world data (IPD-RWD) and to compare these methods with recommendations made in guidelines from European regulatory and health technology assessment (HTA) organizations and with their evaluations described in assessment reports.

Methods: A systematic literature review (until March 1, 2023) in PubMed and Connected Papers was performed to identify analytical methods for comparing uncontrolled trials with external controls from IPD-RWD. These methods were compared descriptively with methods recommended in method guidelines and encountered in assessment reports of the European Medicines Agency (2015-2020) and 4 European HTA organizations (2015-2023).

Results: Thirty-four identified scientific articles described analytical methods for comparing uncontrolled trial data with IPD-RWD–based external controls. The various methods covered controlling for confounding and/or dependent censoring, correction for missing data, and analytical comparative modeling methods. Seven guidelines also focused on research design, RWD quality, and transparency aspects, and 4 of those recommended analytical methods for comparisons with IPD-RWD. The methods discussed in regulatory (n = 15) and HTA (n = 35) assessment reports were often based on aggregate data and lacked transparency owing to the few details provided.

Conclusions: Literature and guidelines suggest a methodological approach to comparing uncontrolled trials with external controls from IPD-RWD similar to target trial emulation, using state-of-the-art methods. External controls supporting regulatory and HTA decision making were rarely in line with this approach. Twelve recommendations are proposed to improve the quality and acceptability of these methods.

Keywords: external controls, health technology assessment, indirect treatment comparisons, individual patient data, marketing authorization, real-world data.

VALUE HEALTH. 2024; ■(■):■–■

Highlights

- Scientific literature on uncontrolled trials with external controls has expanded substantially in recent years, given that regulatory and health technology assessment (HTA) decision making increasingly relies on uncontrolled trials. To enable external control of these trials, individual patient data real-world data (IPD-RWD) can be valuable data, but a comprehensive overview of analytical methods for comparing uncontrolled trials with IPD-RWD–based external controls is lacking.
- To our knowledge, this article is the first to provide a systematic literature review of analytical methods proposed to compare uncontrolled trials with external controls from IPD-RWD. It demonstrates that many analytical methods are described in scientific literature and guidelines of European regulatory and HTA organizations. However, a large methodological gap was identified between these recommended state-of-the-art methods and the methods discussed in regulatory and HTA reports.
- Twelve recommendations for regulatory and HTA authorities were formulated to help them improve the quality and acceptability of the analytical methods used in submissions of IPD-RWD–based externally controlled trials. For externally controlled trials to be acceptable, it is critical to a priori develop a protocol using the target trial emulation approach to minimize bias and increase trust in the results. Advice on the analytical methods should be provided early and continuously through guidelines and scientific consultations.

Introduction

To assess the efficacy and safety of new treatments for regulatory or health technology assessment (HTA) decision making, randomized controlled trials (RCTs) remain the gold standard. Ethical or practical arguments sometimes justify the choice for an uncontrolled design in the case of rare diseases, a high unmet medical need, or a substantial expected benefit.[1] Uncontrolled trials include single-arm trials, randomized trials using multiple arms without control between those arms, or nonrandomized trials.[2,3] Owing to the lack of randomization, confounding and selection bias can affect the causal inference of efficacy and safety outcomes and complicate relative effectiveness assessments.[4,5]

Uncontrolled trials may be compared with external control arms. External control arms can be extracted from other RCTs or from real-world data (RWD) sources such as disease registries, electronic health records, or claims data.[6,7] RWD-based external controls are increasingly submitted to regulatory

and HTA authorities.[2,8,9] RWD allow for using individual patient data (IPD), which may provide more reliable results than aggregated data. RWD also have the potential to better represent the patients treated in clinical practice, improving the generalizability of the study findings.[10] Still, using RWD comes with challenges. The different data sources (ie, RWD vs clinical trial data) may essentially represent different patient populations that may be treated differently, for example, because they are treated at a different location or a different moment in time. When patient characteristics between uncontrolled trial and control arm differ or when treatment characteristics are not representative for current clinical practice, estimates of safety and efficacy may be biased.[3] Furthermore, the necessary data are often scattered over multiple sources and obtaining access to the IPD is not always feasible.[11] Data sources may not include all the relevant endpoints, and if they do, the methods for measuring and reporting these endpoints often differ. Besides, the real-world patient usually differs substantially from the trial patient and missing data are generally higher owing to the lack of an experimental setting.[12] These measurement inconsistencies and data missingness can introduce biases.

Additional bias is sometimes introduced by the methods that aim to correct for them. Adjusting for confounding through multiple regression modeling or propensity score matching assumes that all relevant confounding variables are included in the model. Not meeting this assumption leads to residual or unmeasured confounding.[13] Imputation methods for missing data can introduce bias if the missingness is not completely at random.[14] Time-related biases occur when the follow-up time is not properly classified or if time-varying confounders and outcomes are not adjusted for.[15]

The risk of bias and difficulty interpreting the results from externally controlled trials may lead to divergent regulatory and HTA recommendations across countries and delayed or hampered patient access.[8,16] Simultaneously, pharmaceutical developers have been struggling to define a common approach to applying external control arms owing to the lack of a shared international methodological framework.[17] Despite the wealth of publications on the topic focusing on the previously described biases and methods, no guidelines nor literature provides a complete overview of the available methods with preferences and directions on when and how to use them.[2,7,18-24] Besides, it is unclear whether and how methodological advancements are applied for marketing authorization and reimbursement.[8,24]

Such understanding should provide input for the method alignment efforts under the European HTA regulation and for regulatory guidance development on uncontrolled trials and real-world evidence.[25] Therefore, we aimed to (1) systematically review the analytical methods available for comparing uncontrolled trials with IPD-RWD–based external controls and (2) compare these findings with European regulatory and HTA practice (ie, guidelines and assessment reports). Using these results, we formulated recommendations for regulatory and HTA authorities to help them improve the quality and acceptability of the methods used in submissions.

Methods

This study comprised 2 parts. First, a systematic literature review was performed to identify analytical methods for comparing uncontrolled trials with external controls derived from IPD-RWD (ie, selecting and preparing the cohort and comparative analytical methods). Second, the identified methods in scientific literature were compared with the methods recommended for RWD-based external controls in guidelines from European regulatory and HTA organizations and those discussed in regulatory and HTA reports to identify potential discrepancies between methodological advancements and decision-making practices. In regulatory and HTA guidelines and assessment reports, both IPD-RWD–based and aggregated RWD-based external controls were identified, to place the use of IPD-RWD specifically in perspective.

Scientific Literature Search and Selection Strategy

A systematic literature search was conducted in consultation with a research librarian of the university and following established guidelines for systematic reviews.[26] PubMed was searched for scientific articles up until March 1, 2023, using combinations of the concepts “uncontrolled study” and “external control” as search string (Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002). Reference lists of included articles were also reviewed to identify additional relevant articles (snowballing). In addition to the manual snowball approach, the Connected Papers tool (Connected Papers, n.d., Tel Aviv, Israel) was used to identify additional relevant articles.[27] Four key articles that were considered to exemplify the type of studies relevant to the research question were uploaded in the tool.[28-31] The algorithm generated a list of 160 articles (40 for each uploaded article) that were considered “connected” by the tool through their authors or the topics in the title and abstract.

