Towards causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population

Issa J. Dahabreh[1,2,3], **Lucia C. Petito**[4], **Sarah E. Robertson**[1], **Miguel A. Hernán**[3,5,6], **Jon A. Steingrimsson**[7]

[1]Center for Evidence Synthesis in Health and Department of Health Services, Policy & Practice, School of Public Health, Brown University, Providence, RI

[2]Department of Epidemiology, School of Public Health, Brown University, Providence, RI

[3]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[4]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL

[5]Department of Biostatistics, Harvard School of Public Health, Boston, MA

[6]Harvard-MIT Division of Health Sciences and Technology, Boston, MA

[7]Department of Biostatistics, School of Public Health, Brown University, Providence, RI

Abstract

We take steps towards causally interpretable meta-analysis by describing methods for transporting causal inferences from a collection of randomized trials to a new target population, one-trial-at-a-time and pooling all trials. We discuss identifiability conditions for average treatment effects in the target population and provide identification results. We show that assuming inferences are transportable from all trials in the collection to the same target population has implications for the law underlying the observed data. We propose average treatment effect estimators that rely on different working models and provide code for their implementation in statistical software. We discuss how to use the data to examine whether transported inferences are homogeneous across the collection of trials, sketch approaches for sensitivity analysis to violations of the identifiability conditions, and describe extensions to address non-adherence in the trials. Last, we illustrate the proposed methods using data from the HALT-C multi-center trial.

Keywords

meta-analysis; evidence synthesis; causal inference; conditional average treatment effect; inverse probability weighting

Introduction

For many causal questions, we have effect estimates from several randomized trials that recruit individuals from different underlying populations, typically by convenience sampling. We would then like to synthesize the trial evidence and transport inferences to some target population that is chosen on substantive grounds [1].

Studies that synthesize findings across multiple trials are known as "meta-analyses" [2]. Most meta-analyses combine the trial-specific effect estimates to obtain a pooled effect estimate using either random effects models or common effect models (often referred to as "fixed effect" models) [3–5]. A commonly overlooked problem, however, is that standard meta-analyses may produce results that do not have a clear causal interpretation when each trial includes individuals from a different population and the treatment effect varies across those populations.

As an example, consider the effect of angiotensin-converting enzyme (ACE) inhibitors, compared with other anti-hypertensive agents, in individuals with nondiabetic chronic kidney disease [6–8]. A meta-analysis of 11 trials found that ACE inhibitors improved renal function and reduced progression to end-stage renal disease; the benefits of ACE inhibitors were substantially greater among individuals with high urine protein excretion than those with lower excretion [8, 9]. Across all trials, only 6% of participants were African American and in only 34% of participants the cause of kidney disease was hypertensive nephrosclerosis. Now suppose that we are interested in the effect of ACE inhibitors in a target population of non-diabetic African American adults with chronic kidney disease [10, 11]. In African Americans, chronic kidney disease is most often due to hypertension; and proteinuria is more common compared with whites [12, 13]. In addition, blood pressure in hypertensive African Americans is generally considered to respond better to calcium channel blocker or diuretic monotherapy, rather than ACE inhibitor treatment [14]. Given these differences and the under-representation of African Americans in the trials (in fact, 7 of the 11 trials did not enroll any African Americans), the estimate from a conventional meta-analysis of the 11 trials would have limited relevance for the target population.

In fact, whenever treatment effects are heterogeneous over variables that vary in distribution across trials (as is the case in the ACE inhibitor example) standard meta-analyses of treatment effects from the trials do not produce estimates with a clear causal interpretation for any reasonable target population. The problem arises because meta-analyses combine trial-specific effect estimates using weights that reflect the precision of the estimates (and their variability, for random effects meta-analyses) rather than their relevance to the target population [15]. In fact, most meta-analyses do not even specify their target population, regardless of whether the meta-analysis is based on summary statistics or on individual participant data from each trial. Similar concerns about conventional meta-analysis methods have been recently summarized, independently of our work, in [16].

In this paper, we take steps towards causally interpretable meta-analysis by proposing methods for extending inferences about average treatment effects from a collection of randomized trials to a target population, one-trial-at-a-time and pooling all trials. Our

approach requires individual participant data from the randomized trials together with baseline covariate data from a random sample from the target population. The latter allows us to account for differences in the distribution of effect modifiers between the trial-specific populations and the target population. Besides describing identification results, we show that assuming inferences are transportable from all trials in the collection to the same target population has implications for the law underlying the observed data. We propose average treatment effect estimators that rely on different working models and provide code for their implementation in statistical software. We discuss how to use the data to examine whether transported inferences are homogeneous across the collection of trials, sketch approaches for sensitivity analysis to violations of the identifiability conditions, and describe extensions to address non-adherence in the randomized trials. Last, we illustrate the proposed methods using data from the HALT-C multi-center trial.

Causal quantities of interest

Suppose we have individual-level data from participants in a collection of randomized trials $\mathcal{S} = \{1, \ldots, m\}$. For simplicity, we assume that all trials estimate the effect of assignment to treatment Z that takes values in the finite set $\mathcal{Z}$ on an outcome Y measured at some fixed follow-up time (we do not consider failure-time outcomes). If some treatments of interest are only assessed in a subset of the available trials, we can introduce treatment pair-specific subsets of the collection $\mathcal{S}(z, z') \subseteq \mathcal{S}$ that consist of those trials in $\mathcal{S}$ that compared treatments z and z'. With this change, our results can be extended to so-called "network meta-analyses" [17], in which not all treatments are assessed in all trials.

For each participant in each of the trials in the collection $\mathcal{S}$, we have information on the trial S in which they participated, treatment assignment Z, baseline covariates X, and outcome Y. Therefore, for each trial $s \in \mathcal{S}$, the data consist of realizations of independent random tuples $(X_i, S_i = s, Z_i, Y_i)$, $i = 1, \ldots, n_s$, where n_s denotes the total number of randomized individuals in trial s.

We also obtain a simple random sample from the target population of interest. The individuals in this sample are not participating in any of the trials (either because they were not invited or were invited but declined to participate). We collect baseline covariate data from them, but need not collect treatment or outcome data; we discuss the relationship between the covariate distribution in the target population and the collection of trials in the Identification section. We use the convention that $S = 0$ for the target population so the data from the sample of the target population consist of realizations of random tuples $(X_i, S_i = 0)$, $i = 1, \ldots, n_0$, where n_0 is the total number of sampled individuals. Throughout, we use $I(\cdot)$ to denote the indicator function; for example, $I(S \in \mathcal{S})$ is a random variable that denotes participation in any of the trials in the collection $\mathcal{S}$, such that $I(S \in \mathcal{S}) = 1$ when $S \in \mathcal{S}$; and 0, otherwise.

We form a composite dataset by appending the data from all the trials and the sample from the target population. The dataset consists of independent realizations of the random tuple $(X_i, S_i, I(S_i \in \mathcal{S}) \times Z_i, I(S_i \in \mathcal{S}) \times Y_i)$, $i = 1, \ldots, n$, where n is the total number

of observations in the trials and the sample from the target population, $n = \sum_{s=0}^{m} n_s$. Throughout, we use italicized capital letters to denote random variables and lower case letters to denote the corresponding realizations. We generically use the notation $f(\cdot)$ to denote densities.

We discuss details of the underlying sampling model in eAppendix A. Briefly, we assume that the observed data are obtained by random sampling from an infinite superpopulation of individuals that is stratified by S, with sampling fractions that are unknown and possibly unequal constants for each stratum with $S = s$, for $s \in \{0, 1, …, m\}$. We refer to our model as a biased sampling model [18] because the proportion of individuals in the data with $S = s$, for $s \in \{0, 1, …, m\}$, does not in general reflect the population probability of $S = s$, but is instead shaped by the complex process that drives the design and conduct of actual randomized trials (see references [19–21] for additional details in the single trial case).

