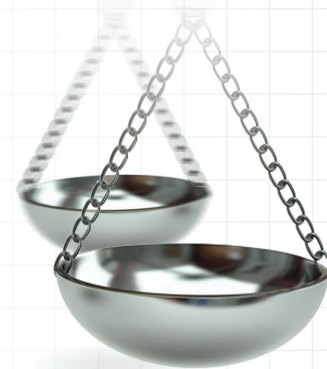




Use of transportability methods for real-world evidence generation: a review of current applications

Journal of **Comparative Effectiveness Research**

Natalie S Levy¹ , Patrick J Arena² , Thomas Jemielita³ , Shahrul Mt-Isa⁴ , Shane McElwee⁵ , David Lenis¹ , Ulka B Campbell^{1,6} , Ashley Jaksa² & Gleicy M Hair^{*,7}

¹Scientific Research & Strategy, Aetion, Inc., New York, NY 10001, USA

²Scientific Research & Strategy, Aetion, Inc., Boston, MA 02109, USA

³Biostatistics & Research Decision Sciences (BARDS), Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065, USA

⁴Biostatistics & Research Decision Sciences (BARDS), MSD Innovation & Development Hub GmbH, Merck Sharp & Dohme, Zürich, 8058, Switzerland

⁵Science & Delivery, Aetion, Inc., New York, NY 10001, USA

⁶Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY 10032, USA

⁷Center for Observational & Real-World Evidence (CORE), Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ 07065, USA

*Author for correspondence: Tel.: +1 917 936 6804; gleicy.hair@merck.com

Aim: To evaluate how transportability methods are currently used for real-world evidence (RWE) generation to inform good practices and support adoption and acceptance of these methods in the RWE context. **Methods:** We conducted a targeted literature review to identify studies that transported an effect estimate of the clinical effectiveness or safety of a biomedical exposure to a target real-world population. Records were identified from PubMed-indexed articles published any time before 25 July 2023 (inclusive). Two reviewers screened abstracts/titles and reviewed the full text of candidate studies to identify the final set of articles. Data on the therapeutic area, exposure(s), outcome(s), original and target populations and details of the transportability analysis (e.g., analytic method used, estimate transported, stated assumptions) were abstracted from each article. **Results:** Of 458 unique records identified, six were retained in the final review. Articles were published during 2021–2023, focused on the US/Canada context, and covered a range of therapeutic areas. Four studies transported an RCT effect estimate, while two transported effect estimates derived from real-world data. Almost all articles used weighting methods to transport estimates. Two studies discussed all transportability assumptions, and one evaluated the likelihood of meeting all assumptions and the impact of potential violations. **Conclusion:** The use of transportability methods for RWE generation is an emerging and promising area of research to address evidence gaps in settings with limited data and infrastructure. More transparent and rigorous reporting of methods, assumptions and limitations may increase the use and acceptability of transportability for producing robust evidence on treatment effectiveness and safety.

Plain language summary: What is this article about?: In this article, we investigated whether and how statistical methods known as ‘transportability methods’ have been applied in published studies using data collected during routine healthcare, known as real-world evidence (RWE) studies. Transportability methods use a result based on data from one population to estimate the result for another population by adjusting for relevant differences in demographic, clinical, and/or other factors between the two populations. These methods may help decision-makers evaluate whether evidence from another location can inform assessments of the safety and effectiveness of medical products in their jurisdiction. We conducted a targeted review of the published literature to understand if and how transportability methods have been applied to RWE studies.

What were the results?: After reviewing 458 potential studies identified in our literature search, we found that only six used transportability methods to generate an estimate of the clinical effectiveness or safety of a biomedical exposure for a target real-world population. These studies were all published during 2021–

2023 using data from the US or Canada and used similar statistical methods. Two studies discussed all the statistical assumptions and limitations of transportability methods.

What do the results mean?: Transportability methods are just beginning to be used for RWE generation but may help fill evidence gaps in places or situations where relevant data are not available. Additionally, clear and thorough reporting of assumptions and limitations may facilitate the use and acceptability of transportability methods for producing robust RWE on treatment effectiveness and safety.

First draft submitted: 15 April 2024; Accepted for publication: 6 September 2024; Published online: 4 October 2024

Keywords: real-world data • real-world evidence • transportability

Regulators and health technology assessment (HTA) bodies rely on high-quality evidence to evaluate the safety and efficacy of biomedical products. HTA bodies additionally consider the economic, organizational and social contexts in which health technologies are used to determine their value and inform coverage and reimbursement decisions [1,2]. Ideally, regulatory and HTA decision-making would be based on robust data reflective of local patient populations and clinical care but, in practice, such data are not always readily available [3,4].

While randomized clinical trials (RCTs) remain the standard for establishing the efficacy of biomedical interventions, they are often conducted in highly selected populations and in a limited number of jurisdictions due to the significant financial resources, sample sizes and infrastructure required to support large-scale trials. Additionally, it is unethical to conduct more RCTs than necessary to establish or refute the efficacy of an intervention on the grounds that patients are put at unnecessary risk by redundant research [5,6]. Regulatory and HTA bodies are therefore increasingly recognizing the value of real-world evidence (RWE) studies – studies that generate evidence on the risks and benefits of biomedical interventions using data collected in the course of routine healthcare (real-world data, RWD) – for informing assessments of clinical effectiveness and safety [7]. However, the robustness and relevance of RWE rely on not only the use of proper epidemiologic and statistical methods (e.g., adjustment for confounders when assessing relative treatment effects) but also high-quality, fit-for-purpose data that may not be available in all jurisdictions [4,8–10]. For example, there are many accessible, high-quality RWD sources in the US whereas data availability is more limited in the European Union, a barrier currently being addressed through initiatives like the Data Analysis and Real World Interrogation Network (DARWIN EU) [11].