We aimed to identify studies providing overviews of analytical methods for comparing uncontrolled trials with external controls from IPD-RWD, including case studies with methodological focus and studies reporting the development of new methods. Inclusion criteria were articles reporting on all of the following: (1) methods for selecting and preparing a control cohort or comparing external control cohorts with uncontrolled trials; (2) use of IPD from RWD; (3) detailed description of a methodological approach, strategy, or technique; and (4) focus on medicines. Exclusion criteria were articles reporting on any of the following: (1) RCTs, (2) studies replicating or augmenting trials using RWD, (3) prediction modeling with RWD, (4) fully observational studies, (5) uncontrolled trials for nonmedicines, or (6) articles with inadequate information on methods. No restrictions were applied to the publication date or language. Two authors (K.S., M.H.) independently screened the titles and abstracts using Rayyan (Cambridge, MA).[32] Uncertainties, discrepancies, and the final list of included articles were discussed with a third author (L.B.) until consensus was reached. Full-text articles were then assessed for eligibility according to the inclusion and exclusion criteria. Any discrepancies among the reviewers were resolved through discussion and consensus. The study selection process was documented using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

Method Guideline Search and Selection Strategy

The websites of the European Medicines Agency (EMA), 4 national HTA organizations, and the European Network on HTA (EUnetHTA) were searched for methodological guidance on uncontrolled trials and comparison with external control arms using RWD (until March 2023). HTA guidelines were searched for the French Haute Autorité de Santé (HAS), the German Gemeinsamer Bundesausschuss and Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), the English and Welsh National Institute for Health and Care Excellence (NICE), and the Dutch National Health Care Institute. These countries have front-running HTA organizations that provide recommendations that are the basis for national reimbursement decisions and publish method guidelines that are publicly available in languages matching the authors’ proficiencies.

Regulatory and HTA Report Search and Selection Strategy

The EMA's European Public Assessment Reports (EPARs) for newly authorized medicines based on uncontrolled pivotal trials (2015-2020) were identified in a previous study.[33] Nonauthorized medicines, diagnostics, vaccines, duplicate marketing authorizations, generics, biosimilars, and fixed-dose combinations were excluded. The EPARs were manually examined to determine whether the assessments were performed solely based on uncontrolled pivotal trials. Missing EPARs were requested from the EMA. To collect comparable HTA reports from the respective 4 national organizations, the EPAR cohort of medicines was also searched for subsequent HTA (2015-2023). Only the initial assessments solely based on uncontrolled pivotal trials were included. All reports assessing uncontrolled trials were reviewed to identify those that used RWD-based external controls and whether IPD or aggregated data were used.

Study Quality and Publication Bias Assessment

Given the focus on methodological approaches and our aim to provide an overview of these methods rather than to perform quantitative assessments or prioritize/weigh studies, a formal quality assessment, risk of bias assessment, including publication bias, was not considered applicable. However, clarity of the included studies was evaluated during the selection process, given that adequate information was one of the inclusion criteria.

Data Extraction and Analysis

Data extraction from eligible studies was performed by M.H. and verified by L.B. for accuracy and completeness; see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002 for the form. Extracted information included the addressed methods, how these were applied, reasons for external control or the specific method, sensitivity analyses, biases, and other comments on the strengths and limitations of the methods. The methods were organized based on guidelines for good pharmacoepidemiology practices (GPP) of the International Society for Pharmacoepidemiology (ISPE).[34] A narrative synthesis approach was used to summarize the extracted information for each step, but mainly focusing on the analytical methods (methods for data analysis under protocol development in the ISPE GPP guideline).[34] The methods were grouped and listed based on how they were described in the scientific literature, ie, with which purpose they were applied or at which point in the process. This purpose was briefly described. Discrepancies in interpretation were resolved through consultation with H.G. and S.B.

The methods recommended in the method guidelines were extracted and presented according to whether they considered the availability of IPD or aggregated data. In addition, data were extracted from EPARs and HTA reports by M.H. and L.B. using a standardized form; see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002. The methods described in literature, guidelines, and regulatory and HTA reports were qualitatively compared. Recommendations to improve the quality and acceptability of analytical methods for external comparisons using IPD-RWD in regulatory and HTA practice were formulated based on the information extracted from literature and the gap analysis. These recommendations may be used by regulators and HTA organizations, for example, to formulate guidance.

Results

Systematic Literature Search

The systematic literature search resulted in 34 relevant articles describing methods for comparing uncontrolled trials with external controls from IPD-RWD (Fig. 1). Most of the relevant articles comprised (nonsystematic) reviews or overview articles[29,35-48] (n = 15) and methodological articles[15,28,30,49-59] (n = 14). Other articles consisted of reflection articles[60-62] (n = 3) and commentaries or editorials[63,64] (n = 2).

The methods and steps for conducting trials with external controls from IPD-RWD were summarized according to the ISPE GPP guideline[34] (Fig. 2). The key is to a priori write a protocol for the entire externally controlled study. The first step is to formulate the research question and determine the estimands (Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002). The research question and estimands will determine what is the most suitable data source and study design. The estimand and design should be informed by the target trial emulation approach, as described by Hernán and Robins,[65] to address selection bias and immortal time bias. With a protocol and selected data, the external cohort can be extracted from the RWD source, to match the uncontrolled trial cohort. Subsequently, the cohorts can be compared with each other by selecting the exact patients that will be compared and using statistical analysis methods. Sensitivity analyses should ensure the robustness of the results. Transparent reporting of protocols, methods, and results is essential for proper interpretation and estimating the remaining uncertainty.

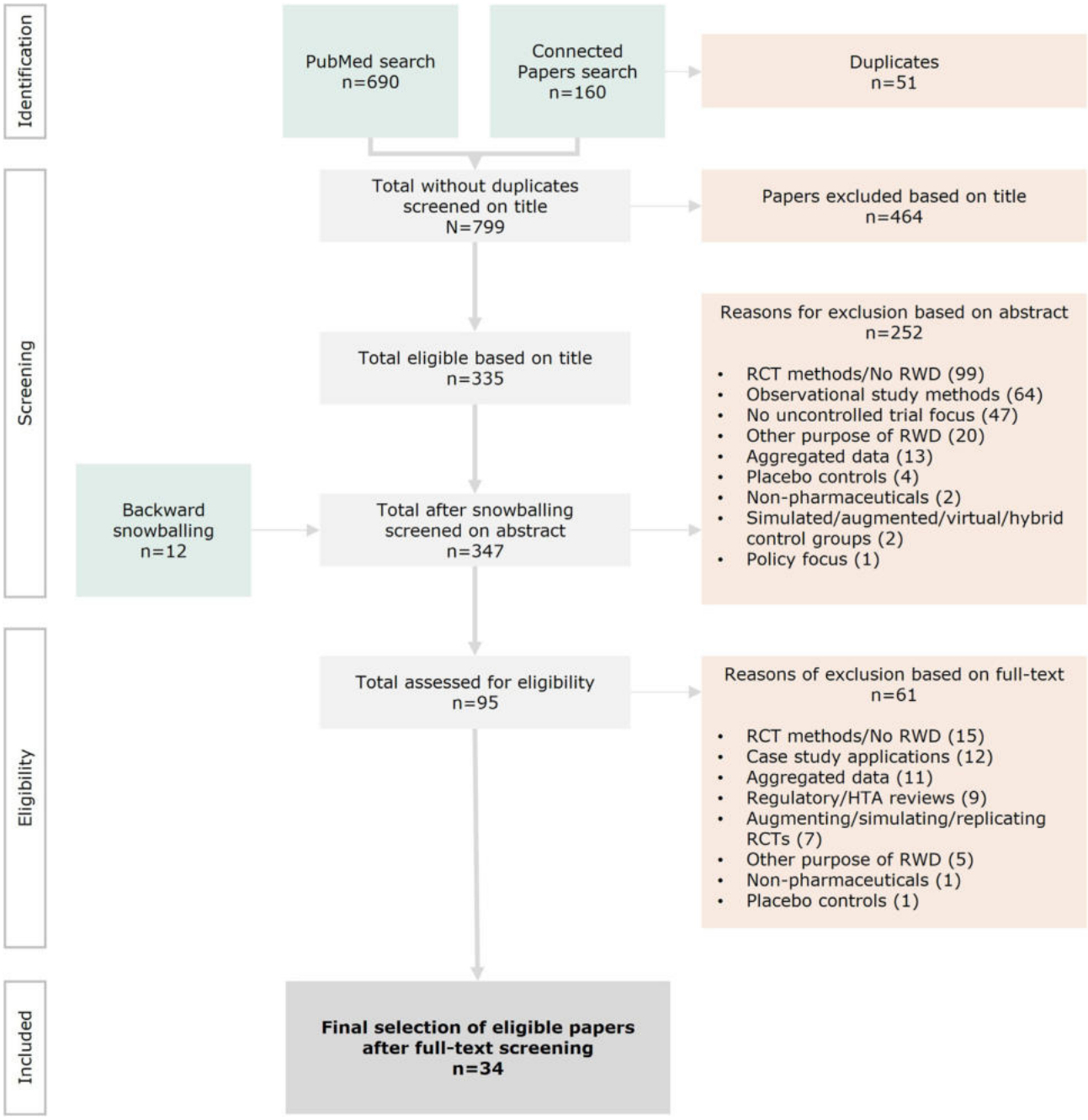
The analytical methods to prepare and compare external cohort data with uncontrolled trial data that were discussed in the 34 scientific articles that were obtained are presented in Table 1. The 4 independent steps and categories of analytical methods that were considered were (1) controlling for confounding, (2) controlling for dependent censoring, (3) correction for missing data, and (4) the analytical comparative modeling method to use. Concerning considerations to control for potential confounding, the major decision identified in the scientific literature was between not controlling for confounding and performing naïve analysis, described in 2 of 34 scientific articles but not considered appropriate, versus controlling through matching (10 of 34), restriction (4 of 34), stratification (7 of 34), correction (ie, including variables in the statistical model; 6 of 34), weighting (13 of 34), and other methods (5 of 34). Of those, several methods may allow the use of a propensity score to efficiently control for multiple potential confounders at the same time, with methods to generate a propensity score discussed in 12 of 34 scientific articles. Concerning considerations to control for dependent censoring, ie, censoring related to the outcome, censoring weights were discussed (1 of 34). Concerning considerations to correct for missing data, complete case selection (1 of 27), full cohort selection (1 of 34), and, when selecting the full cohort, imputation of missing data were discussed (1 of 34). Finally, concerning considerations for the choice of analytical comparative modeling method, regression modeling (4 of 34), meta-analytic methods (1 of 34), advanced exploratory solutions (1 of 34), pseudo-observations (1 of 34), marginal structural modeling (1 of 34), microsimulation (1 of 34), g-computation (1 of 34), machine learning methods (2 of 34), and doubly debiased machine learning (1 of 34) were discussed. A more detailed description and the scientific articles in which a category of methods or specific method was described are described in Table 1 under "source."