Let Y^z denote the potential (counterfactual) outcome under treatment assignment $z \in \mathcal{Z}$ [22, 23]. We are interested in estimating the average treatment effect of treatment assignment (i.e., the intention-to-treat effect) in the target population, $E[Y^z - Y^{z'}|S = 0]$ for every pair of treatments z and z' in $\mathcal{Z}$.

Identification of average treatment effects in the target population using a single trial

We now consider transporting inferences from each trial to the target population. Derivations for all identification results reported in this section are provided in eAppendix B. The derivations are valid under the biased sampling model described in eAppendix A and thus the expectations and probabilities below can be taken over the data law of that model.

Identifiability conditions

The following are sufficient conditions for identifying the average treatment effect in the target population, $E[Y^z - Y^{z'}|S = 0]$, using covariate, treatment, and outcome data from a single trial $s* \in \mathcal{S}$ and baseline covariate data from the sample of the target population.

A1. *Consistency of potential outcomes:* If $Z_i = z$, then $Y_i^z = Y_i$, for every individual i in trial $s*$ or the target population. Implicit in this notation is the absence of any effect of trial engagement on the outcome [24, 25] (e.g., no Hawthorne effects).

A2. *Conditional exchangeability over treatment assignment Z in the trial:* $E[Y^z|X = x, S = s*, Z = z] = E[Y^z|X = x, S = s*]$ for each treatment $z \in \mathcal{Z}$, and each x with $f(x, S = s*) > 0$. This condition is expected to hold because treatment assignment in each trial is randomized (possibly conditional on X).

A3. *Positivity of the treatment assignment probability in the trial:* For each treatment $z \in \mathcal{Z}$, $Pr[Z = z|X = x, S = s*] > 0$ for every x with $f(x, S = s*) > 0$. This condition is also expected to hold because of randomization.

A4. *Conditional exchangeability in measure between the trial and the target population:* For each pair of treatments z and z' in $\mathcal{Z}$, $E[Y^z - Y^{z'}|X =$

$x, S = 0] = E[Y^z - Y^{z'}|X = x, S = s^*]$ for every x with $f(x, S = 0)$ > 0. This condition encodes an assumption of no residual effect measure modification by trial participation, conditional on the baseline covariates X. It is mathematically weaker than the identifiability conditions used in most previous work on transportability (e.g., [26]), but it is sufficient to transport inferences about the average treatment effects from trial s^* to the target population [19]. Exchangeability in measure may be implied by distributional independence conditions, for example, $Y^z \perp\!\!\!\perp I(S = s^*)|X, I(S \in \{0, s^*\}) = 1$, but exchangeability in measure can hold even if such independence conditions do not. Thus, the condition cannot be verified solely using graphical methods such as d-separation-based criteria applied to single world intervention graphs [24] or selection diagrams [27].

A5. *Positivity of the probability of participation in the trial:* $\Pr[S = s^*|X = x]$ > 0 for every x with $f(x, S = 0)$ > 0. Informally, this condition means that, for covariates needed to establish conditional exchangeability in measure, each covariate pattern in the target population should have positive probability of occurring in the trial s^*.

In stating the identifiability conditions, we have used X generically to denote baseline covariates. It is possible however, that strict subsets of X are adequate to satisfy the exchangeability condition for trial s^* and the target population, or conditional exchangeability in measure. For example, if trial s^* is marginally randomized, the mean exchangeability over treatment assignment A among trial participants will hold unconditionally. Also, the identification results we obtain under these conditions apply beyond randomized trials, for example, to pooled analyses of observational studies comparing interventions [28], provided that conditions *A2* and *A3* can be assumed to hold. In this setting, condition *A2* corresponds to the usual "no unmeasured confounding" assumption [29], applied to study s^*.

To focus on issues related to selective trial participation, we assume complete adherence to the trial protocol and no loss-to-follow-up, so that the intention-to-treat effect (i.e., the average effect of treatment assignment) is equal to the per-protocol effect (i.e., the average effect of receiving treatment as indicated in the protocol). If adherence to the protocol is incomplete, the two effects are not equal. The methods we describe here can be extended to account for non-adherence to the trial protocol, but the extensions do not offer additional insights regarding transportability from a collection of trials to a target population. As an example, in eAppendix C, we demonstrate how the methods can be extended to address incomplete adherence in a two-period study. In the remainder of the main text, for brevity, we use the term "average treatment effect" to mean the intention-to-treat average effect in the target population ($S = 0$).

Identification

Under conditions *A1* through *A5*, when transporting inferences comparing treatments z and z' from a trial $s^* \in \mathcal{S}$ to the target population, the average treatment effect $E[Y^z - Y^{z'}|S = 0]$ equals the following functional of the observed data distribution:

$$\psi(z, z', s^*) \equiv E[E[Y \mid X, S = s^*, Z = z] - E[Y \mid X, S = s^*, Z = z'] \mid S = 0]. \quad (1)$$

Under the positivity conditions, we can obtain an expression for $\psi(z, z', s^*)$ that uses weighting,

$$\psi(z, z', s^*) = \frac{1}{\Pr[S = 0]} E\left[\left(\frac{I(Z = z)}{\Pr[Z = z \mid X, S = s^*]} - \frac{I(Z = z')}{\Pr[Z = z' \mid X, S = s^*]} \right) \frac{I(S = s^*) Y \Pr[S = 0 \mid X, I(S \in \{0, s^*\}) = 1]}{\Pr[S = s^* \mid X, I(S \in \{0, s^*\}) = 1]} \right]. \quad (2)$$

The identification results in this section only involve the target population $S = 0$ and the trial $S = s^*$; thus, in addition to holding under the sampling model of eAppendix A they also hold under the sampling model described in reference [20].

Identification when pooling the trials

We now consider transporting inferences to the target population when using the data from all trials in the collection $\mathcal{S}$. Derivations for all identification results reported in this section are also provided in eAppendix B and are valid under the biased sampling model described in eAppendix A.

Identifiability conditions

The following are sufficient conditions for identifying the average treatment effect in the target population, $E[Y^z - Y^{z'} \mid S = 0]$, using covariate, treatment, and outcome data from every trial in the collection $\mathcal{S}$ and baseline covariate data from the sample of the target population.

B1. *Consistency of potential outcomes:* If $Z_i = z$, then $Y_i^z = Y_i$, for every individual i in the target population or the populations underlying the trials in $\mathcal{S}$. As noted before, implicit in this notation is the absence of any effect of trial engagement on the outcome.

B2. *Conditional exchangeability over treatment assignment Z:* $E[Y^z \mid X = x, S = s, Z = z] = E[Y^z \mid X = x, S = s]$ for every trial $s \in \mathcal{S}$, each treatment $z \in \mathcal{Z}$, and every x with $f(x, S = s) > 0$.

B3. *Positivity of the treatment assignment probability in the trials:* For every treatment $z \in \mathcal{Z}$, $\Pr[Z = z \mid X = x, S = s] > 0$ for every trial $s \in \mathcal{S}$ and every x with $f(x, S = s) > 0$.

B4. *Conditional exchangeability in measure between the trial and the target population:* For every pair of treatments z and z' in $\mathcal{Z}$, $E[Y^z - Y^{z'} \mid X = x, S = 0] = E[Y^z - Y^{z'} \mid X = x, S = s]$ for every trial $s \in \mathcal{S}$ and every x with $f(x, S = 0) > 0$.

B5. *Positivity of the probability of participation in the trials:* $\Pr[S = s \mid X = x] > 0$ for every $s \in \mathcal{S}$ and every x with $f(x, S = 0) > 0$.

