A Unified Framework for the Transportability of Population-Level Causal Measures

Ahmed Boughdiri
Premedical Inria
Univ. Montpellier, France
ahmed.boughdiri@inria.fr

Clément Berenfeld
Premedical Inria
Univ. Montpellier, France
clement.berenfeld@inria.fr

Julie Josse
Premedical Inria
Univ. Montpellier, France
julie.josse@inria.fr

Erwan Scornet
Sorbonne Université and Université Paris Cité, CNRS
Laboratoire de Probabilités, Statistique et Modélisation
F-75005 Paris, France
erwan.scornet@sorbonne-universite.fr

Abstract

Generalization methods offer a powerful solution to one of the key drawbacks of randomized controlled trials (RCTs): their limited representativeness. By enabling the transport of treatment effect estimates to target populations subject to distributional shifts, these methods are increasingly recognized as the future of meta-analysis, the current gold standard in evidence-based medicine. Yet most existing approaches focus on the risk difference, overlooking the diverse range of causal measures routinely reported in clinical research. Reporting multiple effect measures—both absolute (e.g., risk difference, number needed to treat) and relative (e.g., risk ratio, odds ratio)—is essential to ensure clinical relevance, policy utility, and interpretability across contexts. To address this gap, we propose a unified framework for transporting a broad class of first-moment population causal effect measures under covariate shift. We provide identification results for both continuous and binary outcome under two conditional exchangeability assumptions, derive both classical and semiparametric estimators, and evaluate their performance through theoretical analysis, simulations, and real-world applications. Our analysis shows the specificity of different causal measures and thus the interest of studying them all: for instance, two common approaches (one-step, estimating equation) lead to similar estimators for the risk difference but to two distinct estimators for the odds ratio.

1 Introduction

Generalization methods [33, 43, 12, 10, 14] have emerged as a powerful response to the restricted external validity [36] of RCTs: due to stringent inclusion and exclusion criteria, RCT populations often exclude key segments of the real-world patient population—such as individuals with comorbidities, pregnant women, or other vulnerable groups—resulting in trial samples that are poorly representative. Consequently, the findings of many RCTs may lack relevance for broader clinical or policy applications. Generalization techniques address this gap by exploiting treatment effect heterogeneity—that is, the fact that treatment efficacy can vary systematically with patient characteristics. By adjusting for

Code & Python library: `Causalipy`. `https://github.com/BoughdiriAhmed/Causalipy`.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

differences in the distribution of these covariates between the trial population and a target population, these methods can estimate treatment effects beyond the original study population. This is especially valuable given the high costs, long timelines, and operational complexity of conducting new trials. As such, generalization approaches are increasingly viewed as a pivotal step toward rethinking the role of clinical trials in modern evidence generation. The implications are far-reaching. For instance, when drug reimbursement decisions are partly tied to estimated real-world efficacy [17], the ability to predict treatment benefits across diverse populations could influence pricing, access, and healthcare policy. Moreover, recent works suggest that generalization methods may also redefine the role of meta-analysis [11, 37, 20], traditionally viewed as the top of the pyramid of evidence-based medicine.

Most generalization methods are dedicated to estimating the average treatment effect (ATE) via the risk difference (RD), owing to its linearity and analytical convenience. Yet this focus is incomplete. Clinical guidelines and regulatory bodies explicitly recommend reporting both absolute and relative effect measures [39, 31], as they capture complementary aspects of treatment impact. Among absolute measures, the RD remains standard, but the number needed to treat (NNT), a direct, clinically intuitive transformation of the RD, offers additional interpretability [29]. On the relative scale, measures such as the risk ratio (RR), widely used in public health research [28], and the odds ratio (OR), very popular in epidemiology [18], play a critical role in framing treatment efficacy. Presenting multiple causal estimands is not simply recommended—it is essential. A treatment effect that appears homogeneous on one scale may reveal heterogeneity on another, a phenomenon that may seem counterintuitive but carries significant implications for personalized decision-making. Furthermore, the perceived magnitude of benefit can shift dramatically depending on the baseline risk and the effect measure used. Consider a treatment that reduces mortality from 3% to 1%: the RD of 0.02 may suggest a minor effect, yet the RR reveals a threefold increase in risk for the untreated, reframing the impact as clinically substantial. This striking contrast illustrates how the choice of causal measure directly shapes interpretation and ultimately, policy and clinical decisions.

Contributions. In Section 2, we present a unified framework for transporting a broad class of *first-moment population causal effect measures*, defined as functionals of the expectations of potential outcomes. This class, composed of more than a dozen widely-used estimands, includes collapsible (RD, RR, etc.) and non-collapsible measures (odds ratio-OR, NNT etc.) the latter posing unique generalization challenges. Building on this formalism, we develop generic identification strategies under two key assumptions: (i) exchangeability in mean (Section 3), and (ii) exchangeability in effect measure (Section 4). The latter assumption, though weaker, requires access to control outcomes in the target population—a condition met, for instance, when treatment has not yet been deployed. Crucially, our identification results extend to non-collapsible measures even under this weaker condition, which to our knowledge has not been previously derived. Within each setting, we derive two broad families of estimators. The first approach adapts classical methods—such as weighting (Horvitz-Thompson) and regression-based strategies (G-formula)—for which we establish asymptotic properties and derive closed-form variance expressions. To the best of our knowledge, no prior work has formally studied these properties using density ratio estimation. The second leverages semiparametric efficiency theory to construct new doubly-robust estimators using either one-step or estimating equation approaches. While these often coincide in the linear case (e.g., RD), we highlight that for nonlinear measures such as the RR or OR, they diverge—leading to fundamentally different estimators. Finally, in Section 5 we conduct an extensive empirical evaluation of all estimators on synthetic and a real-world dataset.