Regulatory and HTA Practice

Guidelines

Seven guidelines that covered (indirect) treatment comparisons were included, of which only one (IQWiG) had a pure RWD focus.[2,7,18,19,21,22,66] All the guidelines emphasized that RCTs are preferred but none of the guidelines explicitly stated in which

Figure 1. PRISMA flow diagram visualizing the results from the systematic literature search.



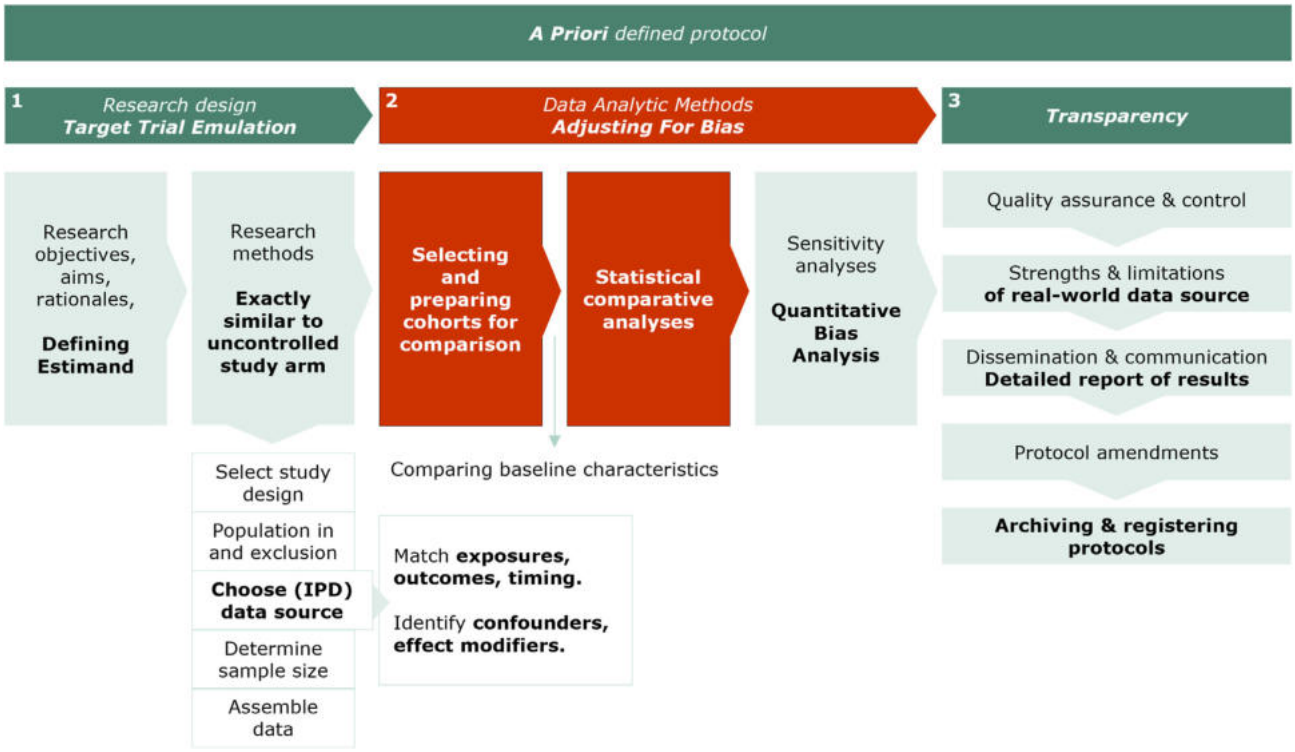
HTA indicates health technology assessment; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomized controlled trial; RWD, real-world data.

cases externally controlled trials are acceptable.[2,7,18,66] The guidelines from the EMA, HAS, and IQWiG had a strong focus on general research design (Appendix 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002). The EMA reflection article stated that isolation of treatment effects can be sufficient if an uncontrolled trial results in a dramatic effect but that "the choice of comparator and comparative data is out of scope of the paper."[2]

An overview of analytical methods recommended for comparing uncontrolled trials with RWD-based external controls is provided in Figure 3. HAS' position statement on uncontrolled trials provided examples of possible IPD-RWD methods based on propensity scores or g-computation.[22] IQWiG did not discuss uncontrolled trials in its guidelines, given that they did not consider them feasible for drawing causal effect-based conclusions without a common (anchored) comparator. This implies that external controls from RWD are in principle not considered sufficient for decision making.[67]

The NICE and EUnetHTA guidelines described more analytical methods including their application, although without stating

Figure 2. A visual summary of the stepwise methodological approach for comparing uncontrolled trials with external controls derived from individual patient real-world data, based on the good pharmacoepidemiology practices guideline of the International Society for Pharmacoepidemiology (ISPE).[34] In bold are the elements that were described in the scientific literature; the orange/red elements are the focus of the rest of the results.



clear preferences for methods (Appendix 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002).[7,19] The NICE guidelines stated a preference for anchored comparisons whenever possible. When IPD is available, options for comparison include (inverse) propensity score methods (weighing), outcome regression, or doubly robust methods.[7] In addition, they indicated specific situations where instrumental variables, panel data models or regression on a matched sample is feasible. If IPD are lacking, matching-adjusted indirect comparisons (MAICs) and simulated treatment comparisons are 2 methods discussed, but for MAIC more limitations and potential for bias are described.[7,19] NICE also discussed the multilevel network meta-regression, which provides more precise treatment effect estimates than simulated treatment comparison and can handle both comparisons based on IPD and aggregated data.[7] However, these methods can only be used if the assumption of "conditional constancy of absolute effects" is met (ie, the absolute outcome in the treatment arms is assumed to be constant at any given level of the prognostic variables and effect modifiers), which is rarely the case in post hoc designed studies using aggregated data.[19] The NICE guideline indicated that unanchored comparisons, ie, using external controls, and the use of aggregated data warrant additional attention and research.[7] Similarly, the EUnetHTA guideline explicitly stated that aggregated data for external controls are not sufficient to produce reliable results because, in the case of unanchored comparisons, the comparison relies on stronger assumptions.[19] In line with the EMA guideline, the EUnetHTA guideline stated that the greater uncertainty associated with nonrandomized data requires a large estimated effect of the treatment.[2,19] Naïve (unadjusted) comparisons should not be used owing to the risk of bias, according to NICE and EUnetHTA.[7,19] HAS and NICE both recommend a doubly robust approach, combining propensity score weighing and regression-based techniques to simultaneously estimate effects while accounting for confounders.[7,22]