Observed data implications

Suppose now that conditions *B1* through *B5* hold, so that inferences are transportable from each trial $s \in \mathcal{S}$ to the target population; if effects are transportable only for a subset of the trials in the collection $\mathcal{S}$ it is easy to modify the results given here to restrict them to the subset of $\mathcal{S}$ that satisfies conditions *B1* through *B5*. We will argue that conditions *B1* through *B5*, have important implications for law underlying the observed data. To see this, note that when conditions *B4* and *B5* hold, then for the pair of treatments z, z', and for all X values in the common support with $S = 0$,

$$\mathrm{E}\left[Y^z - Y^{z'} \mid X, S = 1\right] = \dots = \mathrm{E}\left[Y^z - Y^{z'} \mid X, S = m\right] = \mathrm{E}\left[Y^z - Y^{z'} \mid X, S = 0\right].$$

That is, the conditional causal mean difference of each of the *m* trials in the collection $\mathcal{S}$ is equal to the conditional causal mean difference in the target population. This means that the conditional average causal effect is independent of study participation in $\mathcal{S}$, within strata of baseline covariates.

Using the above result and assumptions *B1* through *B3*, for the pair of treatments z, z' and for all X values in the common support with $S = 0$, we obtain

$$\begin{aligned} &\mathrm{E}[Y \mid X, S = 1, Z = z] - \mathrm{E}[Y \mid X, S = 1, Z = z'] = \dots = \\ &\mathrm{E}[Y \mid X, S = m, Z = z] - \mathrm{E}[Y \mid X, S = m, Z = z'] \equiv \tau(z, z'; X), \end{aligned} \tag{3}$$

where we use the notation $\tau(z, z'; X)$ for the common conditional mean difference.

The above chain of equalities is an observed data implication of conditions *B1* through *B5* because it only involves observed variables. Thus, using the data to evaluate the chain of equalities can be viewed as a falsification test for whether the chain of assumptions *B1* through *B5* fails to hold for the entire collection $\mathcal{S}$. Nevertheless, such assessment cannot prove that the the assumptions hold for the collection or for any subset of trials. Substantive knowledge should be used to determine whether estimates of the conditional mean differences in the above equation are "close enough" across trials. Expert assessment may be aided by statistical testing, because the chain of equalities can be viewed as a null hypothesis to be tested using parametric or non-parametric methods [30–33].

Identification

Suppose that substantive knowledge suggests that assumptions *B1* through *B5* hold, so that all trials in the collection $\mathcal{S}$ are transportable to the target population, and that examination of the observed data implications of assumptions *B1* through *B5* does not raise any concern. Then, the average treatment effect $\mathrm{E}[Y^z - Y^{z'} | S = 0]$ equals the following functional of the observed data distribution:

$$\phi(z, z') \equiv \mathrm{E}[\tau(z, z'; X) \mid S = 0]. \tag{4}$$

Under the positivity conditions, we can obtain an expression for $\phi(z, z')$ that uses weighting,

$$\phi(z, z') = \frac{1}{\Pr[S = 0]} \mathrm{E} \left[\omega(z, z'; X, S, Y) \frac{Y \Pr[S = 0 \mid X]}{\Pr[I(S \in \mathcal{S}) = 1 \mid X]} \right], \qquad (5)$$

where

$$\omega(z, z'; X, S) = \left(\frac{I(Z = z)}{\Pr[Z = z \mid X, S, I(S \in \mathcal{S}) = 1]} - \frac{I(Z = z')}{\Pr[Z = z' \mid X, S, I(S \in \mathcal{S}) = 1]} \right) I(S \in \mathcal{S}).$$

Relaxing positivity condition *B5* when pooling the trials

Intuition suggests that by combining information across multiple trials it should be possible to relax positivity condition *B5*, which, as stated above, requires sufficient overlap between the target population and each one of the trials in the collection $\mathcal{S}$.

As an example, consider two randomized trials, comparing the same anti-hypertensive treatments, one enrolling patients with mild and the other with severe hypertension, and a target population, $S = 0$, that includes individuals with both mild and severe hypertension. It should be possible to obtain inferences about the target population provided one is willing to assume that (1) conditional average treatment effects from each trial are equal to conditional average treatment effects in the corresponding subset of the target population; and (2) every covariate pattern that can occur among individuals in the target population with mild hypertension has positive probability of occurring in the trial of mild hypertension and every covariate pattern that can occur among individuals in the target population with severe hypertension has positive probability of occurring in the trial of severe hypertension.

Using $\mathcal{X}_s$ to denote the support of the distribution of X given $S = s$, the second condition above can be generalized as follows: the union of the intersections of the support sets of the distribution of X in each trial in $\mathcal{S}$ and the target population equals the support set of the distribution of X in the target population. Using $\mathcal{X}_s$ to denote the support of the distribution of X given $S = s$, this more general overlap condition can be written concisely as $\bigcup_{s \in \mathcal{S}} \mathcal{X}_s \cap \mathcal{X}_0 = \mathcal{X}_0$. Though identification is intuitively obvious in simple cases like the two trial hypertension example, handling identification in general cases when condition *B5* is violated but the overlap condition $\bigcup_{s \in \mathcal{S}} \mathcal{X}_s \cap \mathcal{X}_0 = \mathcal{X}_0$ holds, requires modifications to condition *B4* and the introduction of additional notation to keep track of the subsets of $\mathcal{S}$ where each covariate pattern $X = x$ can occur. For that reason, we we do not pursue it here.

Estimation & Inference

Estimation for transporting individual trials

We now use the identification results on the previous section to propose estimators of average treatment effects; when transporting results from a single randomized trial, these estimators relate to the potential outcome mean estimators of [19].

Estimation by modeling the conditional average treatment effect: The first option is to use an estimator based on (1),

$$\widehat{\psi}_{\text{te}}(z, z', s*) = \left\{ \sum_{i=1}^{n} I(S_i = 0) \right\}^{-1} \sum_{i=1}^{n} I(S_i = 0)\widehat{b}(z, z', s*; X_i), \tag{6}$$

where $\widehat{b}(z, z', s*; X)$ is an estimator of E[$Y|X, S = s*, Z = z$] − E[$Y|X, S = s*, Z = z'$]. A simple way to estimate $\widehat{b}(z, z', s*; X)$ is to use an outcome model-based estimator $\widehat{h}(z, s*; X)$ of E[$Y|X, S = s*, Z = z$], for each $z \in \mathcal{Z}$, and then estimate $\widehat{b}(z, z', s*; X)$ by taking the difference $\widehat{h}(z, s*; X) - \widehat{h}(z', s*; X)$. Alternatively, because under assumptions *A1* through *A3*, E[$Y|X, S = s*, Z = z$] − E[$Y|X, S = s*, Z = z'$] = E[$Y^z - Y^{z'}|X, S = s*$], we can use a g-estimator of the conditional average treatment effect in trial $s*$ [34–37]. Regardless of the approach, $\widehat{\psi}_{\text{te}}(z, z', s*)$ converges in probability to $\psi(z, z', s*)$ when $\widehat{b}(z, z', s*; X)$ is a consistent estimator of the conditional average treatment effect in trial $s*$.

Estimation by modeling the probability of trial participation and treatment: A second option is to use a weighting estimator based on (2),

$$\widetilde{\psi}_{\text{w}}(z, z', s*) = \left\{ \sum_{i=1}^{n} I(S_i = 0) \right\}^{-1} \sum_{i=1}^{n} \left(\frac{I(Z_i = z)}{\widehat{\ell}\,(z, s*; X_i)} - \frac{I(Z_i = z')}{\widehat{\ell}\,(z', s*; X_i)} \right) I(S_i = s* \\ ) \frac{\widetilde{p}(X_i)}{1 - \widetilde{p}(X_i)} Y_i, \tag{7}$$

where $\widetilde{p}(X)$ is an estimator of Pr[$S = 0|X, I(S \in \{0, s*\}) = 1$]; $\widehat{\ell}\,(z, s*; X)$ and $\widehat{\ell}\,(z', s*; X)$ are estimators for Pr[$Z = z|X, S = s*$] and Pr[$Z = z'|X, S = s*$], respectively.