As a result, investigators may employ evidence generated in another jurisdiction (country or region) to demonstrate clinical effectiveness and safety to local regulators and HTA bodies [12–16]. The validity of using RWE from one jurisdiction to inform decision-making in another depends on the ‘external validity’ of the RWE. External validity broadly references the extent to which the results of a given study can be applied outside the original study context, such as to another target population of interest. Within external validity, ‘generalizability’ concerns whether study findings can be applied to a target population of which the study population is a subsample. ‘Transportability’, on the other hand, refers to the validity of extending study findings to a target population when there is minimal or no overlap between the study and target populations [17–20]. While the terms transportability and transferability are often used interchangeably in the healthcare research literature, transferability refers to a qualitative evaluation of whether data or findings from one setting may be informative in another context whereas transportability involves formal quantitative methods for applying an effect estimate across populations [15].

Several recent reviews have summarized available transportability methods and their assumptions [17,20–23]. Transportability methods generally fall into three broad categories [17,20]: weighting methods [19,21,23], outcome regression methods [21,24] and doubly-robust methods that combine the two [21,25,26] (Box 1). These approaches all aim to create conditional average exchangeability between the study and target populations by aligning the distribution of effect modifiers in both populations, as differences in these distributions limit the external validity of study findings [17,22,27]. All approaches also require similar identifiability assumptions. In Box 2, we summarize the assumptions required to identify the target population average treatment effect (ATE) – the absolute or relative difference in the average outcome if the entire target population is simultaneously exposed to two different treatments. In most cases, the choice of estimand will affect the conditions required for internal validity while leaving the other key assumptions unchanged.

Box 1. Classes of transportability methods.

Transportability methods are used to extend an estimated effect from one source population to another target population of which it is not a subset. Transportability methods can be applied to many estimands; here we focus on the estimated population average treatment effect (ATE) [17]. The ATE is a comparison of the average outcome if everyone in the population were exposed to a particular treatment with the average outcome if everyone were unexposed to that treatment.

Weighting methods

- Weighting methods transport the ATE from a source to a target population by using inverse odds of sampling weights, where ‘sampling’ refers to the probability of being included in the original study population [19]. Conditional on their effect modifier values, individuals in the study population are weighted up or down to reproduce the distribution of effect modifiers in the target population [17,19,23]. This reweighted population is then used to estimate the ATE in the target population.

Outcome regression methods

- Outcome regression methods use data from the source study to generate models of the outcome conditional on effect modifiers for each treatment group. These models are then used to predict individual outcomes under different exposures (potential outcomes) in the target population. The ATE is then estimated by comparing the average potential outcomes in the target population under each exposure condition [17,20,21].

Combined methods

- Methods that combine both weighting and outcome regression use these two approaches together to build an estimator of the mean potential outcome in the target population under different treatments that can then be used to estimate the ATE [17,20,21]. Many of these combined methods are considered ‘doubly-robust’ because they are asymptotically unbiased (i.e., they are unbiased as the sample size tends to infinity) if either the weighting model or the outcome model is correctly specified [17].

Transportability methods were originally developed in the context of RCTs due to the high internal validity conferred by randomization and rigorous treatment procedures and follow-up, but they can also be applied to findings from RWE studies [7]. Stakeholders acknowledge that RWD from other countries may be used in regulatory and HTA submissions and broadly discuss the need to justify and assess the appropriateness of the imported RWE; however, specific guidance and best practices for performing such evaluations are needed [9,15,31–35]. Transportability methods may provide a rigorous and transparent approach for both assessing the relevance of external RWE and producing local estimates of clinical effectiveness and safety [16]. A first step toward evaluating the potential utility of transportability methods in a regulatory and/or HTA context is to understand if and how these methods are applied, particularly with respect to RWE. To that end, in this paper we conducted a targeted literature review to evaluate how transportability methods are currently being used for RWE generation and to begin developing good practices that may facilitate the adoption and acceptance of these methods in the context of RWE.

Methods**Targeted literature review scope**

As this targeted literature review focused on transportability in the context of RWE studies, we limited our review to RWE transportability investigations that contained an application of the method in the context of studying the clinical effectiveness or safety of a biomedical intervention. We did not pre-specify any explicit populations, exposures, comparators, or outcomes of interest as our primary goal was to identify any associated RWE transportability methods applications. The review was conducted and reported following PRISMA guidelines (Supplemental Table 1).

Search strategy & exclusion criteria

We searched Medical Literature Analysis and Retrieval System Online (MEDLINE)/PubMed for articles published any time up to 25 July 2023 (inclusive). Our search strategy was tailored iteratively by modifying the combination of subject headings/controlled vocabulary and keyword fields (e.g., title and abstract) used to identify all potentially relevant publications, while returning a manageable total number of results for title and abstract screening. To account for potential instances where ‘transferability’ was used in place of ‘transportability’, we included both terms in our search string. Pilot searches returned a large number of genomic (e.g., studies of gene transfer) and transportation studies which were not relevant to this review. Consequently, the final search strategy, as agreed to by all authors, used the following terms: (transportability OR ‘transferability’) AND ((real world) OR (real-world))

Box 2. Identifiability assumptions for transportability methods.

Transportability methods rely on several key identifiability assumptions that must be met to produce valid results. Details of these assumptions are described elsewhere [17,21,22]. Similar assumptions apply to all causal estimands [17]. In most cases, the choice of estimand will affect the conditions required for internal validity while leaving the other key assumptions unchanged. Here, we focus on the necessary assumptions for a transported estimate to equal the ATE in the target population. The ATE is a comparison of the average outcome if everyone in the population were exposed to a particular treatment with the average outcome if everyone were unexposed to that treatment.

Internal validity of the original study

- The estimated effect equals the true ATE in the source population of the original study. This assumption requires conditional exchangeability of the exposed and unexposed in the study population (i.e., no unmeasured confounding, measurement error and selection bias), consistency (i.e., the observed outcome equals the potential outcome under received treatment), positivity of treatment (i.e., the conditional probability of receiving a given treatment is larger than 0 and less than 1 in any patient subgroup as defined by combinations of confounders), no interference between individuals (i.e., treatment received by one individual does not impact the outcome experienced by another individual), treatment version irrelevance (i.e., there is only one version of the treatment or there are multiple versions with the same effect on the outcome) and correct model specification [17,28]. In general, RCTs are assumed to have internal validity due to randomization, although this assumption can be compromised if there are missing data or chance imbalances across study arms.