Related work. Most generalization work focused on RD. While the idea of weighting randomized controlled trials (RCTs) is not new, using external data can be traced back to the foundational work of [5]. [14, 10] provided a comprehensive survey of generalization techniques, and [6] contributed additional consistency results for the main classes of estimators: regression-based, weighting-based, and hybrid approaches. [8] further investigated the inverse probability of sampling weighting (IPSW) estimator, deriving finite-sample bias and variance, as well as an upper bound on its risk under the assumption that the covariates X are categorical. Other studies have addressed the generalization problem by modeling the outcome directly [27, 13]. [11] extended this line of research to settings involving multiple source datasets (i.e., several RCTs) and a set of covariates drawn from a distinct target distribution. They position this framework as a natural evolution of traditional meta-analysis, highlighting its potential to unify evidence across studies. Although the present work focuses on a single RCT and a single target population, all of our results seamlessly extend to the multi-RCT case. This generality allows us to offer a compelling alternative to classical meta-analyses—one

that supports the generalization of both absolute and relative treatment effect measures in a unified framework. We also highlight recent works on the analysis of externally controlled single-arm trials and hybrid trials [2, 30], which aim to enhance the precision of RCT-based estimands by incorporating external information. Similarly, [15] incorporate observational data to RCTs using the prediction-powered inference framework. While these approaches address distinct inferential questions, they share structural similarities with ours in using auxiliary data to improve estimation. Another line of work has focused on estimating alternative causal effect measures beyond the risk difference. For instance, [1] proposes several strategies for estimating the risk ratio (RR) in observational settings, without addressing the generalization to a target population. [7] examined several key properties (e.g., collapsibility) of different causal effect measures (e.g., RD, RR, and OR) to identify which estimands are less sensitive to distributional shifts. While their work emphasizes which causal estimands are easier to identify and generalize, ours provides a unified framework that generalizes all first-moment causal measures, focusing on estimation strategies and efficiency for practical deployment with theoretical guarantees. Their focus is on distinguishing treatment effects from baselines; ours highlights the shared structure among estimands, enabling unified identification and estimation. [34, 44, 40] introduce methods for estimating conditional risk ratios, including approaches based on causal forests. Yet, it is crucial to emphasize that due to the non-linearity of relative measures, estimating the CATE does not directly yield the ATE.

2 Problem setting: notation and identification assumptions

Following the potential outcomes framework [38, 41], we consider random variables $(X, S, A, Y^{(0)}, Y^{(1)})$, where $X \in \mathbb{R}^p$ denotes patient covariates, $S \in \{0, 1\}$ indicates sample membership ($S = 1$ for the source population and $S = 0$ for the target population), and $A \in \{0, 1\}$ denotes treatment assignment ($A = 1$ if treatment is administered and $A = 0$ otherwise). Potential outcomes $Y^{(0)}$ and $Y^{(1)}$ represent outcomes under control and treatment respectively, of which only one is observed per individual depending on the assigned treatment, yielding the observed outcome

$$Y = AY^{(1)} + (1 - A)Y^{(0)}.$$

We observe data from two distinct populations: a Randomized Controlled Trial (RCT) dataset $(X_i, A_i, Y_i)_{i\in[n]}$ from the source population, where treatment assignments and outcomes are recorded, and a target dataset $(X_i)_{i\in[m]}$, where only covariates are observed. This reflects the practical constraint that treatment and outcome data are collected only for individuals enrolled in the trial, while covariate information is available for both populations. We model this two datasets as stemming from one probability distribution P_{obs} and we observe a N-sample:

$$(Z_i)_{i\in[N]} := (S_i, X_i, S_iA_i, S_iY_i)_{i\in[N]} \sim P_{\text{obs}}^{\otimes N},$$

Let α be the Bernoulli parameter of the random variable S. Thus, (n, m) follow a multinomial distribution with parameter N and probability $(\alpha, 1 - \alpha)$. Let P_{S} and P_{T} denote the conditional distributions of Z given $S = 1$ and $S = 0$, respectively, with corresponding expectations $\mathbb{E}_{\text{S}}[\cdot]$ and $\mathbb{E}_{\text{T}}[\cdot]$.

2.1 Causal estimands of interest

Causal effect measures, as formalized by Pearl [32], are expressed in terms of the joint distribution of potential outcomes. Individual-level causal effects rely on this joint distribution, which is generally not identifiable from observed data. Therefore, in this paper, we focus on a specific subclass: the first-moment population causal measures [16]. This class includes commonly employed estimands—such as RD, RR, and OR—and less frequently used quantities, including the Switch Relative Risk [22], Excess Risk Ratio (ERR) [4], Survival Ratio (SR), and Relative Susceptibility (RS), and Log Odds Ratio (log-OR), see Appendix A for a detailed enumeration of measures falling into this class.

Definition 1 (First moment population causal measures)**.** *Let P denote the joint distribution of potential outcomes $(Y^{(0)}, Y^{(1)})$. The quantity τ^P is a first moment population causal measure if there exists an* effect measure $\Phi : D_\Phi \to \mathbb{R}$*, with $D_\Phi \subset \mathbb{R}^2$, such that for all distributions P with $(\mathbb{E}_P[Y^{(0)}], \mathbb{E}_P[Y^{(1)}]) \in D_\Phi$,*

$$\tau^P := \Phi\left(\mathbb{E}_P[Y^{(1)}], \mathbb{E}_P[Y^{(0)}]\right). \tag{1}$$