Estimation for pooling the trials

Estimation by modeling the conditional average treatment effect: The first option is to use an estimator based on (4),

$$\widehat{\phi}_{\text{te}}(z, z') = \left\{ \sum_{i=1}^{n} I(S_i = 0) \right\}^{-1} \sum_{i=1}^{n} I(S_i = 0)\widehat{t}(z, z'; X_i), \tag{8}$$

where $\widehat{t}(z, z'; X)$ is an estimator of the common (across trials) conditional mean difference $\tau(z, z'; X)$. A general way to obtain $\widehat{t}(z, z'; X)$ is to estimate the parameters of a regression model for the following conditional mean function

$$\text{E}\left[\left(\frac{I(Z = z)}{\Pr[Z = z \mid X, S, I(S \in \mathcal{S}) = 1]} - \frac{I(Z = z')}{\Pr[Z = z' \mid X, S, I(S \in \mathcal{S}) = 1]} \right) Y \mid X, I(S \in \mathcal{S}) = 1 \right],$$

where the probabilities in the denominators of the fractions inside the parenthesis are known (or can be estimated) [34–37]. This approach does not require the treatment assignment mechanism to be the same across trials. The estimator $\widehat{\phi}_{\text{te}}(z, z')$ converges in probability to $\phi(z, z')$ when $\widehat{t}(z, z'; X)$ is a consistent estimator of the (common across trials) conditional average treatment effect in the collection of trials $\mathcal{S}$.

Estimation by modeling the probability of trial participation and treatment: A second option is to use a weighting estimator based on (5),

$$
\hat{\phi}_{\mathrm{w}}(z, z') = \left\{ \sum_{i=1}^{n} I(S_i = 0) \right\}^{-1} \sum_{i=1}^{n} \left(\frac{I(Z_i = z)}{\hat{e}(z; X_i, S_i)} - \frac{I(Z_i = z')}{\hat{e}(z'; X_i, S_i)} \right) I(S_i \in \mathcal{S}
$$
$$
) \frac{\hat{p}(X_i)}{1 - \hat{p}(X_i)} Y_i,
$$

(9)

where $\hat{p}(X)$ is an estimator of $\Pr[S = 0 | X]$, and $\hat{e}(z; X, S)$ and $\hat{e}(z'; X, S)$ are estimators for $\Pr[Z = z \mid X, S, I(S \in \mathcal{S}) = 1]$ and $\Pr[Z = z' \mid X, S, I(S \in \mathcal{S}) = 1]$, respectively. This weighted estimator converges in probability to $\phi(z, z')$ when $\hat{p}(X)$ is a consistent estimator of $\Pr[S = 0 | X]$ and $\hat{e}(z; X, S)$ is a consistent estimator of $\Pr[Z = z \mid X, S, I(S \in \mathcal{S}) = 1]$. Correct specification of a model for $\Pr[Z = z \mid X, S, I(S \in \mathcal{S}) = 1]$ is straightforward even if trials have different treatment assignment mechanisms (but may prove more challenging in pooled analyses of observational studies).

Inference

To construct Wald-style confidence intervals for $\hat{\psi}_{\mathrm{te}}(z, z', s^*)$, $\hat{\psi}_{\mathrm{w}}(z, z', s^*)$, $\hat{\phi}_{\mathrm{te}}(z, z')$, or $\hat{\phi}_{\mathrm{w}}(z, z')$, when using parametric models, we can obtain "sandwich" estimators of the sampling variance for each estimator we have described [38]. Alternatively, we can use the non-parametric bootstrap [39]. Straightforward inference approaches are possible because under our sampling model the observations are assumed to be independent. We leave extensions of the sampling model to allow for dependence among the individuals sampled into each trial for future work.

Implementation of the estimators

In eAppendix D we provide a link to R code implementing the estimators described above using the geex package [40]. Our implementation uses parametric working models: we use estimating equations for binary or multinomial logistic regression models to estimate conditional probabilities; and estimating equations for linear regression models to estimate conditional expectations. All estimating equations can be easily replaced by those appropriate for other generalized linear models [41], as needed. Because we estimate the working model parameters jointly with the target parameters of all estimators, the sampling variances appropriately account for uncertainty [38, 42]. The code can be modified to address non-adherence (and incomplete follow-up) as needed.

Violations of conditional exchangeability in measure

Conditional exchangeability in measure (for individual studies or for the collection $\mathcal{S}$, as applicable) will often not hold exactly in applications, when some modifiers of the treatment effect are not included in X. For example, in the case of transporting inferences from a single trial $s^* \in \mathcal{S}$, comparing treatments z and z' in $\mathcal{Z}$, it is possible that for some x with $f(x, S = 0) > 0$, we will have

$$\mathrm{E}\Big[Y^z - Y^{z'} \mid X = x, S = s*\Big] \neq \mathrm{E}\Big[Y^z - Y^{z'} \mid X = x, S = 0\Big],$$

and none of the methods described earlier in this paper will be able to provide valid inferences. Furthermore, substantive experts may disagree on the plausibility of conditional exchangeability in measure in any given application. When conditional exchangeability in measure is implausible or controversial, it is prudent to perform sensitivity analyses to examine how the study conclusions would change under violations of the condition with different magnitudes [43].

When transporting from individual studies in the meta-analysis:

For individual studies in the collection $\mathcal{S}$, we can perform sensitivity analysis by modifying the methods described in reference [44].

Sensitivity analysis model: A convenient way to parameterize the sensitivity analysis is to assume the following *sensitivity analysis model*, conditional on baseline covariates:

$$\mathrm{E}\Big[Y^z - Y^{z'} \mid X, S = 0\Big] = \mathrm{E}\Big[Y^z - Y^{z'} \mid X, S = s*\Big] + u(s*; X).$$

Here, $u(s*; X)$ is a user-specified, possibly study-specific and covariate-dependent, *bias function* that expresses the degree of residual effect modification by trial participation, within levels of the measured covariates in X (see [44] for a similar approach to sensitivity analysis when transporting inferences from a single trial and for suggestions for how to choose bias correction functions in applications).

For the choice of $u(s*; X)$ function for which the sensitivity analysis model holds, by taking expectations, we obtain

$$\begin{aligned}
\mathrm{E}\Big[Y^z - Y^{z'} \mid S = 0\Big] &= \mathrm{E}\Big[\mathrm{E}\Big[Y^z - Y^{z'} \mid X, S = s*\Big] + u(s*; X) \mid S = 0\Big] \\
&= \mathrm{E}\Big[\mathrm{E}\Big[Y^z - Y^{z'} \mid X, S = s*\Big] \mid S = 0\Big] + \mathrm{E}\big[u(s*; X) \mid S = 0\big].
\end{aligned} \tag{10}$$

Suppose that conditions *A1*, *A2*, *A3*, and *A5* hold, but condition *A4* (conditional exchangeability in measure) may not hold. In such cases, using the result in equation (10), we obtain

$$\mathrm{E}\Big[Y^z - Y^{z'} \mid S = 0\Big] = \psi(z, z', s*) + \mathrm{E}\big[u(s*; X) \mid S = 0\big]. \tag{11}$$

Estimation and inference for sensitivity analysis: The result in equation (11) suggest that sensitivity analysis estimators can be obtained by adding the term

$$\left\{ \sum_{i=1}^{n} I(S_i = 0) \right\}^{-1} \sum_{i=1}^{n} I(S_i = 0) u(s*; X_i)$$

to the estimators of $\psi(z, z', s^*)$ in the previous section. This "sensitivity analysis term" converges in probability to $E[u(s^*; X)|S = 0]$, under mild conditions. The function $u(s^*; X)$ is not identifiable from the data and thus sensitivity analysis should be performed over a range of different functions. Confidence intervals can again be obtained using either "sandwich" estimates of the variance or bootstrap methods. For example, the code we provide can be easily modified to obtain "sandwich" estimates of the variance simply by adding the sensitivity analysis term to the summands of the estimating equations.