Conditional exchangeability over selection (S-admissibility)

- Individuals in the study population and in the target population with the same baseline characteristics have the same potential outcomes under treatment and no treatment and are therefore exchangeable. In practice, a weaker assumption of conditional average exchangeability is most often invoked (i.e., the same effect of treatment would be observed if individuals with the same baseline characteristics were moved from the study population to the target population). To be met, this assumption requires that all effect modifiers that have different distributions in the study and target populations are identified, measured and accounted for by the transportability method [17,29].

Positivity of selection

- There is a non-zero probability of being in the original study population in every stratum of effect modifiers needed to ensure conditional exchangeability. For this assumption to be met, every stratum of effect modifiers that differs between the study and target populations must be represented in the study population [30]. This assumption is distinct from the positivity of treatment assumption that is required for internal validity.

Stable unit treatment value assumption (SUTVA) for selection

- There is no interference between units and there is treatment version irrelevance between the study and target populations. This assumption requires that the treatment received by one individual does not impact the outcome experienced by another individual (i.e., there is no interference between individuals) and that the treatment or intervention is applied in the exact same way in the study and target populations [17].

OR (observa*)) NOT ((genom*[Title/Abstract]) OR (transportation[Title/Abstract])). No language restrictions were applied. Google Sheets was used to save and organize references and provide an interface for screening.

Data extraction & synthesis

After searching MEDLINE/PubMed and importing the relevant results, two authors (NSL and PJA) performed title and abstract screening to select articles that discussed RWE applications of transportability methods in the context of a biomedical intervention; after this initial screening round, full text screening was performed by the same two authors to select for studies that included a specific transportability RWE case study. Discrepancies were resolved through consensus discussion between the two reviewers or inclusion of a third reviewer (SM) to serve as tiebreaker if agreement could not be reached. Simulation studies as well as purely (bio)statistical manuscripts were excluded because these article types did not inform our understanding of how transportability methods are being applied for RWE generation. Likewise, we did not include articles that focused exclusively on the cost-effectiveness of medical products due to our pre-specified focus on studies of the clinical effectiveness and safety of biomedical interventions. Cited references from included studies were reviewed, but no additional sources were identified.

Data elements were extracted by NSL and PJA from the selected studies using a standardized template; discrepancies were resolved through consensus discussion between the two reviewers or inclusion of a third reviewer (SM) to serve as tiebreaker if necessary. The following items were extracted: first author, journal name, publication year, therapeutic area, exposure(s), outcome(s), transportability details (i.e., method used, estimate being transported,

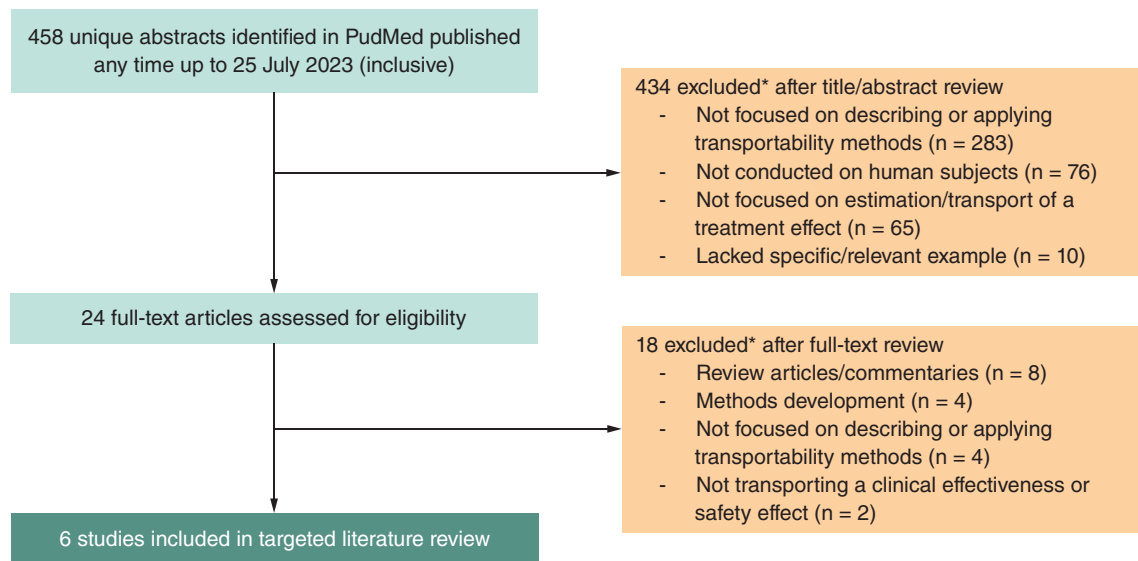


Figure 1. Study selection flow chart. *It was possible for an abstract/article to have more than one reason for exclusion; however, only the primary reason is shown here.

reason for transporting and stated assumptions), original study details (i.e., study design, data source and study estimate), target population details (i.e., study design, data source and study estimate), article conclusions, strengths and limitations (Supplemental Table 2).

Due to the nature of our research question, identified studies were not assessed for their quality or validity via traditional risk of bias tools (such as GRADE [36] or ROBINS-I [37]); however, we did assess their completeness and clarity from a transportability perspective (e.g., the transparency of their presentation and application of a transportability method). Similarly, publication bias was also not assessed. However, a narrative synthesis of the details from the included studies was performed, and a table of key data elements and assumptions from the included studies was produced to summarize the associated information. Lastly, no quantitative syntheses were performed.

Ethical approval & other considerations

This targeted literature review did not require institutional review board approval as the data that informed the research were publicly available and collected from an existing online database (i.e., MEDLINE/PubMed). Moreover, this research did not involve any human subjects, so informed consent was not required.