We require that for all $\psi_0 \in \mathbb{R}$, the map $\psi_1 \mapsto \Phi(\psi_1, \psi_0)$ is injective on its definition domain. It's inverse, when it exists, is denoted by $\Gamma(\cdot, \psi_0)$ and is called the effect function*. This definition extends to subpopulations by conditioning on covariates X, yielding, when it exists, the conditional effect*

$$\tau^P(X) := \Phi\left(\mathbb{E}_P[Y^{(1)}|X], \mathbb{E}_P[Y^{(0)}|X]\right). \tag{2}$$

Example 1. For these well-known causal measures, the effect measures and functions are given by:

Measure	**Effect Measure Φ**	**Effect Function Γ**
Risk Difference (RD)	$\Phi(\psi_1, \psi_0) = \psi_1 - \psi_0$	$\Gamma(\tau, \psi_0) = \psi_0 + \tau$
Risk Ratio (RR)	$\Phi(\psi_1, \psi_0) = \frac{\psi_1}{\psi_0}$	$\Gamma(\tau, \psi_0) = \tau \cdot \psi_0$
Odds Ratio (OR)	$\Phi(\psi_1, \psi_0) = \frac{\psi_1}{1-\psi_1} \cdot \frac{1-\psi_0}{\psi_0}$	$\Gamma(\tau, \psi_0) = \frac{\tau \cdot \psi_0}{1+\tau \cdot \psi_0 - \psi_0}$

The existence of Γ ensures that for a fixed baseline, distinct treatment responses yield different causal measures. In the following, we use τ_Φ^P to denote any first moment population causal measure evaluated under distribution P. Our objective is to estimate $\tau_\Phi^{\mathrm{T}} := \Phi\left(\mathbb{E}_{\mathrm{T}}[Y^{(1)}], \mathbb{E}_{\mathrm{T}}[Y^{(0)}]\right)$, for all effect measures Φ, where the expectations are taken with respect to the target population distribution.

2.2 Identification assumptions

Because the effect τ_Φ^{T} is defined in terms of potential outcomes in the target population, it is not directly identifiable from trial data alone. The core difficulty lies in the fact that while treatment assignment in the trial is randomized, the trial sample may not be representative of the broader target population. To ensure identification of the estimand from observed data, we introduce standard causal inference assumptions (Assumption 1) and assume covariate overlap (Assumption 2).

Assumption 1 (Trial's Internal Validity)**.** *The RCT is assumed to be internally valid, that is:*

A. ***Ignorability:*** $(Y^{(1)}, Y^{(0)}) \perp\!\!\!\perp A \mid S = 1$*;*
B. ***Stable Unit Treatment Value Assumption (SUTVA):*** $Y = AY^{(1)} + (1-A)Y^{(0)}$*;*
C. ***Positivity and Randomized Assignment:*** $A \sim \mathcal{B}(\pi)$*, where* $0 < \pi < 1$ *(typically* $\pi = 0.5$*).*

Assumption 2 (Overlap)**.** *For all $x \in \mathrm{supp}(P_{\mathrm{T}})$, we have $\mathbb{P}(S = 1 \mid X = x) > 0$.*

3 Generalization under exchangeability in mean

In addition to internal validity and covariate overlap, identification under this setting requires a transportability condition that links the distribution of potential outcomes between the trial and target populations. This condition can be expressed as conditional exchangeability in mean, which states that, conditional on covariates, the average potential outcomes are the same across populations.

Assumption 3 (Exchangeability in mean)**.** *For all $x \in \mathrm{supp}(P_{\mathrm{S}}) \cap \mathrm{supp}(P_{\mathrm{T}})$ and for all $a \in \{0, 1\}$, we have $\mu_{(a)}^{\mathrm{S}}(x) = \mu_{(a)}^{\mathrm{T}}(x)$ where $\mu_{(a)}^P(x) = \mathbb{E}_P[Y^{(a)} \mid X = x]$.*

Leveraging observations from the randomized controlled trial and under Assumption 1 to 3, three identification formulas that express $\mathbb{E}_{\mathrm{T}}[Y^{(a)}]$ in terms of source population quantities can be derived:

$$\mathbb{E}_{\mathrm{T}}[Y^{(a)}] = \mathbb{E}_{\mathrm{T}}\left[\mathbb{E}_{\mathrm{S}}[Y^{(a)}|X]\right] \qquad \textit{(transporting conditional outcomes)} \tag{3}$$

$$= \mathbb{E}_{\mathrm{S}}\left[\frac{P_{\mathrm{T}}(X)}{P_{\mathrm{S}}(X)} Y^{(a)}\right] \qquad \textit{(weighting outcomes)} \tag{4}$$

$$= \mathbb{E}_{\mathrm{S}}\left[\frac{P_{\mathrm{T}}(X)}{P_{\mathrm{S}}(X)} \mathbb{E}_{\mathrm{S}}[Y^{(a)}|X]\right] \qquad \textit{(weighting conditional outcomes)}, \tag{5}$$

see Appendix B.1 for details. In the next section, we present several estimators of any first moment population causal measure, based on the three identification formulas above.

3.1 Weighting and regression strategies under exchangeability in mean

The Horvitz–Thompson estimator [21] is probably one of the most simple estimators to use in a RCT. Based on equation (4), we construct the weighted Horvitz-Thompson estimator. The following assumption is also required for technical reasons.

Assumption 4. *For $a \in \{0,1\}$, we assume that $Y^{(a)}$ and $P_{\mathrm{T}}(X)Y^{(a)}/P_{\mathrm{S}}(X)$ are square-integrable.*

Definition 2 (Weighted Horvitz-Thompson)**.** *For any first moment population causal measure Φ, we define the weighted Horvitz-Thompson estimator $\widehat{\tau}_{\Phi,wHT}$ as follows:*

$$\widehat{\tau}_{\Phi,wHT} = \Phi\left(\frac{1}{n}\sum_{S_i=1}\widehat{r}(X_i)\frac{A_iY_i}{\pi}, \frac{1}{n}\sum_{S_i=1}\widehat{r}(X_i)\frac{(1-A_i)Y_i}{1-\pi}\right), \tag{6}$$

where $\widehat{r}(X)$ is any estimator of the density ratio between the target and source covariate distributions.