Sensitivity analysis for transporting from multiple trials:

Sensitivity analysis for transporting inferences from multiple trials would follow the same principles as outlined for the case of individual studies. In the absence of conditional exchangeability in measure from one or more of the trials in the collection $\mathcal{S}$ trials (i.e., if assumption *B4* does not hold), the observed data implications, and the chain of equalities in (3) in particular, would not be expected to hold. Furthermore, the choice of bias correction functions for each study $u(s; X)$ for each trial $s \in \mathcal{S}$ would be more complicated because the functions would need to satisfy the following constraint (if all trials in the collection are transportable):

$$
\begin{aligned}
&E[Y \mid X, S = 1, Z = z] - E[Y \mid X, S = 1, Z = z'] + u(1; X)\\
&= E[Y \mid X, S = 2, Z = z] - E[Y \mid X, S = 2, Z = z'] + u(2; X)\\
&= \ldots\\
&= E[Y \mid X, S = m, Z = z] - E[Y \mid X, S = m, Z = z'] + u(m; X).
\end{aligned}
$$

We leave the development of methods for harnessing expert knowledge to choose appropriate $u(s; X)$ functions for sensitivity analysis as future work.

Homogeneity of transported effects

When investigators are willing to assume (and do not have contrary evidence) that exchangeability in measure holds for two or more trials in the collection $\mathcal{S}$, the observed data implications about conditional average treatment effects, have additional implications for the transported effects. Suppose that identifiability conditions *B1* through *B5* hold, then it has to be that

$$
\begin{aligned}
&E[E[Y \mid X, S = 1, Z = z] - E[Y \mid X, S = 1, Z = z'] \mid S = 0]\\
&= E[E[Y \mid X, S = 2, Z = z] - E[Y \mid X, S = 2, Z = z'] \mid S = 0]\\
&= \ldots\\
&= E[E[Y \mid X, S = m, Z = z] - E[Y \mid X, S = m, Z = z'] \mid S = 0],
\end{aligned}
$$

where, as defined earlier, m is the total number of trials in $\mathcal{S}$. This implication of conditions *B1* through *B5* can be interpreted as a condition of homogeneity of transported inferences among the trials in $\mathcal{S}$.

It is instructive to compare the homogeneity of transported inferences implication against the usual meta-analytic assumption of homogeneity of the study-specific average treatment effects in the same collection of trials,

$$\text{E}[\text{E}[Y \mid X, S = 1, Z = z] - \text{E}[Y \mid X, S = 1, Z = z'] \mid S = 1]$$
$$= \text{E}[\text{E}[Y \mid X, S = 2, Z = z] - \text{E}[Y \mid X, S = 2, Z = z'] \mid S = 2]$$
$$= \dots$$
$$= \text{E}[\text{E}[Y \mid X, S = m, Z = z] - \text{E}[Y \mid X, S = m, Z = z'] \mid S = m].$$

This chain of equalities represents the null hypothesis adopted for most "tests for heterogeneity" used in applied meta-analyses [3]. The critical difference is this: the homogeneity of transported inferences implication of conditions *B1* through *B5* is a comparison of functionals that involve marginalization over (i.e., standardization to) the same baseline covariate distribution, that of the target population $S = 0$. In contrast, the usually homogeneity assumption in meta-analyses is a comparison of functionals that involve marginalization over a different baseline covariate distribution, that of each trial $S = s$. The latter may not be equal, even if assumptions *B1* through *B5* hold, when the distribution of X varies across trials.

Application of the methods to data from the HALT-C trial

Using data from a multicenter trial to emulate a meta-analysis

The HALT-C trial enrolled 1050 patients with chronic hepatitis C and advanced fibrosis who had not responded to previous therapy and randomized them to treatment with peginterferon alfa-2a ($z = 1$) versus no treatment ($z = 0$). Patients were recrutied in 10 research centers and followed up every 3 months after randomization. We used data on the secondary outcome of platelet count at 9 months of follow-up; we report all outcome measurements as platelets $\times 10^3$/ml. For simplicity, we restricted our analysis to 974 patients with complete baseline covariate and outcome data.

We used the HALT-C trial data to emulate a meta-analysis: First, we treated the data from the center that contributed the most patients as a sample from the target population ($S = 0$; 202 patients). Second, we treated the data from the remaining 9 centers ($S = 1, \dots, 9$; 772 patients) as if derived from a collection of separate randomized trials ($\mathcal{S}$). Third, we transported inferences from $\mathcal{S}$ to $S = 0$ using the methods described in this paper. Fourth, we transported inferences from each trial s in the collection $\mathcal{S}$ to $S = 0$ using the methods described in [19].

The benefit of our approach for evaluating the methods is that by re-purposing data from a multi-center trial we have access to information on the randomly assigned treatment and outcome from $S = 0$, allowing us to compare analyses using data exclusively from $S = 0$ to the results of transportability analyses. These comparisons are informative because, provided the conditions needed for transporting inferences from the collection $\mathcal{S}$ (or from each center $s \in \mathcal{S}$) to $S = 0$ hold, we expect that estimates from transportability analyses should agree with the estimates from the analyses that exploit randomization in $S = 0$ (up to sampling variability).

Our HALT-C analysis was not considered human subjects research because it used de-identified secondary data.

Methods implemented and comparisons

We applied the estimators $\hat{\phi}_{\text{te}}(1,0)$ and $\hat{\phi}_{\text{w}}(1,0)$ to transport inferences from the collection of trials to $S = 0$. We also transported inferences separately from each trial to $S = 0$ using the estimators $\hat{\psi}_{\text{te}}(1,0,s)$ and $\hat{\psi}_{\text{w}}(1,0,s)$ for each $s \in \mathcal{S} = \{1, ..., 9\}$.

We specified parametric working models, as needed for each estimator (binary or multinomial logistic regression models for discrete outcomes, and linear regression models for continuous quantities).All working models used the following baseline covariates as main effects: baseline platelet count, age, sex, previous use of pegylated interferon, race, white blood cell count, history of injected recreational drugs, ever receiving a transfusion, body mass index, creatinine levels, and smoking status (to avoid numerical issues in the smallest centers, when transporting inferences from center 1 alone, we did not consider sex, previous use of pegylated interferon, and creatinine; when transporting inferences from center 9 alone, we did not consider sex, previous use of pegylated interferon, race, and creatinine).

Data analysis results

The results from the meta-analysis emulation are summarized in Table 1 and graphed in the forest plot [45] of Figure 1. When transporting inferences from the collection $\mathcal{S}$ to $S = 0$, the treatment effect and weighting estimator results were –43.7 and –42.4, respectively (the confidence interval was slightly narrower for the treatment effect estimator). These transportability results agree with the randomization-based analysis from $S = 0$ (using treatment and outcome data only from the target center) that produced an estimate of –45.7. It is interesting to note that the confidence interval for the crude mean difference in $S = 0$ was substantially wider than the confidence intervals for the two transportability estimates from $\mathcal{S}$ to $S = 0$; this reflects the fact that the collection $\mathcal{S}$ contains many more individuals with covariate, treatment, and outcome data (772 individuals) compared to the $S = 0$ sample (202 individuals). In this particular dataset, the additional information is evidently enough to overcome any imprecision induced by differences in the covariate distribution between the trials in $\mathcal{S}$ and $S = 0$).