Results

Targeted search results

In total, we identified six articles that addressed the aims of this targeted literature review. Of 458 unique records initially identified, 434 were excluded after title and abstract review (Figure 1). Most (n = 283) were excluded because they were not explicitly focused on describing or applying transportability methods; these records instead discussed topics common in other academic disciplines, such as materials science (e.g., cadmium transfer from soils), psychology (e.g., transferable telepsychiatry models) and education (e.g., transfer of knowledge from the classroom to practice) that were not applicable for our purposes. Other records were excluded because they studied non-human subjects (n = 76) or were not focused on the estimation or transport of a treatment effect (n = 65); examples of the latter category included qualitative studies and investigations describing the transferability of prediction/machine learning algorithms. A further 18 articles were excluded after full-text screen; these included eight review articles or commentaries without an example or application [7,20,21,38–42] and four methods development papers where applications were only illustrative [25,43–45] (Figure 1). Thus, six case studies applying transportability methods to RWE were ultimately included.

Table 1. Overview of key study characteristics from the six identified case studies.

First author	Study year	Therapeutic area	Primary reason for transporting	Data type		Transported estimate	Transportability method	Ref.
				Original	Target			
Basu	2023	Mental health	Limited RWD available	US RCT data	US RWD (claims)	Mean difference	Weighting	[48]
Cook	2023	Substance use	Unrepresentative RCTs	US RCT data	US RWD (multiple) [†]	Hazard ratio	Weighting	[47]
Inoue	2021	Oncology	Demonstrate utility/feasibility of transportability	US RCT data	US RCT data [‡]	Hazard ratio	Weighting	[30]
Mollan	2021	Infectious disease	Discrepancies between RCTs & observational data	US RCT data	US RWD (EMR)	Hazard ratio & incidence rate difference	Weighting	[49]
Montez-Rath	2022	Rheumatology/ inflammation	Limited RWD available	US RWD (claims)	US RWD (claims)	Incidence rate	Weighting	[46]
Ramagopalan	2022	Oncology	Demonstrate utility/feasibility of transportability	US RWD (EMR)	Canada RWD (EMR, registry)	Overall survival	Outcome regression	[50]

[†] This study used survey data, substance use treatment admissions/discharges data and research consortium data comprising both survey and laboratory data.
[‡] This study did not use RWD but rather used RCT data to demonstrate how transportability methods can extend estimation beyond RCT trial participants to external populations of interest that more closely resemble real-world populations.
EMR: Electronic medical record; RCT: Randomized clinical trial; RWD: Real-world data.

Case study characteristics

The main features of the six included case studies are summarized in Table 1. All identified case studies were published between 2021 and 2023 and these investigations were conducted across diverse therapeutic areas: mental health (n = 1), substance use (n = 1), oncology (n = 2), infectious disease (n = 1) and rheumatology/inflammation (n = 1). Four studies were performed to address gaps in available RCT data and/or RWD. For example, Montez-Rath *et al.* [46] investigated the safety of Janus kinase inhibitors (JAK-Is) among adults with atopic dermatitis (AD) using RWD from a rheumatoid arthritis (RA) cohort due to limited availability of safety data on AD patients treated with JAK-Is. Likewise, Cook *et al.* [47] conducted their transportability study due to concerns that certain patient populations (e.g., people living in rural areas) were underrepresented in RCTs of treatment for substance use disorders. The remaining two transportability studies were performed to demonstrate the utility and feasibility of transportability methods for RWE generation.

Three case studies transported estimates from RCT populations to real-world target populations, while two case studies used transportability methods with both original and target populations constructed from RWD sources. One case study performed by Inoue *et al.* [30] used RCT data exclusively to illustrate how “transportability methods extend estimation of RCTs’ utility beyond trial participants, to external populations of interest, including those that more closely mirror real-world populations.” In five of the case studies, both the original and target populations were US-based. In the remaining case study by Ramagopalan *et al.* [50], an original estimate derived from US RWD was transported to a Canadian target population. Administrative claims and electronic medical record (EMR) data were the most common type of RWD used across the case studies.

Transportability methods used in case studies

Five of the case studies used a weighting approach as their transportability method and one used outcome regression modeling (Box 1). These approaches all require access to individual-level data, though only Mollan *et al.* [49] and Inoue *et al.* [30] explicitly indicated that this type of data was used. We also observed that authors described their approaches differently, despite applying equivalent methods. For instance, Basu *et al.* [48] framed their weighting method as a calibration approach in which inverse probability weights were used to ‘calibrate’ the distribution of patient characteristics across the original and target populations. The other four case studies that utilized a weighting methodology first estimated associations between the exposure and the outcome in the original population and then transported the associated estimate (e.g., the hazard ratio, incidence rate, or incidence rate difference) to the target population via a weighting procedure. For example, Mollan *et al.* [49] used weights to transport the hazard ratio and incidence rate difference calculated from four AIDS Clinical Trial Group RCTs to an observational cohort of adults living with HIV and receiving care at eight academic medical center sites. Of note, these four case studies used varying terminology to identify their selected transportability method despite almost all citing Westreich *et al.*’s 2017 paper on inverse odds of sampling weights for transportability [19]. Specifically, Montez-Rath *et al.* [46] and

Table 2. Discussion and evaluation of transportability assumptions in the six identified case studies.

Transportability assumption [†]	Case study					
	Basu (2023) [48]	Cook (2023) [47]	Inoue (2021) [30]	Mollan (2021) [49]	Montez-Rath (2022) [46]	Ramagopalan (2022) [50]
Internal validity of the original study: The estimated effect equals the true estimand in the source population of the original study.		✓	✓✓	✓✓	✓✓	✓
Conditional exchangeability over selection (S-admissibility): Exchangeability of individuals in the study population and in the target population with the same baseline characteristics.	✓✓	✓✓	✓✓	✓✓	✓✓	
Positivity of selection: There is a non-zero probability of being in the original study population in every stratum of effect modifiers needed to ensure conditional exchangeability of individuals in the study and target populations.		✓✓	✓✓	✓		
Stable unit treatment value assumption for selection: There is no interference between units and there is treatment version irrelevance between the study and target populations.			✓✓	✓		
[†] Similar assumptions apply to all causal estimands. In most cases, the choice of estimand will affect the conditions required for internal validity while leaving the other key assumptions unchanged. ✓ Indicates that the assumption was mentioned in the text. ✓✓ Indicates that the assumption was mentioned and the authors either discussed the likelihood of the assumption being met or conducted sensitivity analyses to investigate the potential impact of violations.						