When $\widehat{r}$ is replaced by the oracle quantity r, we show that $\widehat{\tau}_{\Phi,\mathrm{wHT}}$ is asymptotically normal and an unbiased estimator of τ^{T}_{Φ} (see Proposition 8 in Appendix B.2). Since

$$r(X) = \frac{P_{\mathrm{T}}(X)}{P_{\mathrm{S}}(X)} = \frac{\mathbb{P}(S=1)\,\mathbb{P}(S=0|X=x)}{\mathbb{P}(S=0)\,\mathbb{P}(S=1|X=x)},$$

one can estimate $r(X)$ by estimating $\mathbb{P}(S=1 \mid X=x)$ via a logistic regression model. By doing so, we avoid imposing a parametric model, such as a Gaussian distribution, on the distribution of X [see also 24]. In this context, the M-estimation framework [42] can then be applied to derive asymptotic variances for this estimator.

Logistic model. *We assume $\mathbb{E}[XX^\top]$ is positive definite, X is Sub-Gaussian, and $\exists\beta_\infty = (\beta^0_\infty, \beta^1_\infty) \in \mathbb{R}^{p+1}$ s.t. $\mathbb{P}(S=1|X) = \sigma(X,\beta_\infty) = \{1+\exp(-X^\top\beta^1_\infty - \beta^0_\infty)\}^{-1}$, a.s.*

Proposition 1 (asymptotic normality of weighted Horvitz-Thompson estimator)**.** *Grant Assumption 1 to 4. Let $\widehat{\beta}_N$ denote the maximum likelihood estimate (MLE) obtained from logistic regression of the selection indicator S on covariates X. Define the estimated density ratio as*

$$r(x,\widehat{\beta}_N) = \frac{n}{N-n}\cdot\frac{1-\sigma(x,\widehat{\beta}_N)}{\sigma(x,\widehat{\beta}_N)}, \qquad \textit{with} \quad n = \sum_{i=1}^{N} S_i. \tag{7}$$

Under the logistic model, the Horvitz-Thompson estimator $\widehat{\tau}_{\Phi,HT}$ weighted by the estimated ratio $r(x,\widehat{\beta}_N)$ is asymptotically unbiased and satisfies $\sqrt{N}\left(\widehat{\tau}_{\Phi,wHT} - \tau^{\mathrm{T}}_{\Phi}\right) \xrightarrow{d} \mathcal{N}(0, V_{\Phi,HT})$.

Appendix B.3 provides full proofs and asymptotic variances for all first moment population causal measures, including results for the Neyman estimator with estimated treatment probabilities—results that, to our knowledge, are novel even for the RD.

Alternatively, using the identification results in equations (3) and (5), one can adapt the regression-based approach of Robins [35] to derive weighted or transported G-formula estimators. While the transported version appears in Dahabreh et al. [12] for the RD, the weighted version, to the best of our knowledge, has not been previously presented.

Definition 3 (Weighted/Transported G-formula)**.** *For any first moment population causal measure Φ, we define the weighted G-formula estimator $\widehat{\tau}_{\Phi,\mathrm{wG}}$ and the transported G-formula estimator $\widehat{\tau}_{\Phi,\mathrm{tG}}$ as:*

$$\widehat{\tau}_{\Phi,\mathrm{wG}} = \Phi\left(\frac{1}{n}\sum_{S_i=1}\widehat{r}(X_i)\,\widehat{\mu}^{\mathrm{S}}_{(1)}(X_i), \frac{1}{n}\sum_{S_i=1}\widehat{r}(X_i)\,\widehat{\mu}^{\mathrm{S}}_{(0)}(X_i)\right), \tag{8}$$

$$\widehat{\tau}_{\Phi,\mathrm{tG}} = \Phi\left(\frac{1}{N-n}\sum_{S_i=0}\widehat{\mu}^{\mathrm{S}}_{(1)}(X_i), \frac{1}{N-n}\sum_{S_i=0}\widehat{\mu}^{\mathrm{S}}_{(0)}(X_i)\right), \tag{9}$$

where $\widehat{r}, \widehat{\mu}^{\mathrm{S}}_{(1)}, \widehat{\mu}^{\mathrm{S}}_{(0)}$ are any estimators of $r, \mu^{\mathrm{S}}_{(1)}, \mu^{\mathrm{S}}_{(0)}$.

Using the M-estimation framework [42], one can derive asymptotic variance estimates for both G-formula estimators. Furthermore, assuming a linear model for the outcomes, we prove that regression adjustment leads to a lower asymptotic variance compared to the weighted Horvitz-Thompson.