In transportability analyses from each center $s \in \mathcal{S}$ to $S = 0$, point estimates from the treatment effect and weighting estimator were generally similar (and the former had narrower associated confidence intervals). For some centers, transportability analyses produced estimates that were reasonably close to the overall meta-analysis estimate, but the pattern was variable. For example, for $S = 1$, the transportability estimates (–50.9 for the treatment effect estimator; –48.9 for the weighting estimator) were closer to the overall transportability result (–43.7 for the treatment effect estimator; –42.4 for the weighting estimator) and the randomization-based analysis using $S = 0$ data (–45.7), compared to the analysis using only data from $S = 1$ (–29.3). Given the small sample size in some of the centers, analyses transporting inferences one trial at a time may be affected by sampling variability, over-fitting the data from individual centers, or model misspecification (because we relied on relatively simple parametric models).

Discussion

In our experience, users of evidence syntheses are not interested in the aggregate of the populations underlying the completed randomized trials, which is often ill-defined because of convenience sampling. Instead, for decision-making, the users typically have a new target population in mind. Traditional approaches to meta-analysis, however, yield pooled effect estimates that are not generally interpretable as the causal effect in that target population of interest.

Here, we described the conditions under which estimates from individual studies and pooled estimates can be causally interpreted as treatment effects in a target population which is chosen on substantive grounds and may be different from any of the of the populations sampled in the randomized trials. For example, our methods can be used when policy-makers want to examine the implications of a collection of trials for treatment effectiveness in a well-defined target population from which baseline covariate information is routinely collected (e.g., electronic health record data gathered by a healthcare system). The methods can also be used in planning a new trial when two or more related trials are already available: once the new trial's target population is specified, covariate data from it can be collected, and our methods can be used to obtain a treatment effect estimate in the target population to determine the feasibility of the new trial (e.g., through sample size calculations). In the rare case where one of the randomized trials in $\mathscr{S}$ is representative of the target population, the methods can be modified to treat the trial data as a sample from the target population.

The methods we propose relate to the general theory of causal identification when transporting inferences from multiple trials to a new setting [46, 47]. We address transportability from multiple trials by considering the underlying sampling model and provide both conditional treatment effect modeling and weighting approaches that can be used to reduce dependence on the specification of models for the conditional average treatment effect (or the conditional outcome mean) in the trials [48].

Other work on the causal interpretation of meta-analyses has focused on identifiability conditions and their use to examine the presence of heterogeneity across studies [49]; the causal interpretation of meta-analyses using published aggregate (summary) data [50]; or the use of aggregate data to estimate causal quantities in a "meta-population" that contains the individual superpopulations from each study included in a meta-analysis [51]. The estimators we propose are different from those used to illustrate a recently proposed framework for causally interpretable meta-analysis [49]; estimation in that work relied on conventional multivariable regression models for conditional average treatment effects.

A recent unpublished technical report [16] addresses the same limitations of conventional meta-analysis methods as our work. Our approach is complementary to that proposed in the report: our identifiability conditions are stronger, but sufficient for point identification of causal effects in the target population (and, as noted earlier, sensitivity analysis can be done); the assumptions in the report appear weaker, but only allow partial (interval) identification. Our estimators allow statistical inference using standard methods (e.g., M-

estimation or the bootstrap) and require access to individual-level data. In contrast, we expect that inference about the bounds in the report will often prove challenging because the target quantities are non-smooth and study-level data are limited (most meta-analyses include only a few studies, often with small sample sizes).

Our approach delineates the (non-parametric) identifiability conditions from any additional modeling assumptions that may be needed for estimation, especially when the vector of baseline covariates is high dimensional [52]. In this paper, we focused on simple conditional treatment effect and weighting estimators that can be easily implemented in standard statistical packages. A downside of these estimators is that, to produce valid results, the working models each estimator relies on need to be correctly specified. Also, it is possible to obtain doubly robust [53] estimators of the causal quantities of interest.

In summary, we have taken steps towards causally interpretable meta-analysis by considering the identification and estimation of average treatment effects when transporting inferences from a collection of randomized trials to a well-defined target population. To deploy the methods for applied evidence synthesis, extensions will be needed to address failure-time outcomes, systematically missing data, and measurement error in covariates or outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

References

[1]. Dahabreh Issa J and Hernán Miguel A. Extending inferences from a randomized trial to a target population. European Journal of Epidemiology, pages 1–4, 2019. [PubMed: 30547255]

[2]. Cooper Harris, Hedges Larry V, and Valentine Jeffrey C. The handbook of research synthesis and meta-analysis. Russell Sage Foundation, 2009.

[3]. Laird Nan M and Mosteller Frederick. Some statistical methods for combining experimental results. International Journal of Technology Assessment in Health care, 6(1):5–30, 1990. [PubMed: 2361819]

[4]. Higgins Julian PT, Thompson Simon G, and Spiegelhalter David J. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 172(1):137–159, 2009.

[5]. Rice Kenneth, Higgins Julian, and Lumley Thomas. A re-evaluation of fixed effect (s) meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 181(1):205–227, 2018.

[6]. Giatras Ioannis, Lau Joseph, and Levey Andrew S. Effect of angiotensin-converting enzyme inhibitors on the progression of nondiabetic renal diseasea meta-analysis of randomized trials. Annals of Internal Medicine, 127(5):337–345, 1997. [PubMed: 9273824]

[7]. Jafar Tazeen H, Schmid Christopher H, Landa Marcia, Giatras Ioannis, Toto Robert, Remuzzi Giuseppe, Maschio Giuseppe, Brenner Barry M, Kamper Annelise, Zucchelli Pietro, et al. Angiotensin-converting enzyme inhibitors and progression of nondiabetic renal disease: a meta-analysis of patient-level data. Annals of Internal Medicine, 135(2):73–87, 2001. [PubMed: 11453706]

[8]. Jafar Tazeen H, Stark Paul C, Schmid Christopher H, Landa Marcia, Maschio Giuseppe, de Jong Paul E, de Zeeuw Dick, Shahinfar Shahnaz, Toto Robert, and Levey Andrew S. Progression of chronic kidney disease: the role of blood pressure control, proteinuria, and angiotensin-converting enzyme inhibition: a patient-level meta-analysis. Annals of Internal Medicine, 139(4):244–252, 2003. [PubMed: 12965979]

[9]. Schmid Christopher H, Stark Paul C, Berlin Jesse A, Landais Paul, and Lau Joseph. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. Journal of Clinical Epidemiology, 57(7):683–697, 2004. [PubMed: 15358396]

[10]. Wright Jackson T Jr, Kusek John W, Toto Robert D, Lee Jeannette Y, Agodoa Lawrence Y, Kirk Katharine A, Randall Otelio S, and Glassock Richard. Design and baseline characteristics of participants in the african american study of kidney disease and hypertension (aask) pilot study. Controlled Clinical Trials, 17(4):S3–S16, 1996.

[11]. Wright Jackson T Jr, Bakris George, Greene Tom, Agodoa Larry Y, Appel Lawrence J, Charleston Jeanne, Cheek DeAnna, Douglas-Baltimore Janice G, Gassman Jennifer, Glassock Richard, et al. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the aask trial. JAMA, 288(19):2421–2431, 2002. [PubMed: 12435255]

[12]. Hsu Chi-yuan, Lin Feng, Vittinghoff Eric, and Shlipak Michael G. Racial differences in the progression from chronic renal insufficiency to end-stage renal disease in the united states. Journal of the American Society of Nephrology, 14(11):2902–2907, 2003. [PubMed: 14569100]

[13]. McClellan William M, Warnock David G, Judd Suzanne, Muntner Paul, Kewalramani Reshma, Cushman Mary, McClure Leslie A, Newsome Britt B, and Howard George. Albuminuria and racial disparities in the risk for esrd. Journal of the American Society of Nephrology, 22(9):1721–1728, 2011. [PubMed: 21868498]

[14]. Mann Johannes FE and Bakris George L. Antihypertensive therapy and progression of nondiabetic chronic kidney disease in adults. https://www.uptodate.com/contents/antihypertensive-therapy-and-progression-of-nondiabetic-chronic-kidney-disease-in-adults, 2019. Accessed: 09/10/2019.