Cook *et al.* [47] used the phrase ‘inverse probability of selection weighting’ to describe their weighting method, Inoue *et al.* [30] referred to their approach as ‘inverse-odds weighting’, and Mollan *et al.* [49] called their method an applied use of ‘inverse odds of participation weights’.

The remaining case study by Ramagopalan *et al.* [50] was the only one to employ an outcome regression approach as their transportability method. The authors fit pooled logistic regression models for overall survival (noting that these models produce coefficients equivalent to those from a Cox regression analysis under certain assumptions) as a function of covariates in their original sample from the US; they then standardized the models to the covariate distributions in the target (Canadian) population to obtain transported marginal survival probabilities. Ramagopalan *et al.* [50] was also the only identified case study that provided a justification for their choice of transportability method: the inability to pool data sets from the US and Canada, which prevented the use of a weighting approach.

Evaluation of transportability assumptions

As mentioned previously and summarized in Box 2, transportability methods aim to create exchangeability over selection between two populations and rely on several other assumptions to produce valid estimates. Two of the six articles discussed all four transportability assumptions (Table 2). Inoue *et al.* [30] included a table of key transportability assumptions with illustrative examples in the context of their research question and evaluated the likelihood of each assumption being met in their study and the potential impact of violations on their results. Mollan *et al.* [49] listed all required assumptions when presenting their methods, evaluated the internal validity of the estimates being transported and discussed challenges to achieving conditional exchangeability between the original and target populations due to missing data and differences in measurement. Of the remaining studies, two evaluated a subset of the identifiability assumptions [46,47], one evaluated the conditional exchangeability of selection assumption [48] and one discussed but did not evaluate the internal validity of the original study [50]. Overall, the stable unit treatment value assumption (SUTVA) for selection (i.e., the assumption that there is no interference between units and there is treatment version irrelevance between the study and target populations) [17] was the least discussed and evaluated across the identified case studies.

Study authors also mentioned and assessed other potential threats to validity. For example, four articles used multiple imputation by chained equations to account for potential bias due to missing data [46,47,49,50]. These articles provided varying levels of detail about their imputation approach (e.g., number of imputed data sets, imputation method for continuous vs categorical variables, etc.). Lastly, Ramagopalan *et al.* [50] performed quantitative bias analyses to assess the potential impact of measurement error (i.e., through a tipping point analysis where they imputed values for mismeasured metastases and comorbidities) and treatment pattern differences between the original and target populations (i.e., through estimation of marginal risk under two hypothetical dynamic treatment regimens via G-computation).

Discussion

This targeted literature review investigated how transportability methods are currently being applied for RWE generation. We identified six case studies meeting our inclusion criteria. These applications spanned a wide range of therapeutic areas, exposures and outcomes, likely reflecting a growing acceptance of RWE by healthcare researchers for assessing treatment effectiveness and safety more generally. However, these case studies were all published in the past few years using data from the US or Canada, and none of the studies were explicitly conducted in a regulatory or HTA context. This finding suggests that applications of transportability for regulatory and HTA decision-making are limited and that the use of transportability methods for RWE generation is an emerging field.

Four of the six case studies used an RCT effect estimate as the original study measure in their transportability analyses (Table 1). This finding was unsurprising given that transportability methods were originally designed to address the limited external validity of RCTs resulting from stringent inclusion criteria. Furthermore, the use of randomization and tightly-controlled study procedures and follow-up increase confidence in the internal validity of RCTs, a key assumption for transported estimates to be valid (Box 2). The two articles that transported estimates using only RWD focused on descriptive rather than comparative estimates, for which there is no assumption of internal validity. More specifically, Montez-Rath *et al.* [46] transported adverse event incidence rates for an RA population to an AD population and Ramagopalan *et al.* [50] transported overall survival estimates for US patients with advanced non-small-cell lung cancer to a Canadian patient population – an instance that should be highlighted as the only example of cross-country transportability that we identified. These examples potentially highlight gaps in the availability of high-quality RWD in certain patient populations and jurisdictions and represent promising potential applications of transportability for RWE generation.

Almost all the included articles used weighting as their primary transportability method. This trend likely reflects the fact that weighting is an intuitive and less computationally-intensive approach that parallels widely-used methods for confounding control in epidemiology and health outcomes research, such as inverse probability weighting. However, most authors did not provide a justification for their selected method or discuss alternative approaches. Relatedly, only two studies explicitly discussed all the assumptions required to generate valid transported estimates, and few took additional steps to evaluate the likelihood of meeting key assumptions or the impact of potential violations.

Lessons learned

Overall, our review suggested that the use of transportability methods for RWE generation is an emerging area of research. The identified case studies suggest that these methods may have promising applications in regulatory and/or HTA contexts, particularly in jurisdictions where RCTs are not conducted or where high-quality and fit-for-purpose RWD are limited. Our review also highlighted some areas for improvement to increase the acceptability of transportability methods in these settings. These suggestions build on considerations proposed by Turner *et al.* for incorporating transportability into HTA applications [16]. First, we found that it was sometimes challenging to understand the specific analytical approaches used in the reviewed articles. Thorough and transparent descriptions of and justifications for the transportability methods used would allow readers to better understand their application and implementation. This includes providing details on the granularity of data sources used, as there are fewer options for transporting results when only summary- rather than individual-level data are available [17].