Regulatory assessment reports

We identified 15 EPARs in which submitted uncontrolled trials were compared with external controls from RWD (Table 2). IPD were available for most comparisons (n = 12); however, these were often applied in a naïve comparative way (n = 8), without statistical bias adjustments. Other methods for comparison were regression techniques (n = 5) and matching techniques (n = 3) to improve the comparability of the 2 arms. The EMA mostly considered external controls to contextualize results often as a benchmark or for "exploratory" purposes. A specific scenario for external controls was to assess time-to-event outcomes, more specifically survival outcomes. The EPARs sometimes stated that owing to the difficulty of getting access to high-quality IPD-RWD, comparisons were complicated because estimates could not be adequately adjusted for baseline characteristics.

HTA reports

We identified 41 HTA reports in which the submitted uncontrolled trials were compared with external controls from RWD using one or more analytical methods (Table 2). Most submitted comparisons with external controls were based on aggregated data (n = 26). Aggregated data approaches were more frequently disregarded in the recommendation than the IPD approaches (23% [6 of 26] vs 5% [1 of 19]). Some reports did not clearly describe which methods for comparison were used (n = 15). Owing to the

Table 1. Methods to prepare and compare the external cohort to the uncontrolled trial.

Steps and Methods	Description	Source
1. Consider controlling for potential confounding		
1.1. Not controlling for potential confounding		
1.1.1. Naive analysis	Includes comparing simple means, medians or fixed-effect pooling, but not considering patient characteristics and confounding factors is often not appropriate.	19,65
1.2. Controlling for potential confounding		19,27,37-39,42,48,52,54,56,59,63,65
1.2.1. Matching		
1.2.1.1. (Partial) Matching	As a common approach, matching involves selecting individuals (one or more) from the external RWD cohort who are similar to those in the uncontrolled cohort based on covariates such as age, sex, or comorbidities, to balance these covariates. Matching can be performed using various techniques such as nearest neighbor matching, or propensity score matching.	27,29,42,44,48,52,56,58,63,65
1.2.2. Restriction		
1.2.2.1. Restriction or trimming	Exclude patients for whom there is no clinical equipoise or nonoverlapping regions of certain variables before (naively) comparing.	27,29,37,38
1.2.3. Stratification		
1.2.3.1. (Fine) stratification	Stratification involves dividing the data into subgroups based on specific baseline characteristics, or based on propensity scores, before comparing. Stratification can be useful when there are a few important covariates that strongly influence the outcome of interest.	27,29,37,38,52,56,58
1.2.4. Correction (including variable(s) in the statistical model)		
1.2.4.1. Regression or (imbalance) adjustment	The propensity score can be used as an adjustment variable in a regression (for example, Cox) model to estimate the treatment effect while adjusting for the covariates, as an alternative to including all the variables individually. This is suitable when there are multiple covariates to be considered and when the relationship between covariates and the outcome is complex. A standardized (mean) difference or covariate balance diagnostics may be used to measure imbalance. This is useful when there are specific covariates known to be associated with the outcome and need to be accounted for in the analysis.	19,27,29,52,56,65

continued on next page

Table 1. Continued

Steps and Methods	Description	Source
1.2.5. Weighting		
1.2.5.1. Weighing methods	Weighing involves assigning weights to individuals in the RWD cohort based on their propensity scores (inverse probability treatment weighing) or rather straightforward covariate balancing without a direct estimation of propensity scores. Weighing is suitable when achieving balance on covariates is the primary concern, and the goal is to create a weighed sample that is representative of the uncontrolled cohort or to determine the value of individual patients in the analysis. Inverse probability treatment weighing is one of the most common approaches, together with matching on the PS, and estimates the average treatment effect on the treated and the average treatment effect estimands.	15,27,35-38,44,48,52,56-58,65
1.2.6. Other methods		
1.2.6.1. IV methods	IVs can account for unmeasured confounding by leveraging the relation between an IV and the assignment of treatment.	37,38,48,56
1.2.6.2. Doubly robust method	Doubly robust methods are an alternative to pure PS methods, combining PS weighing with an outcome model, which results in improved robustness, namely that only one of the 2 models (PS or outcome model) has to be specified correctly to obtain unbiased results.	56
1.2.6.3. Complex adjustment and weighing (Bayesian)	Bayesian methods can be applied with the intent to discount historical information, for example, due to low data quality or strong heterogeneity when using different data sources.	19,56
1.3. If required (based on earlier design choices): generate propensity score		15,27,29,42,43,48,51,52,54,65
1.3.1. Methods to estimate a propensity score		
1.3.1.1. Logistic regression	The most frequently used technique to calculate the predicted probabilities of being exposed to experimental treatment versus a given comparator, conditional on the manually selected covariates. Standard or multivariable regression, examples include lasso and ridge regression.	27,29,42,48

continued on next page

Table 1. Continued

Steps and Methods	Description	Source
1.3.1.2. Machine learning techniques/ High-dimensional algorithms	Calculates the predicted probabilities of being exposed to experimental treatment versus a given comparator, conditional on the artificial intelligence–selected covariates. Simultaneously models propensity scores and selects variables for adjustment. Can be used to generate high-dimensional propensity scores. High-dimensional propensity scores algorithms are a specific application of machine learning. It creates binary variables describing the frequency of diagnoses, procedures, and medication dispensations. A bias approximation is calculated to estimate the expected bias in the treatment effect estimate for each variable. This approximation considers the variable's prevalence and its univariate associations with both treatment and outcome. The bias calculation is utilized to prioritize variables for inclusion in the propensity score model.	42,48
1.3.2. Methods to optimize the estimated propensity score		
1.3.2.1. Trimming propensity scores	Exclude patients for whom there is no clinical equipoise or nonoverlapping regions of the propensity score.	27,29,37,38
2. Consider controlling for dependent censoring (censoring related to the outcome)		
2.1. Censoring weights (IPCW)	These scores adjust the analysis by giving higher weights to censored individuals who share similar characteristics with noncensored individuals and is used to address right censoring of the outcome.	15
3. Consider correction for missing data		43,66
3.1. Complete case selection	Manages missing data by restricting the inclusion of patients in the external arm to those without any missing data to increase direct comparability between uncontrolled cohort and external control. Decreases sample size (thus power) and may introduce bias if the missingness is not completely at random.	27
3.2. Full cohort selection	Includes all patients, also if data are missing. Used if the missingness is not completely at random or the cohort is very small. It may require imputation (of missing data) and sensitivity analysis during the analysis phase to validate the results.	15
3.2.1. If selecting the full cohort: consider imputation of missing data	The usual recommendation for imputation is the technique of multiple imputation, which is a standard procedure available in many statistical software packages under the assumption that data are missing at random.	27