[15]. Eddy David M, Hasselblad Victor, Shachter Ross D, et al. Meta-analysis by the confidence profile method. Academic Press, 1992.

[16]. Manski Charles F. Meta-analysis for medical decisions. Technical report, National Bureau of Economic Research, 2019.

[17]. Lumley Thomas. Network meta-analysis for indirect treatment comparisons. Statistics in Medicine, 21(16):2313–2324, 2002. [PubMed: 12210616]

[18]. Bickel Peter J, Klaassen Chris AJ, Wellner Jon A, and Ritov Ya'acov. Efficient and adaptive estimation for semiparametric models. Johns Hopkins University Press Baltimore, 1993.

[19]. Dahabreh Issa J, Robertson Sarah E, Steingrimsson Jon A, Stuart Elizabeth A, and Hernán Miguel A. Extending inferences from a randomized trial to a new target population. arXiv preprint arXiv:1805.00550, 2019.

[20]. Dahabreh Issa J, Haneuse Sebastien J-PA, Robins James M, Robertson Sarah E, Buchanan Ashley L, Stuart Elisabeth A, and Hernán Miguel A. Study designs for extending causal inferences from a randomized trial to a target population. arXiv preprint arXiv:1905.07764, 2019.

[21]. Dahabreh Issa J, Steingrimsson Jon A, Robertson Sarah E, Petito Lucia C, and Hernán Miguel A. Efficient and robust methods for causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a target population. arXiv preprint arXiv:1908.09230, 2019.

[22]. Rubin Donald B. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5):688, 1974.

[23]. Robins James M and Greenland Sander. Causal inference without counterfactuals: comment. Journal of the American Statistical Association, 95(450):431–435, 2000.

[24]. Dahabreh Issa J, Robins James M, Haneuse Sebastien J-PA, and Hernán Miguel A. Generalizing causal inferences from randomized trials: counterfactual and graphical identification. arXiv preprint arXiv:1906.10792, 2019.

[25]. Hernán Miguel A and VanderWeele Tyler J. Compound treatments and transportability of causal inference. Epidemiology (Cambridge, Mass.), 22(3):368, 2011.

[26]. Westreich Daniel, Edwards Jessie K, Lesko Catherine R, Stuart Elizabeth, and Cole Stephen R. Transportability of trial results using inverse odds of sampling weights. American Journal of Epidemiology, 186(8):1010–1014, 2017. [PubMed: 28535275]

[27]. Pearl Judea and Bareinboim Elias. External validity: from do-calculus to transportability across populations. Statistical Science, 29(4):579–595, 2014.

[28]. Blettner Maria, Sauerbrei Willi, Schlehofer Brigitte, Scheuchenpflug Thomas, and Friedenreich Christine. Traditional reviews, meta-analyses and pooled analyses in epidemiology. International Journal of Epidemiology, 28(1):1–9, 1999. [PubMed: 10195657]

[29]. Hernán Miguel A and Robins James M. Causal inference (forthcoming). Chapman & Hall/CRC, Boca Raton, FL, 2020.

[30]. Delgado Miguel A. Testing the equality of nonparametric regression curves. Statistics & probability letters, 17(3):199–204, 1993.

[31]. Neumeyer Natalie, Dette Holger, et al. Nonparametric comparison of regression curves: an empirical process approach. The Annals of Statistics, 31(3):880–920, 2003.

[32]. Racine Jeffery S, Hart Jeffrey, and Li Qi. Testing the significance of categorical predictor variables in nonparametric regression models. Econometric Reviews, 25(4):523–544, 2006.

[33]. Luedtke Alex, Carone Marco, and van der Laan Mark J. An omnibus non-parametric test of equality in distribution for unknown functions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 81(1):75–99, 2019. [PubMed: 31024219]

[34]. Robins James M. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. Health service research methodology: a focus on AIDS, pages 113–159, 1989.

[35]. Robins James M, Blevins Donald, Ritter Grant, and Wulfsohn Michael. G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of aids patients. Epidemiology, pages 319–336, 1992. [PubMed: 1637895]

[36]. Vansteelandt Stijn, Joffe Marshall, et al. Structural nested models and g-estimation: the partially realized promise. Statistical Science, 29(4):707–731, 2014.

[37]. Tian Lu, Alizadeh Ash A, Gentles Andrew J, and Tibshirani Robert. A simple method for estimating interactions between a treatment and a large number of covariates. Journal of the American Statistical Association, 109(508):1517–1532, 2014. [PubMed: 25729117]

[38]. Stefanski Leonard A and Boos Dennis D. The calculus of m-estimation. The American Statistician, 56(1):29–38, 2002.

[39]. Efron Bradley and Tibshirani Robert J. An introduction to the bootstrap, volume 57 of Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 1994.

[40]. Saul Bradley. geex: An API for M-Estimation, 2018. R package version 1.0.11.

[41]. McCullagh Peter and Nelder John A. Generalized linear models, volume 37 of Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 1989.

[42]. Lunceford Jared K and Davidian Marie. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in Medicine, 23(19):2937–2960, 2004. [PubMed: 15351954]

[43]. Robins James M, Rotnitzky Andrea, and Scharfstein Daniel O. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In Statistical models in epidemiology, the environment, and clinical trials, pages 1–94. Springer, 2000.

[44]. Dahabreh Issa J, Robins James M, Haneuse Sebastien J-PA, Saeed Iman, Robertson Sarah E, Stuart Elisabeth A, and Hernán Miguel A. Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population. arXiv preprint arXiv:1905.10684, 2019.

[45]. Lewis Steff and Clarke Mike. Forest plots: trying to see the wood and the trees. BMJ, 322(7300):1479–1480, 2001. [PubMed: 11408310]

[46]. Bareinboim Elias and Pearl Judea. Meta-transportability of causal effects: A formal approach. In Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS), pages 135–143, 2013.

[47]. Bareinboim Elias and Pearl Judea. Causal transportability with limited experiments. In Twenty-Seventh AAAI Conference on Artificial Intelligence, 2013.

[48]. Robins James M and Rotnitzky Andrea. Comments. Statistica Sinica, 11(4):920–936, 2001.

[49]. Sobel Michael, Madigan David, and Wang Wei. Causal inference for meta-analysis and multi-level data structures, with application to randomized studies of Vioxx. Psychometrika, 82(2):459–474, 2017. [PubMed: 27388291]

[50]. Kabali Conrad and Ghazipura Marya. Transportability in network meta-analysis. Epidemiology, 27(4):556–561, 2016. [PubMed: 26963292]

[51]. Schnitzer Mireille E, Steele Russell J, Bally Michèle, and Shrier Ian. A causal inference approach to network meta-analysis. Journal of Causal Inference, 4(2), 2016.

[52]. Robins James M and Ritov Ya'acov. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. Statistics in Medicine, 16(3):285–319, 1997. [PubMed: 9004398]

[53]. Bang Heejung and Robins James M. Doubly robust estimation in missing data and causal inference models. Biometrics, 61(4):962–973, 2005. [PubMed: 16401269]

[54]. DerSimonian Rebecca and Laird Nan. Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3):177–188, 1986. [PubMed: 3802833]

[55]. Robins James M. Confidence intervals for causal parameters. Statistics in Medicine, 7(7):773–785, 1988. [PubMed: 3406605]

[56]. Robins James M, Hernán Miguel A, and Brumback Babette. Marginal structural models and causal inference in epidemiology. Epidemiology (Cambridge, Mass.), 11(5):550–560, 2000.