Second, we observed a lack of consistent terminology when referring to transportability methods, with authors referring to equivalent weighting methods as ‘inverse probability of selection weighting’, [46,47] ‘inverse-odds weighting’, [30] and ‘inverse odds of participation weights’. [49] As previously discussed by Westreich *et al.* [19], use of precise and consistent terminology is important for clear communication between researchers and decision-makers, particularly given the lack of widespread familiarity with and use of transportability methods in the realm of RWE. We therefore recommend that future transportability applications include thorough and transparent descriptions of the methods employed, model statements and formulas as applicable and justification for the selected approach (e.g., the rationale for using inverse odds) to enhance understanding of these methods and interpretation of findings.

Finally, as discussed previously, we noted that only a few articles discussed the assumptions needed to generate valid transported estimates. We believe that explicit discussion of transportability assumptions, evaluation of their likelihood of being met and assessment of the potential impact of any violations are critical to the interpretation and acceptability of transported estimates. Indeed, these recommendations align with recent updates to the National Institute for Health and Care Excellence’s RWE framework, in which the agency recommends sensitivity analyses to explore potential violations of transportability study assumptions [9]. In Table 3, we present a proposed template

Table 3. Template for evaluation of key transportability assumptions.

Transportability assumption [†]	Study operationalization: describe how each assumption applies in the specific context of the study being conducted to transport an effect.	Assessment of whether assumption is likely met: discuss why or why not each assumption is likely to be met given the specific study context.	Potential impact of violations on transported estimate: discuss the potential impact of violations of each assumption on the transported estimate in terms of direction and magnitude. If a violation is not expected, explain why not.
Internal validity of the original study: The estimated effect equals the true estimand in the source population of the original study.	<i>"Exposed and unexposed individuals in the original study are exchangeable on all causes of the outcome [conditional on baseline confounders]."</i>	<i>"The original study likely has [high/low] internal validity due to [use of randomization, minimal/differential loss to follow-up, measurement of exposure/outcome, etc.]"</i>	<i>"Due to the use of randomization in the original study, no violations are expected on average." "The inability to control for [confounder] in the original study may result in a [slight/significant] [positive/negative] bias in the transported effect estimate."</i>
Conditional exchangeability over selection (S-admissibility): Exchangeability of individuals in the study population and in the target population with the same baseline characteristics.	<i>"The distribution of all relevant effect modifiers is the same or can be made the same in the study and target populations."</i>	<i>"Conditional exchangeability over selection [is/is not] likely given [similar/dissimilar distributions of effect modifiers after applying method]."</i>	<i>[Discussion of the potential for and impact of unmeasured or unknown effect modifiers, model misspecification, limitations of transportability method, results of sensitivity analyses, etc.]</i>
Positivity of selection: There is a non-zero probability of being in the original study population in every stratum of effect modifiers needed to ensure conditional exchangeability of individuals in the study and target populations.	<i>"Every stratum of relevant effect modifiers is populated with exposed and unexposed individuals in the original study population."</i>	<i>"Positivity of selection is likely/unlikely to be met due to [availability or lack of observations in all strata of relevant effect modifiers]."</i>	<i>[Discussion of whether any known effect modifiers were unobserved or had zero counts for some strata in the original study population or if there was substantial missing data for the target population]</i>
Stable unit treatment value assumption for selection: There is no interference between units and there is treatment version irrelevance between the study and target populations.	<i>"The [outcome] of one individual is not influenced by the exposure status of another and the exposure is measured and implemented in the same way in both populations."</i>	<i>"[No] interference between units is expected. Treatment version irrelevance between the two populations [is/is not] likely due to [specificity of the exposure definition, objectivity/subjectivity of measurement, etc.]"</i>	<i>[Discussion of potential differences in the definition or implementation of the exposure across the two populations, e.g., a drug administered according to a protocol in a clinical trial vs self-administration in routine real-world practice]</i>
[†] Similar assumptions apply to all causal estimands. In most cases, the choice of estimand will affect the conditions required for internal validity while leaving the other key assumptions unchanged. Adapted from the work of Inoue <i>et al.</i> [30].			

and example language informed by the approach of Inoue *et al.* to aid in the evaluation and presentation of key transportability assumptions [30]. For additional guidance on conducting transportability studies, and particularly for interpreting findings, we refer researchers to the instructive workflow developed by Ling *et al.* [20].

Strengths & limitations

To our knowledge, this is the first targeted literature review to assess the current use of transportability methods in the context of RWE investigations. Moreover, to capture relevant articles that are representative of this area of research, we created a tailored and iteratively refined search strategy and placed no limitations on the time frame for publication. However, articles were only retrieved from one database (i.e., MEDLINE/PubMed). This choice was made to allow us to focus on biomedical applications of transportability methods, but exclusion of literature from other disciplines may have resulted in some relevant case studies being overlooked. We also encountered some challenges when refining our search strategy. Our research team initially considered a broader search strategy – using transport instead of transportability in our search strategy – but this returned an exceedingly large number of studies to screen ($n = 35,837$). Using more inclusive forms of the term ‘transportability’ (e.g., ‘transport’, ‘transporting’, or ‘transported’) further increased the number of articles returned ($n > 50,000$) and a preliminary review suggested that fewer than 2% of the additional studies retrieved would be potentially relevant to our study based on title and abstract screening. Additionally, there is no specific Medical Subject Headings term to identify articles that employ transportability methods in PubMed. While we tried to maintain as broad a search as possible, we did have to place some restrictions on our search terms to return a manageable number of articles for screening, which likely resulted in the exclusion of some potentially pertinent articles. While we believe that our search strategy successfully targeted representative articles, our results should not be interpreted as an exhaustive review of the applied transportability literature. Our focus on RWE investigations of clinical effectiveness and safety likely also excluded transportability applications from other contexts (e.g., behavioral interventions) and investigations that focused solely on RCTs.

Finally, due to the nature of our question, we did not perform a validated risk of bias assessment for individual studies and did not assess potential publication bias.