continued on next page

Table 1. Continued

Steps and Methods	Description	Source
4. Consider the analytical comparative modeling method to use: performing the analysis and answering the research question		27
4.1. Regression modeling	This may include ordinary multivariable regression, linear regression, logistic regression, proportional hazards, and others. Standard regression may risk overfitting and requires an adequate number of patients/events per covariate, which is particularly difficult for external controls with often limited sample sizes and numerous covariates.	19,27,48,56
4.2. Meta-analytic methods	Meta-analytic methods model the unexplained heterogeneity between the source data and account for this when extrapolating to the target, based on aggregate data. If more complex meta-analytic approaches are needed to synthesize evidence, meta-regression methods may be applied.	65
4.3. Advanced exploratory solutions	Cluster analysis (such as Gaussian mixture models) can help identify subgroups or patterns within the cohorts, highlighting differences or similarities that may be of interest. Random forests and neural networks can be useful tools in comparing cohorts as they can handle complex interactions between variables and nonlinear relationships. Neural networks can capture complex patterns and relationships in the data but can be computationally intensive. These methods provide associations rather than causal inferences.	19
4.4. Pseudo-observations	The pseudo-observations approach creates a hypothetical comparison group from the RWD cohort. Pseudo-observations aim to mimic the outcomes that would have been observed if the RWD cohort had been treated with the same intervention as the uncontrolled trial cohort. This allows for the creation of a counterfactual group that serves as a comparison to the uncontrolled trial cohort.	56
4.5. Marginal structural modeling	Based on the full external cohort, time-varying covariates are updated, and inverse probability treatment weights and censoring weights are applied to create a pseudo-population balanced concerning the confounding and censoring factors.	15
4.6. Microsimulation	Microsimulations of synthetic controls can be linked to uncontrolled trials or uncontrolled prospective patient registries to aid in the estimation of a comparative effect, particularly in settings where a long-term clinical trial is infeasible.	19
4.7. G-computation	G-computation is based on the counterfactual framework in which we posit that we can predict a patient outcome if the patient would have been enrolled in the control arm instead of the experimental one or vice-versa. This makes the inference of a causal effect theoretically possible.	54

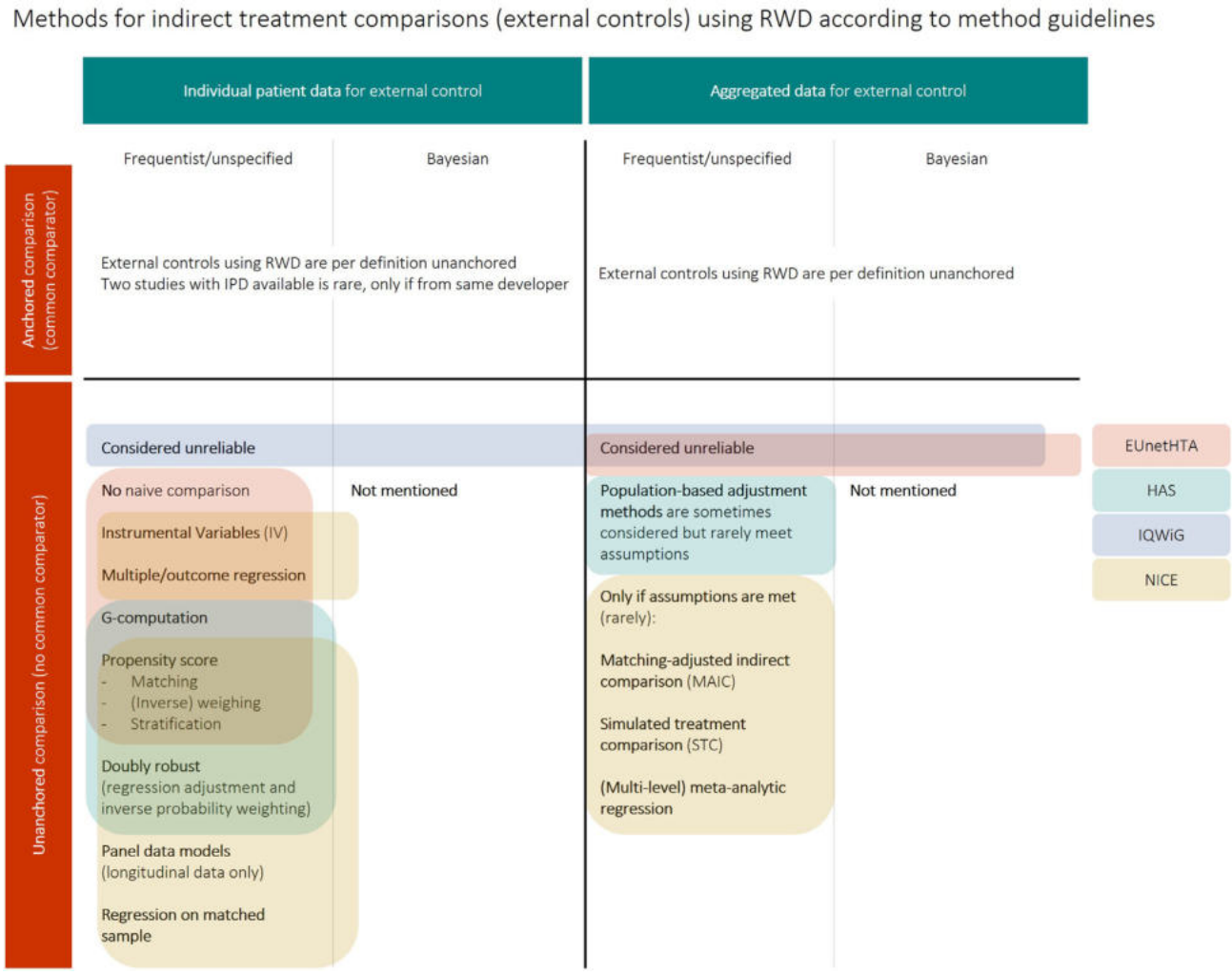
continued on next page

Table 1. Continued

Steps and Methods	Description	Source
4.8. Machine learning methods	Outcome prediction methods involve estimating the expected clinical outcome based on patient covariates. Flexible machine learning models such as boosted trees or neural networks can be used to fit the model.	54,56
4.9. Doubly debiased machine learning	Doubly debiased machine learning is related to G-computation, but it further accounts for the possible bias of machine learning outcome models.	54

See also Figure 2. The strategy for these steps should be explored a priori but are not always required altogether. These methods are described as they were described in the corresponding scientific literature reviewed for this study. More options may be available. A selection of the steps and/or a combination of methods could be applied. Multiple methods within each step could be applied as scenarios or sensitivity analyses. This depends on the estimand and strategic choices. References are reported at the table level that corresponds to the level of detail in which the method was initially described.
IV indicates instrumental variable; IPCW, inverse probability of censoring weighting; RWD, real-world data.

Figure 3. Overview of methods in regulatory and HTA method guidelines.



EUnetHTA indicates, European Network on HTA; HAS, Haute Autorité de Santé; HTA, health technology assessment; IPD, individual patient data; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; NICE, National Institute for Health and Care Excellence; RWD, real-world data.

Table 2. Overview of the number of EMA and HTA reports based on uncontrolled trials and the methods that were used to compare these to an external control.

Data source	Individual patient data for external control					Aggregated data for external control				
N = submitted (considered)	EMA n = 12	HAS n = 4	IQWiG n = 6	NICE n = 5	ZIN n = 0	EMA n = 3	HAS n = 5	IQWiG n = 9	NICE n = 9	ZIN n = 3
Naive or unadjusted	8	0	2 (1)	1 (1)	0	3	0	4 (0)	6 (6)	0
Regression techniques	5	1 (1)	0	2 (2)	0	0	1 (1)	1 (1)	1 (0)	0
Matching/weighing techniques	3	2 (2)	1 (1)	3 (3)	0	1	3 (3)	1 (1)	7 (7)	3 (3)
Unclear	0	2 (1)	4 (4)	1 (1)	0	0	2 (2)	4 (3)	1 (1)	1 (1)

In some assessments, multiple comparisons were considered; hence, the figures do not always add up. The numbers represent the comparisons that were submitted; the numbers between brackets represent the comparisons that were considered for the recommendation.
EMA indicates European Medicines Agency; HAS, Haute Autorité de Santé; HTA, health technology assessment; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; NICE, National Institute for Health and Care Excellence; ZIN, Dutch National Health Care Institute.

brief descriptions, we were only able to distinguish methods to control for confounding as "regression techniques" (n = 6) and weighing and matching (n = 20). The remaining analyses comprised naïve comparisons (n = 13) that were most frequently disregarded in the recommendation (38% [5 of 13] vs 17% [1 of 6] of regressions, 0% [0 of 20] of matching/weighing techniques, and 13% [2 of 15] of the unclear methods). Across all institutions, a priori defined analyses were better appreciated than post hoc analyses (see case studies on axicabtagene ciloleucel and tisagenlecleucel in Appendix 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.08.002).