Linear model 1. *For all $a, s \in \{0,1\}$, let $Y_s^{(a)} = V^\top \beta_s^{(a)} + \epsilon_s^{(a)}$, where $V^\top = [1, X^\top]$ with $\mathbb{E}[\epsilon_s^{(a)} \mid X] = 0$ and Var$(\epsilon_s^{(a)} \mid X) = \sigma^2$.*

Proposition 2 (Asymptotic Normality of weighted and Transported G-formula Estimators)**.** *Let $\widehat{\tau}_{\Phi,\mathrm{tG}}$ and $\widehat{\tau}_{\Phi,\mathrm{wG}}$ denote respectively the weighted G-formula where $\widehat{r}$ is a logistic regression estimator* (7) *and $\widehat{\mu}^{\mathrm{S}}_{(1)}, \widehat{\mu}^{\mathrm{S}}_{(0)}$ are ordinary least squares estimator. Then, under Assumptions 1 to 4,*

$$\sqrt{N}\left(\widehat{\tau}_{\Phi,\mathrm{wG}} - \tau^{\mathrm{T}}_{\Phi}\right) \xrightarrow{d} \mathcal{N}\left(0, V^{OLS}_{\Phi,\mathrm{wG}}\right), \qquad \sqrt{N}\left(\widehat{\tau}_{\Phi,\mathrm{tG}} - \tau^{\mathrm{T}}_{\Phi}\right) \xrightarrow{d} \mathcal{N}\left(0, V^{OLS}_{\Phi,\mathrm{tG}}\right).$$

Besides, $V^{OLS}_{\Phi,\mathrm{tG}} \leq V^{OLS}_{\Phi,\mathrm{wG}} \leq V_{\Phi,HT}$, where all variances are explicit Appendix B.4.

Given these results, one might wonder whether it's better to avoid estimating the density ratio altogether, as the transported G-formula is more efficient under correct specification. However, when models are misspecified, weighting the outcomes may perform better. This motivates doubly robust estimators, which retain consistency if either model is correctly specified.

3.2 Semiparametric efficient estimators under exchangeability in mean

A common approach to constructing doubly robust estimators relies on semiparametric efficiency theory. By deriving the efficient influence function (EIF) of the target parameter [see, e.g. 25], one can build estimators that are not only robust to mispecification, but also achieve the lowest possible asymptotic variance among unbiased estimators. We denote by $\varphi_\Phi(Z, \eta, \psi_1^{\mathrm{T}}, \psi_0^{\mathrm{T}})$ the EIF of τ_Φ, which depends on (i) the nuisance parameters $\eta = (\mu_{(0)}, \mu_{(1)}, r)$ and (ii) the values of ψ_1^{T} and ψ_0^{T}.

Based on estimators $\widehat{\eta}, \widehat{\psi}_1^{\mathrm{T}}, \widehat{\psi}_0^{\mathrm{T}}$, one can use the EIF framework via one of the following two techniques to build new estimators of τ_Φ^{T} whose properties are described below:

(i) **One-step estimators** $\widehat{\tau}_\Phi^{\mathrm{OS}}$ consist in applying a first-order bias correction to an initial plug-in estimator $\widehat{\tau}_\Phi^{\mathrm{T}} = \Phi(\widehat{\psi}_1^{\mathrm{T}}, \widehat{\psi}_0^{\mathrm{T}})$, resulting in $\widehat{\tau}_\Phi^{\mathrm{OS}} = \widehat{\tau}_\Phi^{\mathrm{T}} + \frac{1}{N}\sum_{i=1}^N \varphi_\Phi(Z_i, \widehat{\eta}, \psi_1^{\mathrm{T}}, \widehat{\psi}_0^{\mathrm{T}})$.

(ii) **Estimating equation estimators** are obtained by setting the empirical mean of the EIF to zero. This amounts to finding an estimators $\widehat{\tau}_\Phi^{\mathrm{EE}} = \Phi(\widehat{\psi}_1^{\mathrm{T}}, \widehat{\psi}_0^{\mathrm{T}})$ such that $\widehat{\psi}_1^{\mathrm{T}}, \widehat{\psi}_0^{\mathrm{T}}$ are solutions to $\sum_{i=1}^N \varphi_\Phi(Z_i, \widehat{\eta}, \widehat{\psi}_1^{\mathrm{T}}, \widehat{\psi}_0^{\mathrm{T}}) = 0$ (Estimating Equation).

In practice and in both case, it is usual to resort to crossfitted techniques [3] by estimating $\widehat{\tau}_\Phi$ and $\widehat{\eta}$ and evaluating the EIF φ_Φ on two different datasets to enforce independence. We drop the T superscript in the rest of this section for the sake of clarity. Both above approaches require knowing the EIF φ_Φ. As $\tau_\Phi = \Phi(\psi_1, \psi_0)$, letting φ_1 (resp. φ_0) be the EIF of ψ_1 (resp. ψ_0), standard calculations on EIF functions (see e.g. [26, Sec 3.4.3]) show that

$$\varphi_\Phi(Z, \eta, \psi_1, \psi_0) = \partial_1\Phi(\psi_1, \psi_0)\varphi_1(Z, \eta, \psi_1) + \partial_0\Phi(\psi_1, \psi_0)\varphi_0(Z, \eta, \psi_0).$$

With this equality, and the following proposition, one can compute the influence function φ_Φ.

Proposition 3. *Grant Assumption 1 to 3. For all $a \in \{0,1\}$, we have*

$$\varphi_a(Z, \eta, \psi_a) := \frac{1-S}{1-\alpha}(\mu_{(a)}(X) - \psi_a) + \frac{S\mathbb{1}\{A=a\}}{\alpha P(A=a)} r(X)(Y - \mu_{(a)}(X)).$$

The exact expressions of one-step and estimating equation estimators depend on φ_Φ, which in turns depends on the effect measure Φ. While the first approach leads to explicit expressions, estimating equation estimators are defined implicitly through the estimating equation in (ii), which writes

$$\partial_1\Phi(\widehat{\psi}_1, \widehat{\psi}_0)\frac{1}{N}\sum_{i=1}^N \varphi_1(Z, \widehat{\eta}, \widehat{\psi}_1) + \partial_0\Phi(\widehat{\psi}_1, \widehat{\psi}_0)\frac{1}{N}\sum_{i=1}^N \varphi_0(Z, \widehat{\eta}, \widehat{\psi}_0) = 0. \tag{10}$$

One way — though not the only — to satisfy (10) is to find two estimators $\widehat{\psi}_1, \widehat{\psi}_0$ that cancels the empirical mean of their EIF (second and fourth term in (10)), that is the corresponding EE estimator for ψ_1 and ψ_0. We end up with a plug-in estimator of the form $\Phi(\widehat{\psi}_1^{\mathrm{EE}}, \widehat{\psi}_0^{\mathrm{EE}})$, whose precise expression is given below.