Conclusion

The results of this targeted literature review suggest that the application of transportability methods for RWE generation is in its infancy. Nevertheless, the identified case studies demonstrate the potential utility of using transportability methods to address evidence gaps that may limit the assessment of treatment effectiveness and safety in jurisdictions with fewer resources and data sources or among underrepresented populations. The lessons learned from this review may inform the development of recommended practices for the design and reporting of RWE transportability studies, a necessary first step to facilitate their use and acceptability for addressing gaps in jurisdiction-specific evidence needed for regulatory and HTA decision-making.

Summary points

- Transportability methods were developed to overcome the limited external validity of randomized controlled trials that arises from stringent inclusion criteria, but they can also be applied to other types of studies and data sources, including real-world data.
- Transportability methods extend effect estimates from an original study population to a broader, non-overlapping target population by adjusting for differences in the distribution of effect modifiers between the two populations.
- We found that the use of transportability methods for real-world evidence generation is a nascent area of research.
- The six studies that met our inclusion criteria were all published during 2021–2023, used data from the US or Canada and primarily applied weighting methods to produce transported estimates.
- Few articles explicitly discussed and evaluated the assumptions needed to generate valid transported estimates.
- We propose the use of a structured template for presenting transportability methods, including explicit discussion of the required assumptions, their likelihood of being met in a real-world context and the potential impact of violations.
- Developing good practices for the design and reporting of real-world evidence transportability studies may help increase their acceptability for addressing evidence gaps in regulatory and health technology assessment contexts.
- The use of transportability methods may be particularly valuable for jurisdictions where patient populations are too small or local real-world data sources are limited and in rapidly-evolving public health scenarios where timely evidence generation is necessary.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: <https://bpl-prod.literatumonline.com/doi/10.57264/cer-2024-0064>

Author contributions

NS Levy, PJ Arena, A Jaksa, GM Hair, T Jemielita, S Mt-Isa, D Lenis and UB Campbell were responsible for study conception and design. NS Levy, PJ Arena and S McElwee were responsible for acquisition, analysis and interpretation of data. NS Levy and PJ Arena were responsible for drafting the manuscript and A Jaksa, GM Hair, T Jemielita, S Mt-Isa, D Lenis, S McElwee and UB Campbell were responsible for providing critical revisions and reviewing the final manuscript.

Acknowledgments

The authors would like to thank Victoria Faur for ensuring this study was completed in a timely and efficient manner.

Financial disclosure

This work was supported by Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc., NJ, USA. The authors have received no other financial and/or material support for this research or the creation of this work apart from that disclosed. NS Levy, PJ Arena, A Jaksa, S McElwee, D Lenis and UB Campbell are employees of and/or have ownership stake in Aetion, Inc., which works in collaboration with several pharmaceutical companies, government organizations and payers in healthcare. GM Hair, T Jemielita and S Mt-Isa are employees of and/or have ownership stake in Merck and Co., Inc. 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.

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. O'Rourke B, Oortwijn W, Schuller T. International Joint Task Group. The new definition of health technology assessment: a milestone in international collaboration. *Int. J. Technol. Assess Health Care* 36(3), 187–190 (2020).
2. Goodman C. *HTA 101 - Introduction to Health Technology Assessment*. United States National Library of Medicine, MD, USA (2014).
3. O'Donnell JC, Pham SV, Pashos CL, Miller DW, Smith MD. Health Technology Assessment: Lessons learned from around the world – an overview. *Value Health* 12, S1–S5 (2009).
4. Hogervorst MA, Pontén J, Vreman RA, Mantel-Teeuwisse AK, Goettsch WG. Real World Data in Health Technology Assessment of Complex Health Technologies. *Front. Pharmacol.* 13, (2022).
5. De Meulemeester J, Fedyk M, Jurkovic L *et al.* Many randomized clinical trials may not be justified: a cross-sectional analysis of the ethics and science of randomized clinical trials. *J. Clin. Epidemiol.* 97, 20–25 (2018).
6. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 334(7589), 349–351 (2007).
7. Simpson A, Ramagopalan SV. R WE ready for reimbursement? A round up of developments in real-world evidence relating to health technology assessment: part 10. *J. Comp. Eff. Res.* 12(1), e220194 (2023).
8. Center for Drug Evaluation and Research, Center for Biologics Evaluation, Oncology Center of Excellence. Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products, U.S. Food and Drug Administration. (2023). <https://www.fda.gov/media/171667/download>.
9. National Institute for Health and Care Excellence. NICE real-world evidence framework, National Institute for Health and Care Excellence. (2022). <https://www.nice.org.uk/corporate/ecd9/chapter/methods-for-real-world-studies-of-comparative-effects>.
10. Gatto NM, Campbell UB, Rubinstein E *et al.* The structured process to identify fit-for-purpose data: a data feasibility assessment framework. *Clin. Pharmacol. Ther.* 111(1), 122–134 (2022).
11. European Medicines Agency, European Medicines Regulatory Network. “DARWIN EU” (2023). <https://www.darwin-eu.org/>.
12. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. E5 Ethnic Factors in the Acceptability of Foreign Clinical Data., U.S. Food and Drug Administration. (1998). <https://www.fda.gov/media/71287/download>.
13. Drummond M, Barbieri M, Cook J *et al.* Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force Report. *Value Health* 12(4), 409–418 (2009).
14. Goeree R, He J, O'Reilly D *et al.* Transferability of health technology assessments and economic evaluations: a systematic review of approaches for assessment and application. *Clin. Outcomes Res. CEOR* 3, 89–104 (2011).
15. Jaksa A, Arena PJ, Chan KKW, Ben-Joseph RH, Jónsson P, Campbell UB. Transferability of real-world data across borders for regulatory and health technology assessment decision-making. *Front. Med.* 9 (2022).
16. Turner AJ, Sammon C, Latimer N *et al.* Transporting comparative effectiveness evidence between countries: considerations for health technology assessments. *Pharmacoeconomics* (2023).
17. Degtiar I, Rose S. A review of generalizability and transportability. *Annu. Rev. Stat. Its Appl.* 10(1), 501–524 (2023).
18. Pearl J, Bareinboim E. Transportability of causal and statistical relations: a formal approach. Presented at: *Proceedings of the 25th AAAI Conference on Artificial Intelligence*. AAAI Press, CA, USA (August, 2011).
19. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *Am. J. Epidemiol.* 186(8), 1010–1014 (2017).
20. Ling AY, Montez-Rath ME, Carita P *et al.* An overview of current methods for real-world applications to generalize or transport clinical trial findings to target populations of interest. *Epidemiol. Camb. Mass* 34(5), 627–636 (2023).
21. Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernán MA. Extending inferences from a randomized trial to a new target population. *Stat. Med.* 39(14), 1999–2014 (2020).
22. Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Stat. Sci.* 29(4), 579–595 (2014).
23. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev. Sci. Off. J. Soc. Prev. Res.* 16(3), 475–485 (2015).