Discussion

We aimed to (1) systematically review the analytical methods available for comparing uncontrolled trials with IPD-RWD–based external controls and (2) compare these findings with European regulatory and HTA practice (ie, guidelines and assessment reports). More advanced methods for comparing uncontrolled studies with external controls using IPD-RWD were described in literature than were mentioned in guidelines or encountered in regulatory or HTA reports.

Many guidelines focused on research design aspects rather than the analytical methods for comparison but if analytical methods were discussed, these were also discussed in literature. The NICE and EUnetHTA guidelines discussed analytical methods based on whether IPD or a common comparator is available. However, these guidelines and literature suggested different preferred methods (g-computation, instrumental variables) than those that were encountered in regulatory and HTA reports (naive, MAICs, simple regression methods). None of the guidelines recommended methods for aggregated external control data unless the underlying assumptions were met, which rarely happens. Thus, despite these discrepancies between scientific literature and guidance from regulators and HTA organizations, guidelines were generally in line with recommendations in scientific literature. In contrast, aggregated external control data were often discussed in regulatory and HTA reports, suggesting that guidelines may not be followed by developers. Notably, guidelines often did not describe a preference for any of the methods. Differences among guidelines of different organizations may be explained by differences in timing of publication and focus and perceptions on the acceptability of risk and uncertainty by organizations. Based on our findings, we formulated recommendations for regulatory and HTA authorities to help them improve the quality and the acceptability of the methods used in submissions (Table 3).

Two recent studies qualitatively assessed regulatory and HTA evaluations of uncontrolled trials using an external control from RWD.[23,24] Sola-Morales et al[23] identified operational and methodological aspects for which guidance is required: early engagement with regulators and HTA organizations, handling missing data, and the selection of real-world endpoints. They concluded that the extent to which regulators and HTA organizations consider the same external controls from RWD is highly variable and calls for alignment across institutions, which is in line with our findings. The negative recommendations they identified were mostly owing to research design and data quality aspects (eg, the lack of randomization or differences in endpoints, exposure, or population between the compared cohorts). The authors did not highlight analytical comparative methods as the reason for a negative recommendation, although we found that not all methods were considered in the recommendations (mostly for aggregated data and naive comparisons). Curtis et al[24] outlined methodological considerations for the design, conduct, and reporting of external controls from RWD. They focused on the research design and data quality elements, describing that for regulatory and HTA submissions it is critical to consider the appropriateness of external controls, ensure adequate sample sizes, implement a clear strategy for addressing data quality, select appropriate endpoints, and conduct sensitivity analyses. The choice of analytical methods was not considered in this study. In line with our recommendations (Table 3), they also concluded that early engagement with regulatory and HTA organizations was essential.

Many other institutions, working groups, and academics published on the use of RWD to generate external controls for regulatory and HTA decision making.[3,23,68-71] Recommendations in (draft) guidelines from the Food and Drug Administration (FDA) and Canadian Agency for Drugs and Technologies in Health reiterate our findings on the need for a priori defined protocols, including analyses and data collection procedures.[3,69] The FDA guideline does not specify analytical methods, stating that "no single statistical or analytical method will be suitable for all trials involving external control arms, and potential approaches should be discussed with the FDA."[3] This underlines the added value of our current methodological overview. ISPE has endorsed a manuscript of its members that discussed types of external control arms based on RWD and how to mitigate biases. In line with our results, they describe that when sufficiently justified

Table 3. Based on our results, several recommendations to regulatory and HTA authorities were formulated to improve the quality and acceptability of analytical methods for externally controlling trials using real-world data.

Research design
1. To reduce bias related to research design, a priori (at the time of trial design) defined study protocols are required, including an analysis plan for the indirect comparison to external control from IPD-RWD.
2. The study protocol and design should emulate a target RCT, in which the uncontrolled trial and the comparison with an external control from IPD-RWD are treated as one trial, ie, design as an externally controlled trial rather than an uncontrolled trial.
3. The protocol should clearly define the research question, including the more specific estimand before designing the study.
4. Guidance or advice for situations in which the use of an external control from IPD-RWD is acceptable, and in which situations it is not acceptable, may build trust and enhance acceptance.
Data quality
5. Using RWD for external controls imposes different limitations compared with using external controls from other trials. An overview of specific considerations for RWD limitations (and the methods that may be used to address these limitations, see Analytical methods) could inform trial conductors when choosing IPD-RWD sources.
6. High-quality IPD-RWD should be better available and accessible.
Analytical methods
7. An overview of when to use what method could guide developers or academics when conducting comparisons because not all methods are suitable for answering every research question or RWD limitation.
8. Given that a doubly robust method is preferred, an overview of methods that can or should be combined should be created.
Transparency and communication
9. Guidance on consistent reporting of methods and results for externally controlled trials using IPD-RWD in regulatory and HTA reports is required to improve the interpretation of results.
10. Aligning terminology for methods for externally controlled trials could support understanding across stakeholders.
Stakeholder alignment
11. Discussion among stakeholders (regulators and HTA organizations) to find the desired balance between using sophisticated and complex methods, and maintaining the interpretability and transparency of these methods may enhance the uptake of state-of-the-art methods for external controls from IPD-RWD.
12. Because there is no "one-size-fits-all" approach to externally controlled trials using IPD-RWD, there is a need for early and iterative scientific and statistical consultations by regulators and HTA organizations throughout the process.

HTA indicates health technology assessment; IPD-RWD, individual patient data real-world data; RCT, randomized controlled trial.

(ethically or practically) and with an a priori target trial emulation approach, externally controlled trials are sometimes acceptable. External controls can only be applied if all the underlying assumptions are met (which may substantially differ for anchored non–RWD-based and unanchored [non–]RWD-based approaches), an actual comparator is available (other than placebo or standard of care), an extreme benefit is expected, and the natural history of disease is known. However, these situations rarely occur at the same time.

Given that there seems to be no "one-size-fits-all" approach to applying external controls, this highlights why early dialogues and scientific advice among all stakeholders in the medicine lifecycle are critical.[23,24] Some guidance on the use of RWD in regulatory decision making exists but early multistakeholder dialogues with developers, regulators, HTA organizations, patients, and clinicians and subsequent (iterative, statistical) scientific consultations may be able to address some of the uncertainty that arises when considering the use of (IPD-)RWD as external control.[70-73] Both dialogues and advice provide opportunities to discuss the justification for an externally controlled trial, the acceptability of an external control for decision making, the most suitable and feasible trial design to answer the relevant questions, and the gaps that may remain. In addition, the evidence that needs to be generated after authorization or other risk management strategies may be discussed here.[74] The overview in this article may be used as a starting point for these discussions.

Strengths and Limitations

This article is the first to provide an overview of analytical methods for comparing uncontrolled trials with external controls from IPD-RWD and their role in regulatory and HTA decision making. The performed search included scientific literature on external controls using IPD-RWD only. Therefore, other methods that can also be applied in the context of IPD-RWD as external control may be available, particularly methods that are described in other contexts for unanchored comparisons. Our results can be used as an overview of the available options, but the use of (combinations of) methods should be considered case by case given that specific use cases were often not provided in the literature. Importantly, owing to the brief and inconsistent reporting of methods in EPARs and HTA reports, it was difficult to fully understand how the methods had been applied and to assess the potential remaining biases. Finally, we included EPARs until 2020. Given the increased submissions using uncontrolled trials with external controls, the most recent cases are not included. These cases may include more "state-of-the-art" methods in their submissions than are currently expressed in this article.