Proposition 4 (Estimating equation estimators.)**.** *Given estimators $\widehat{\mu}_{(a)}$ (resp. $\widehat{r}$) of $\mu_{(a)}$ (resp. r), an estimating equation estimator $\widehat{\tau}_\Phi^{\mathrm{EE}}$ of τ_Φ is given by $\widehat{\tau}_\Phi^{\mathrm{EE}} = \Phi(\widehat{\psi}_1^{\mathrm{EE}}, \widehat{\psi}_0^{\mathrm{EE}})$ where for all $a \in \{0,1\}$*

$$\widehat{\psi}_a^{\mathrm{EE}} := \frac{1}{m}\sum_{S_i=0} \widehat{\mu}_{(a)}(X_i) + \frac{1-\alpha}{m\alpha}\sum_{S_i=1} \frac{\mathbb{1}\{A=a\}}{\mathbb{P}(A=a)}\widehat{r}(X_i)(Y - \widehat{\mu}_{(a)}(X_i)). \tag{11}$$

A similar estimator already appears in [12]. We prove below that $\widehat{\tau}_\Phi^{\mathrm{EE}}$ is *(weakly) doubly-robust*.

Proposition 5. *Let $\widehat{\mu}_{(a)}$ and $\widehat{r}$ be two estimators independent from $Z_1, \ldots, Z_n$. Then, under boundedness of $\widehat{\mu}^{(a)}$, $\widehat{r}$ and Y, and assuming that Φ is continuous, the estimator $\widehat{\tau}_\Phi^{\mathrm{EE}} = \Phi(\widehat{\psi}_1^{\mathrm{EE}}, \widehat{\psi}_0^{\mathrm{EE}})$ is consistent as soon as either $\widehat{\mu}_{(a)} = \mu$ for all $a \in \{0,1\}$ or $\widehat{r} = r$.*

It would be easy to extend this result to a *strong* robustness property, meaning that $\sqrt{N}(\widehat{\tau}_\Phi^{\mathrm{EE}} - \tau_\Phi^{\mathrm{T}}) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}\,\varphi_\Phi(Z))$ under mild convergence requirements of $\widehat{\mu}_{(a)}$ and $\widehat{r}$ towards $\mu_{(a)}$ and r. Concerning the OS estimators, we cannot derive any formal robustness results in full generality, and this property needs to be assessed on a case-by-case basis. For instance, [1] shows that the OS estimator for the RR is doubly robust based on usual requirements on the estimation of the nuisance parameters and some extra assumptions on $\widehat{\psi}_0^{\mathrm{T}}$. In light of these observations, we recommend using EE estimators rather than OS estimators, when possible, as they also usually lead to better results empirically, see Figure 1.

When Φ is a linear functional (e.g., RD), the two approaches, one-step and estimating equations yields the same estimators, up to a scaling term depending on α and N. For nonlinear functionals however (such as the RR or OR), the estimators generally differ (see below and Appendix B.6).

(RD) For the risk difference, the estimating equation approach yields $\widehat{\tau}_{\mathrm{RD}}^{\mathrm{EE}} = \widehat{\psi}_1^{\mathrm{EE}} - \widehat{\psi}_0^{\mathrm{EE}}$. The one step approach, based on initial estimators $\widehat{\psi}_1$ and $\widehat{\psi}_0$ yields an estimate of the form:

$$\widehat{\tau}_{\mathrm{RD}}^{\mathrm{OS}} = \frac{m}{N(1-\alpha)}\widehat{\tau}_{\mathrm{RD}}^{\mathrm{EE}} + \left(1 - \frac{m}{N(1-\alpha)}\right)\widehat{\tau}_{\mathrm{RD}}.$$

In particular, estimators $\widehat{\psi}_a$ of the form G-formula or weighted Horwitz-Thomson yield a final one-step estimator that has the same structure as the estimating equation estimator.

(RR) For the risk ratio, the estimating equation approach yield $\widehat{\tau}_{\mathrm{RR}}^{\mathrm{EE}} = \widehat{\psi}_1^{\mathrm{EE}}/\widehat{\psi}_0^{\mathrm{EE}}$. In contrast, the one-step approach, based on initial estimators $\widehat{\psi}_1$ and $\widehat{\psi}_0$ yields

$$\widehat{\tau}_{\mathrm{RR}}^{\mathrm{OS}} = \frac{\widehat{\psi}_1}{\widehat{\psi}_0} + \frac{1}{\widehat{\psi}_0}\frac{m}{N(1-\alpha)}(\widehat{\psi}_1^{\mathrm{EE}} - \widehat{\psi}_1) - \frac{\widehat{\psi}_1}{\widehat{\psi}_0^2}\frac{m}{N(1-\alpha)}(\widehat{\psi}_0^{\mathrm{EE}} - \widehat{\psi}_0).$$

In general, $\widehat{\tau}_{\mathrm{RR}}^{\mathrm{OS}} \neq \widehat{\tau}_{\mathrm{RR}}^{\mathrm{EE}}$, unless of course we initially picked $\widehat{\psi}_a = \widehat{\psi}_a^{\mathrm{EE}}$.

4 Generalization under exchangeability in effect measure

Estimators in Section 3 were derived under the exchangeability in mean assumption (Assumption 3). Under this assumption, generalizing necessitates full access to all prognostic covariates whose distributions is shifted between the source and target populations [8]. A weaker assumption consists in transportability of conditional treatment effects.