[57]. Robins James M. Marginal structural models versus structural nested models as tools for causal inference. In Statistical models in epidemiology, the environment, and clinical trials, pages 95–133. Springer, 2000.
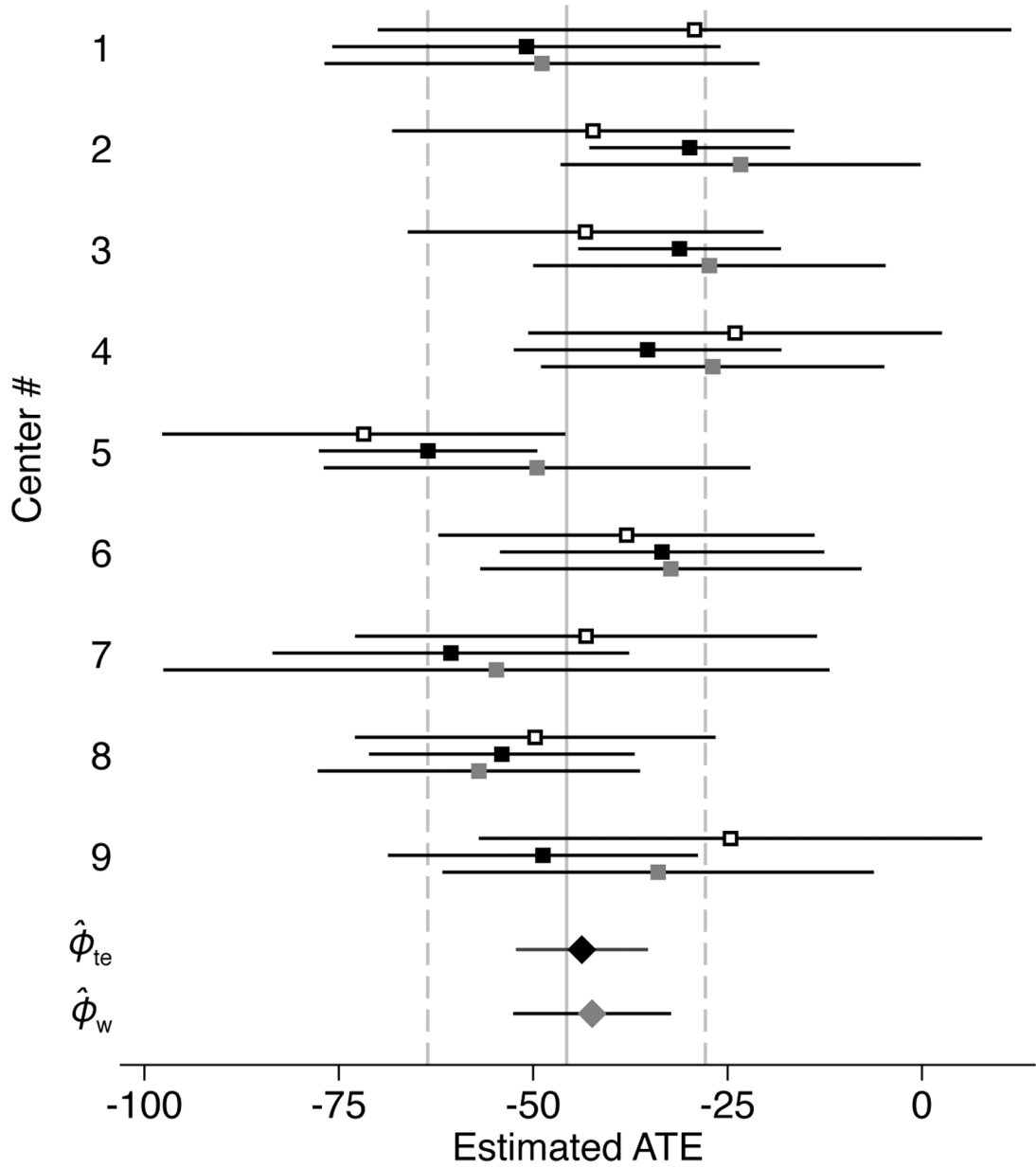
Figure 1:

Forest plot summarizing analyses using the multi-center HALT-C trial data to emulate a meta-analysis.

Point estimates (markers) and 95% confidence intervals (extending lines) from analyses using the multi-center HALT-C trial data to emulate a meta-analysis. White squares, □ = unadjusted analyses that use data only from each center $s \in \mathcal{S} = \{1, ..., 9\}$; black squares, ■ = transportability analyses using conditional outcome mean modeling and estimator (6) to transport inferences from each center $S = s, s \in \{1, ..., 9\}$ to $S = 0$; gray squares, ■ = transportability analyses using inverse odds weighting and estimator (7) to transport inferences from each center $S = s, s \in \{1, ..., 9\}$ to $S = 0$; diamonds denote analyses transporting inferences from the collection of centers $\mathcal{S} = \{1, ..., 9\}$ to $S = 0$: black diamond,

◆ = treatment effect estimator (8), $\hat{\phi}_{\text{te}}(1,0)$; gray diamond, ◆ = weighting estimator (9), $\hat{\phi}_{\text{w}}(1,0)$. The solid vertical line indicates the point estimate and the dashed vertical lines indicate the limits of the confidence interval from the randomization-based analysis using only data from $S = 0$.

Table 1:

Results from analyses using the multi-center HALT-C trial data to emulate a meta-analysis.

S	Sample size	Unadjusted	Treatment effect modeling	Weighting
1	48	−29.3 (−70.0, 11.5)	−50.9 (−75.8, −25.9)	−48.9 (−76.8, −20.9)
2	98	−42.3 (−68.1, −16.4)	−29.9 (−42.8, −17.0)	−23.3 (−46.5, −0.2)
3	133	−43.3 (−66.1, −20.4)	−31.2 (−44.2, −18.2)	−27.4 (−50.0, −4.7)
4	69	−24.0 (−50.6, 2.6)	−35.3 (−52.5, −18.1)	−26.9 (−49.0, −4.8)
5	77	−71.8 (−97.7, −45.9)	−63.5 (−77.6, −49.5)	−49.5 (−76.9, −22.1)
6	110	−38.0 (−62.2, −13.8)	−33.4 (−54.3, −12.6)	−32.3 (−56.8, −7.8)
7	94	−43.2 (−72.9, −13.5)	−60.6 (−83.5, −37.7)	−54.7 (−97.6, −11.9)
8	100	−49.7 (−72.9, −26.6)	−54.0 (−71.1, −37.0)	−57.0 (−77.7, −36.3)
9	43	−24.6 (−57.0, 7.7)	−48.8 (−68.7, −28.8)	−33.9 (−61.7, −6.2)
0	202	−45.7 (−63.6, −27.9)	−43.7 (−52.2, −35.2)	−42.4 (−52.6, −32.3)

Point estimates and 95% confidence intervals (in parentheses) from analyses using the multi-center HALT-C trial data to emulate a meta-analysis. Unadjusted results are obtained using the crude mean difference in each trial for all rows. Treatment effect modeling results are obtained using the estimator of equation (6) in rows 1 through 9 and estimator (8) in the bottom row; weighting results are obtained using the estimator of equation (7) in rows 1 through 9 and estimator (9) in the bottom row. In rows 1 through 9, for each trial unadjusted results use only treatment and outcome data from each trial $s \in \mathcal{S}$; transportability analyses use covariate, treatment, and outcome data from each trial $s \in \mathcal{S}$ and baseline covariate data from $S = 0$. In the bottom row, the unadjusted analysis uses only treatment and outcome data from the center $S = 0$; transportability analyses use covariate, treatment, and outcome data from all trials in the collection $\mathcal{S}$ and baseline covariate data from $S = 0$.