Conclusion

External controls using RWD discussed in regulatory and HTA reports often concern aggregated data and rarely make use of

state-of-the-art analytical methods described in literature and/or recommended in regulatory and HTA guidelines. The literature suggested a target trial emulation-like approach: a priori designed studies based on one interventional and one IPD-RWD arm, mimicking target RCTs to the extent feasible to minimize bias. We developed recommendations for regulatory and HTA organizations to improve the quality and acceptability of external controls.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.08.002.

Article and Author Information

Accepted for Publication: August 16, 2024

Published Online: xxxx

doi: https://doi.org/10.1016/j.jval.2024.08.002

Author Affiliations: Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Utrecht, The Netherlands (Hogervorst, Soman, Gardarsdottir, Goettsch, Bloem); Division Laboratory and Pharmacy, Department of Clinical Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands (Gardarsdottir); Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland (Gardarsdottir); National Health Care Institute (ZIN), Diemen, The Netherlands (Goettsch).

Correspondence: Lourens T. Bloem, PharmD, PhD, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Universiteitsweg 99, 3584 CG, Utrecht, The Netherlands. Email: l.t.bloem@uu.nl

Author Contributions: *Concept and design:* Hogervorst, Goettsch, Bloem
Acquisition of data: Hogervorst, Soman, Bloem
Analysis and interpretation of data: Hogervorst, Soman, Gardarsdottir, Bloem
Drafting of the manuscript: Hogervorst, Soman, Gardarsdottir
Critical revision of the paper: Hogervorst, Gardarsdottir, Goettsch, Bloem
Statistical analysis: Hogervorst
Obtaining funding: Goettsch, Bloem
Administrative, technical, or logistic support: Hogervorst, Gardarsdottir
Supervision: Hogervorst, Goettsch, Bloem

Funding/Support: This work was commissioned and funded by the GetReal Institute (Utrecht, The Netherlands).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, and analysis of the data; and preparation and review of the manuscript. The GetReal External Comparator Working Group was informed of the results twice during 1-hour meetings, where their practical perspectives on the relevance of the results were collected.

Acknowledgment: The authors extend their appreciation to Jasmin Schelhaas, MSc, and Karla Weber, MSc, for their diligent efforts in creating the databases that included the products assessed by the European Medicines Agency and European health technology assessment organizations based on uncontrolled trials. In addition, the authors are grateful to Dr Rick Vreman for his guidance and critical insights that were instrumental in the supervision of Karla throughout the data collection process. Moreover, Dr Sophie Bots greatly contributed to the discussions on the classification of the comparative methods. Finally, the authors sincerely thank Prof Dr Aukje Mantel-Teeuwisse for critically reviewing the manuscript.

REFERENCES

1. Eichler HG, Pignatti F, Schwarzer-Daum B, et al. Randomized controlled trials versus real world evidence: neither magic nor myth. *Clin Pharmacol Ther*. 2021;109(5):1212–1218.
2. EMA. Single-arm trials as pivotal evidence for the authorisation of medicines in EU. European Medicines Agency. https://www.ema.europa.eu/en/news/single-arm-trials-pivotal-evidence-authorisation-medicines-eu; Published 2023. Accessed May 31, 2023.
3. US Department of Health and Human Services, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Oncology Center of Excellence (OCE). Considerations for the design and conduct of externally controlled trials for drug and biological products. Food and Drug Administration (FDA). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products; Published 2023. Accessed February 25, 2023.
4. Sterne JA, Hernán MA, Reeves BC, et al. Robins-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
5. Jiang Y, Ni W. Performance of unanchored matching-adjusted indirect comparison (MAIC) for the evidence synthesis of single-arm trials with time-to-event outcomes. *BMC Med Res Methodol*. 2020;20(1):241.
6. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W, (on behalf of GetReal Work Package 1). What Is Real-World Data? A review of definitions based on literature and stakeholder interviews. *Value Health J Int Soc Pharmacoecon Outcomes Res*. 2017;20(7):858–865.
7. Phillippo DM, Ades AE, Dias S, et al. NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE report by the decision support unit. www.nicedsu.org.uk; Published 2016. Accessed February 25, 2023.
8. Jaksa A, Louder A, Maksymiuk C, et al. A comparison of seven oncology external control arm case studies: critiques from regulatory and health technology assessment agencies. *Value Health J Int Soc Pharmacoecon Outcomes Res*. 2022;25(12):1967–1976.
9. Patel D, Grimson F, Mihaylova E, et al. Use of external comparators for health technology assessment submissions based on single-arm trials. *Value Health*. 2021;24(8):1118–1125.
10. Rudrapatna VA, Butte AJ. Opportunities and challenges in using real-world data for health care. *J Clin Invest*. 2020;130(2):565–574.
11. Grimberg F, Asprion PM, Schneider B, Miho E, Babrak L, Habbabeh A. The real-world data challenges radar: a review on the challenges and risks regarding the use of real-world data. *Digit Biomark*. 2021;5(2):148–157.
12. Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N Engl J Med*. 2020;382(7):674–678.
13. Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*. 2005;162(3):279–289.
14. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;17(1):162.
15. Suissa S. Single-arm trials with historical controls: study designs to avoid time-related biases. *Epidemiology*. 2021;32(1):94–100.
16. Nicod E, Maynou L, Visintin E, Cairns J. Why do health technology assessment drug reimbursement recommendations differ between countries? A parallel convergent mixed methods study. *Health Econ Policy Law*. 2020;15(3):386–402.
17. Seo SK. External control arms: we're not on cruise control yet. *Clin Pharmacol Ther*. 2023;114(2):249–251.
18. Abbas S, Bender R, Gorris R, et al. Concepts for the generation of routine practice data and their analysis for the benefit assessment of drugs according to §35a social code book V (SGB V). Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. https://www.iqwig.de/download/a19-43_routine-practice-data-for-the-benefit-assessment-of-drugs_rapid-report_v1-0.pdf; Published 2019. Accessed February 25, 2023.
19. EUnetHTA 21. *Methodological Guideline on Direct and Indirect Comparisons V1.0*. EUnetHTA 21; 2022:1–33. https://www.eunethta.eu/wp-content/uploads/2022/08/EUnetHTA-21-Deliverable-D4.3.2-Methodological-Guideline-on-Direct-and-indirect-comparisons-V1.0.pdf. Accessed February 25, 2023.
20. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR task force on indirect treatment comparisons good research practices: Part 2. *Value Health*. 2011;14(4):429–437.
21. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg (versie 2016). Zorginstituut Nederland. https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg; Published 2016. Accessed February 25, 2023.
22. 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 Evid Based Med*. 2024;29(1):1–5.
23. Sola-Morales O, Curtis LH, Heidt J, et al. Effectively leveraging RWD for external controls: a systematic literature review of regulatory and HTA decisions. *Clin Pharmacol Ther*. 2023;114(2):325–355.
24. Curtis LH, Sola-Morales O, Heidt J, et al. Regulatory and HTA considerations for development of real-world data derived external controls. *Clin Pharmacol Ther*. 2023;114(2):303–315.