24. Kern HL, Stuart EA, Hill J, Green DP. Assessing methods for generalizing experimental impact estimates to target populations. *J. Res. Educ. Eff.* 9(1), 103–127 (2016).
25. Rudolph KE, van der Laan MJ. Robust estimation of encouragement-design intervention effects transported across sites. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 79(5), 1509–1525 (2017).
26. Dong N, Stuart EA, Lenis D, Quynh Nguyen T. Using propensity score analysis of survey data to estimate population average treatment effects: a case study comparing different methods. *Eval. Rev.* 44(1), 84–108 (2020).
27. Lu H, Cole SR, Howe CJ, Westreich D. Toward a clearer definition of selection bias when estimating causal effects. *Epidemiol. Camb. Mass* 33(5), 699–706 (2022).
28. Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC, FL, USA (2020).
29. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *J. R. Stat. Soc. Ser. A Stat. Soc.* 174(2), 369–386 (2001).
30. Inoue K, Hsu W, Arah OA, Prosper AE, Aberle DR, Bui AAT. Generalizability and transportability of the national lung screening trial data: extending trial results to different populations. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* 30(12), 2227–2234 (2021).
31. Effective Health Care Program. Research report: developing a protocol for observational comparative effectiveness research: a user's guide. Agency for Healthcare Research and Quality, MD, USA (2019). <https://effectivehealthcare.ahrq.gov/products/observational-cer-protocol/research>.
32. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. US Food and Drug Administration (2013). <https://www.fda.gov/media/79922/download>.
33. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products. US Food and Drug Administration (2021). <https://www.fda.gov/media/152503/download>.
34. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. [A19-43] Development of Scientific Concepts for the Generation of Routine Practice Data and Their Analysis for the Benefit Assessment of Drugs According to §35a Social Code Book V—rapid report, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. (2020). <https://www.iqwig.de/en/projects/a19-43.html>.
35. Jaksa A, Wu J, Jónsson P, Eichler H-G, Vittoe S, Gatto NM. Organized structure of real-world evidence best practices: moving from fragmented recommendations to comprehensive guidance. *J. Comp. Eff. Res.* 10(9), 711–731 (2021).
36. Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. *BMJ* 328(7454), 1490 (2004).
37. Sterne JA, Hernán MA, Reeves BC et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 355, i4919 (2016).
38. Webster-Clark M, Toh S, Arnold J, McTigue KM, Carton T, Platt R. External validity in distributed data networks. *Pharmacoepidemiol. Drug Saf.* 32(12), 1360–1367 (2023).
39. Westreich D, Edwards JK. Invited commentary: every good randomization deserves observation. *Am. J. Epidemiol.* 182(10), 857–860 (2015).
40. Dahabreh IJ, Haneuse SJ-PA, Robins JM et al. Study designs for extending causal inferences from a randomized trial to a target population. *Am. J. Epidemiol.* 190(8), 1632–1642 (2021).
41. Gershman B, Guo DP, Dahabreh IJ. Using observational data for personalized medicine when clinical trial evidence is limited. *Fertil. Steril.* 109(6), 946–951 (2018).
42. Dahabreh IJ, Matthews A, Steingrimsson JA, Scharfstein DO, Stuart EA. Using trial and observational data to assess effectiveness: trial emulation, transportability, benchmarking, and joint analysis. *Epidemiol. Rev.* mxac011 (2023).
43. Josey KP, Yang F, Ghosh D, Raghavan S. A calibration approach to transportability and data-fusion with observational data. *Stat. Med.* 41(23), 4511–4531 (2022).
44. Wu Y, Hui J, Deng Q. Empirical profile Bayesian estimation for extrapolation of historical adult data to pediatric drug development. *Pharm. Stat.* 19(6), 787–802 (2020).
45. Lee D, Yang S, Dong L, Wang X, Zeng D, Cai J. Improving trial generalizability using observational studies. *Biometrics* 79(2), 1213–1225 (2023).
46. Montez-Rath ME, Lubwama R, Kapphahn K et al. Characterizing real world safety profile of oral Janus kinase inhibitors among adult atopic dermatitis patients: evidence transporting from the rheumatoid arthritis population. *Curr. Med. Res. Opin.* 38(8), 1431–1437 (2022).
47. Cook RR, Foot C, Arah OA et al. Estimating the impact of stimulant use on initiation of buprenorphine and extended-release naltrexone in two clinical trials and real-world populations. *Addict. Sci. Clin. Pract.* 18(1), 11 (2023).

48. Basu A, Patel C, Fu AZ, Brown B, Mavros P, Benson C. Real-world calibration and transportability of the Disease Recovery Evaluation and Modification (DREaM) randomized clinical trial in adult Medicaid beneficiaries with recent-onset schizophrenia. *J. Manag. Care Spec. Pharm.* 29(3), 293–302 (2023).
49. Mollan KR, Pence BW, Xu S *et al.* Transportability from randomized trials to clinical care: on initial HIV treatment with efavirenz and suicidal thoughts or behaviors. *Am. J. Epidemiol.* 190(10), 2075–2084 (2021).
50. Ramagopalan SV, Popat S, Gupta A *et al.* Transportability of overall survival estimates from US to Canadian patients with advanced non-small-cell lung cancer with implications for regulatory and health technology assessment. *JAMA Netw. Open* 5(11), e2239874 (2022).