Assumption 5 (Exchangeability in effect measure)**.** $\forall x \in \mathrm{supp}(P_{\mathrm{T}}) \cap \mathrm{supp}(P_{\mathrm{S}})$, $\tau_\Phi^{\mathrm{S}}(x) = \tau_\Phi^{\mathrm{T}}(x)$.

Under Assumption 5, the effect of the treatment depends on the patient's features in the same way in the source and target population. While exchangeability in mean implies exchangeability of the effect measure for any Φ, the reverse does not generally hold.

4.1 Transporting causal measures under exchangeability in effect measure

To ensure identification, we typically require a relationship between the conditional average treatment effect (CATE) and the average treatment effect (ATE). A classical concept that supports this

relationship is collapsibility: a measure is said to be collapsible if it can be expressed as a weighted average of conditional effects [see, e.g., 19, 23]. However, some measure (like OR) are not collapsible, thus questioning their transportability under this Assumption 5. However, it turns out that under Assumption 5, any first-moment causal measure is identifiable assuming access to the control potential outcomes $Y^{(0)}$ for all individuals in the target population:

$$\tau_\Phi^{\mathrm{T}} = \Phi\left(\mathbb{E}_{\mathrm{T}}\left[\Gamma\left(\tau_\Phi^{\mathrm{S}}(X), \mu_{(0)}^{\mathrm{T}}(X)\right)\right], \mathbb{E}_{\mathrm{T}}\left[Y^{(0)}\right]\right) \qquad \textit{(transporting)} \tag{12}$$

$$= \Phi\left(\mathbb{E}_{\mathrm{S}}\left[\frac{p_{\mathrm{T}}(X)}{p_{\mathrm{S}}(X)}\Gamma\left(\tau_\Phi^{\mathrm{S}}(X), \mu_{(0)}^{\mathrm{T}}(X)\right)\right], \mathbb{E}_{\mathrm{T}}\left[Y^{(0)}\right]\right) \qquad \textit{(weighting)} \tag{13}$$

where Γ is the effect function (see Definition 1). Note that for collapsible measures such as RD or RR, expressions (12) and (13) reduce to the identification results derived by [7]. Thus, our framework can be viewed as a natural extension of their approach to any first moment population causal measures even non-collapsible measures such as the OR. The detailed identification results for RD, OR, and RR are provided in Appendix C. Using the identification formula (12), we derive Γ-formula estimators, which, to the best of our knowledge, are novel contributions.

Definition 4 (Transported Γ-formula)**.** *For any first moment population causal measure Φ, we define the transported Γ-formula estimator $\hat{\tau}_{\Phi,\mathrm{t}\Gamma}$ as follows:*

$$\hat{\tau}_{\Phi,\mathrm{t}\Gamma} = \Phi\left(\frac{1}{N-n}\sum_{S_i=0}\Gamma(\tau_\Phi^{\mathrm{S}}(X_i), \mu_{(0)}^{\mathrm{T}}(X_i)), \frac{1}{N-n}\sum_{S_i=0}\mu_{(0)}^{\mathrm{T}}(X_i)\right). \tag{14}$$

We can also define, as before, a reweighted version of the Γ-formula (see Definition 3).

4.2 Semiparametric efficient estimators under exchangeability of treatment effect

Under Assumption 5, the identifiability formula (12) serves as the basis for constructing the EIF. Given access to the target baseline distribution, $\mu_{(0)}^{\mathrm{T}}$ and ψ_0^{T} are known. To construct one-step and estimating equation estimators (see Section 3.2), we require EIF of τ_Φ^{T}, denoted $\varphi_\Phi(Z, \eta, \tau_\Phi^{\mathrm{T}})$ which is is related to φ_1, the EIF of ψ_1^{T}, via the chain rule:

$$\varphi_\Phi(Z, \eta, \tau_\Phi^{\mathrm{T}}) = \varphi_1(Z, \eta, \psi_1^{\mathrm{T}})\partial_1\Phi(\psi_1^{\mathrm{T}}, \psi_0^{\mathrm{T}}),$$

where $\eta = (\mu_{(0)}^{\mathrm{S}}, \tau_\Phi, r)$ is the nuisance parameters.

Proposition 6. *The influence function of ψ_1^{T} at P_{obs} is given by*

$$\varphi_1(Z, \eta, \psi_1^{\mathrm{T}}) := \frac{Sr(X)}{\alpha}\left(\frac{A}{\pi}(Y - \Gamma(\tau_\Phi(X), \mu_{(0)}^{\mathrm{S}}(X)))\right)$$
$$- \frac{Sr(X)}{\alpha}\left(\frac{1-A}{1-\pi}(Y - \mu_{(0)}^{\mathrm{S}}(X))\partial_0\Gamma(\tau_\Phi(X), \mu_{(0)}^{\mathrm{T}}(X))\right) + \frac{1-S}{1-\alpha}\left(\Gamma(\tau_\Phi(X), \mu_{(0)}^{\mathrm{T}}(X)) - \psi_1^{\mathrm{T}}\right).$$

As in Section 3.2, we can either find the estimating equation estimator of ψ_1 and plug it into Φ to obtain $\widehat{\tau}_\Phi^{\mathrm{EE}} = \Phi(\widehat{\psi}_1^{\mathrm{EE}}, \psi_0^{\mathrm{T}})$, or apply a one step correction to a initial estimator $\widehat{\psi}_1$ of the form $\widehat{\tau}_\Phi^{\mathrm{OS}} = \Phi(\widehat{\psi}_1, \psi_0^{\mathrm{T}}) + (1/N)\sum_{i=1}^N \varphi_\Phi(Z_i, \widehat{\eta}, \widehat{\tau}_\Phi^{\mathrm{T}})$. Like in Section 3.2, the estimating equation estimators and one-step estimators have no reason to coincide in general. Exact computations of the RD, RR and OR are provided in Appendix C.3.