25. Regulation on health technology assessment. European Commission. https://health.ec.europa.eu/health-technology-assessment/regulation-health-technology-assessment_en. Accessed December 20, 2022.
26. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10(1):89.
27. Connected Papers. Find and explore academic papers. https://www.connectedpapers.com/. Accessed June 7, 2023.
28. Seeger JD, Davis KJ, Iannacone MR, et al. Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials. *Pharmacoepidemiol Drug Saf*. 2020;29(11):1382–1392.
29. Wang CY, Berlin JA, Gertz B, et al. Uncontrolled extensions of clinical trials and the use of external controls—scoping opportunities and methods. *Clin Pharmacol Ther*. 2022;111(1):187–199.
30. Abrahami D, Pradhan R, Yin H, Honig P, Baumfeld Andre E, Azoulay L. Use of real-world data to emulate a clinical trial and support regulatory decision making: assessing the impact of temporality, comparator choice, and method of adjustment. *Clin Pharmacol Ther*. 2021;109(2):452–461.
31. Phillippo DM, Dias S, Ades AE, Welton NJ. Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. *Stat Med*. 2020;39(30):4885–4911.
32. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
33. Schelhaas J, Mantel-Teeuwisse AK, Hoekman J, Hennik PB, Herberts C, Bloem LT [unpublished]. *Justification for the Acceptance of Uncontrolled Pivotal Trials to Support Marketing Authorisation of Medicines in Europe*; 1995-2020.
34. Public Policy Committee, International Society of Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practice (GPP). *Pharmacoepidemiol Drug Saf*. 2016;25(1):2–10.
35. Burger HU, Gerlinger C, Harbron C, et al. The use of external controls: to what extent can it currently be recommended? *Pharm Stat*. 2021;20(6):1002–1016.
36. Mishra-Kalyani PS, Amiri Kordestani L, Rivera DR, et al. External control arms in oncology: current use and future directions. *Ann Oncol*. 2022;33(4):376–383.
37. Gross AM. Using real world data to support regulatory approval of drugs in rare diseases: a review of opportunities, limitations & a case example. *Curr Probl Cancer*. 2021;45(4):100769.
38. Evans SR, Harris AD. Methods and issues in studies of CRE. *Virulence*. 2017;8(4):453–459.
39. Evans SR, Powers J. Evaluating anti-infective drugs in the resistant pathogen setting: can we use external controls? *Stat Commun Infect Dis*. 2017;9(1):20160003.
40. Franklin JM, Liaw KL, Iyasu S, Critchlow CW, Dreyer NA. Real-world evidence to support regulatory decision making: new or expanded medical product indications. *Pharmacoepidemiol Drug Saf*. 2021;30(6):685–693.
41. Franklin JM, Glynn RJ, Martin D, Schneeweiss S. Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther*. 2019;105(4):867–877.
42. Thorlund K, Dron L, Park JJH, Mills EJ. Synthetic and external controls in clinical trials – a primer for researchers. *Clin Epidemiol*. 2020;12:457–467.
43. Yap TA, Jacobs I, Baumfeld Andre E, Lee LJ, Beaupre D, Azoulay L. Application of real-world data to external control groups in oncology clinical trial drug development. *Front Oncol*. 2022;11:695936.
44. Gray CM, Grimson F, Layton D, Pocock S, Kim J. A framework for methodological choice and evidence assessment for studies using external comparators from real-world data. *Drug Saf*. 2020;43(7):623–633.
45. Lambert J, Lengliné E, Porcher R, Thiébaut R, Zohar S, Chevret S. Enriching single-arm clinical trials with external controls: possibilities and pitfalls. *Blood Adv*. 2023;7(19):5680–5690.
46. Li Y, Izem R. Novel clinical trial design and analytic methods to tackle challenges in therapeutic development in rare diseases. *Ann Transl Med*. 2022;10(18):1034.
47. Wu J, Wang C, Toh S, Pisa FE, Bauer L. Use of real-world evidence in regulatory decisions for rare diseases in the United States—current status and future directions. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1213–1218.
48. Schmidli H, Häring DA, Thomas M, Cassidy A, Weber S, Bretz F. Beyond randomized clinical trials: use of external controls. *Clin Pharmacol Ther*. 2020;107(4):806–816.
49. Franklin JM, Platt R, Dreyer NA, et al. When can nonrandomized studies support valid inference regarding effectiveness or safety of new medical treatments? *Clin Pharmacol Ther*. 2022;111(1):108–115.
50. Backenroth D. How to choose a time zero for patients in external control arms 2021;. 2021;20(4):783–792.
51. Ramsey SD, Adamson BJ, Wang X, et al. Using electronic health record data to identify comparator populations for comparative effectiveness research. *J Med Econ*. 2020;23(12):1618–1622.
52. Tan K, Bryan J, Segal B, et al. Emulating control arms for cancer clinical trials using external cohorts created from electronic health record-derived real-world data. *Clin Pharmacol Ther*. 2022;111(1):168–178.
53. Li Q, Lin J, Chi A, Davies S. Practical considerations of utilizing propensity score methods in clinical development using real-world and historical data. *Contemp Clin Trials*. 2020;97:106123.
54. Webster-Clark M, Jonsson Funk M, Stürmer T. Single-arm trials with external comparators and confounder misclassification: how adjustment can fail. *Med Care*. 2020;58(12):1116–1121.
55. Loiseau N, Trichelair P, He M, et al. External control arm analysis: an evaluation of propensity score approaches, G-computation, and doubly debiased machine learning. *BMC Med Res Methodol*. 2022;22(1):335.
56. Zhu J, Tang RS. A proper statistical inference framework to compare clinical trial and real-world progression-free survival data. *Stat Med*. 2022;41(29):5738–5752.
57. Rippin G, Ballarini N, Sanz H, Largent J, Quinten C, Pignatti F. A review of causal inference for external comparator arm studies. *Drug Saf*. 2022;45(8):815–837.
58. Popat S, Liu SV, Scheuer N, et al. Addressing challenges with real-world synthetic control arms to demonstrate the comparative effectiveness of pralsetinib in non-small cell lung cancer. *Nat Commun*. 2022;13(1):3500.
59. Oksen D, Prince P, Boutmy E, et al. Treatment effectiveness in a rare oncology indication: lessons from an external control cohort study. *Clin Transl Sci*. 2022;15(8):1990–1998.
60. Burcu M, Dreyer NA, Franklin JM, et al. Real-world evidence to support regulatory decision-making for medicines: considerations for external control arms. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1228–1235.
61. Jarow JP. Use of external controls in regulatory decision-making. *Clin Pharmacol Ther*. 2017;101(5):595–596.
62. Prasad V. Reliable, cheap, fast and few: what is the best study for assessing medical practices? Randomized controlled trials or synthetic control arms? *Eur J Clin Investig*. 2021;51(8):e13580.
63. Honig PK. Real-world evidence and the regulation of medicines. *Clin Pharmacol Ther*. 2021;109(5):1169–1172.
64. Davi R, Mahendraratnam N, Chatterjee A, Dawson CJ, Sherman R. Informing single-arm clinical trials with external controls. *Nat Rev Drug Discov*. 2020;19(12):821–822.
65. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758–764.
66. Criteria voor beoordeling therapeutische waarde. Zorginstituut Nederland. https://www.zorginstituutnederland.nl/publicaties/publicatie/2014/07/01/criteria-voor-beoordeling-therapeutische-waarde; Published 2014. Accessed June 10, 2023.
67. IQWiG. General methods. version 6.0. https://www.iqwig.de/methoden/general-methods_version-6-0.pdf; Published 2020. Accessed February 25, 2023.
68. Working group XIII real-world data and real-world evidence in regulatory decision making. Council for International Organizations of Medical Sciences. https://cioms.ch/working-groups/real-world-data-and-real-world-evidence-in-regulatory-decision-making/. Accessed June 13, 2023.
69. Guidance for reporting real-world evidence. CADTH. https://www.cadth.ca/guidance-reporting-real-world-evidence. Accessed June 13, 2023.
70. FDA. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug; Published 2023. Accessed September 1, 2023.
71. Guideline on registry-based studies-scientific guideline. European Medicines Agency. https://www.ema.europa.eu/en/guideline-registry-based-studies-scientific-guideline; Published 2021. Accessed September 1, 2023.
72. Hogervorst MA, Møllebæk M, Vreman RA, et al. Perspectives on how to build bridges between regulation, health technology assessment and clinical guideline development: a qualitative focus group study with European experts. *BMJ Open*. 2023;13(8):e072309.
73. Hogervorst MA, Vreman R, Heikkinen I, et al. Uncertainty management in regulatory and health technology assessment decision-making on drugs: guidance of the HTAi-DIA Working Group. *Int J Technol Assess Health Care*. 2023;39(1):e40.
74. Schünemann HJ, Reinap M, Piggott T, et al. The ecosystem of health decision making: from fragmentation to synergy. *Lancet Public Health*. 2022;7(4):e378–e390.