5 Simulations

5.1 Synthetic data

We generate data $(S, X, A, Y^{(0)}, Y^{(1)})$ using the following binary outcome model: for all $a, s \in \{0, 1\}$, $\mathbb{P}(Y^{(a)} = 1 \mid X, S = s) = p_s^{(a)}(V)$ where $V^\top = [1, X^\top]$ and $X|S = s \sim \mathcal{N}(\nu_s, I_d)$. We set $d = 5$ and $S \sim \mathrm{B}(0.3)$ to reflect limited RCT data relative to the target population, and $A \mid S = 1 \sim \mathrm{B}(0.5)$. We evaluate the estimators from Section 3 and 4, estimating nuisance components—regression surfaces and density ratios—using parametric methods (linear/logistic

regression). The red (resp. gray, when displayed) dotted line represents the treatment effect in the target (resp. source) population. A basic linear setting, in which all estimators perform well, is presented in Appendix D.

Experiment 1 (Exchangeability in mean and non-linear/non-logistic response): We consider a setting under which Assumption 3 holds and $\forall a, s \in \{0,1\}, \quad p_s^{(a)}(V) = \sigma(\beta_0^\top V \cdot (V^\top \beta_1)^a)$. Both G-formula-based estimators exhibit substantial bias across all evaluation metrics, which is expected given that the non-linear response surfaces are misspecified by using linear regression. In contrast, the estimating equation–based estimators remain unbiased across all measures, benefiting from their double robustness property. Among the one-step estimators, only the RD variant is accurate in this setting. This is also anticipated, as one-step estimators generally do not retain double robustness—except in cases where they coincide with the corresponding estimating equation estimator, which holds true for the RD variant.

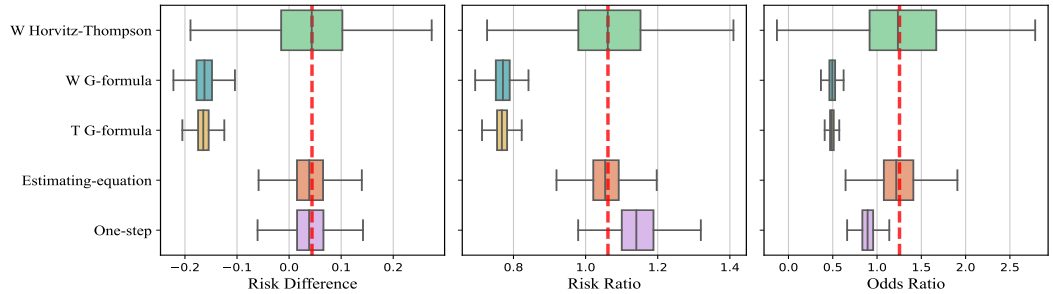

Figure 1: Comparison of estimators across different causal measures under a non-linear outcome model with a sample size of $N = 50{,}000$ and 3,000 repetitions. Source values are 0.45 / 3.2 / 7.5.

Experiment 2 (Exchangeability in effect and linear/logistic response): We now consider a setting for which Assumption 5 holds and such that, depending on the causal measure: (RD) Model 1 and $\beta_{\mathrm{T}}^{(a)} = \beta_{\mathrm{S}}^{(a)} + \theta$, (RR) $p_s^{(a)}(X) = \sigma(X^\top \beta_s) \cdot \sigma(X^\top \gamma)^a$, (OR) $p_s^{(a)}(X) = \sigma(X^\top(\beta_s + a \cdot \gamma))$. In this setting, Assumption 3 is no longer satisfied, as the nuisance functions depend on S. However, one can verify that Assumption 5 holds for each model. Estimators introduced in Section 3 still converge for the RD, despite the violation of strong transportability, thanks to the linearity of the RD. In contrast, the RR and OR estimators fail to converge, since these measures are nonlinear. On the other hand, all estimators introduced in Section 4 remain unbiased, as expected.

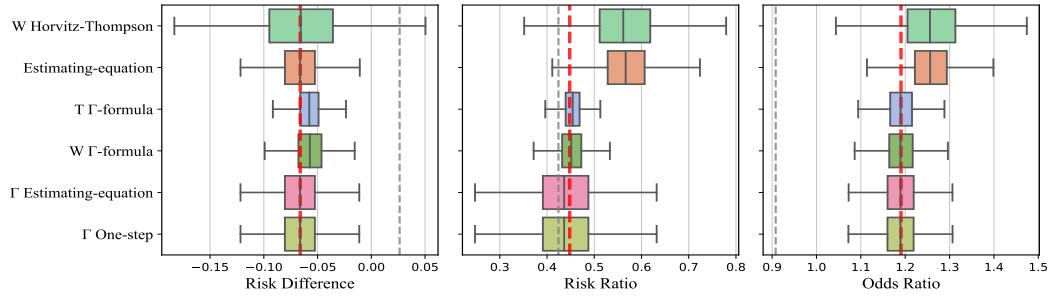

Figure 2: Comparison of estimators across different causal measures under a linear outcome model with a sample size of $N = 50{,}000$ and 3,000 repetitions.

5.2 Real-World Experiment

We evaluate estimators using a case study on the effectiveness of tranexamic acid on mortality for brain injury patients, combining data from the CRASH-3 trial and the Traumabase registry. CRASH-3, is a RCT with over 9,000 TBI patients from 29 countries, while Traumabase provides detailed clinical data on 8,000+ patients from 23 French trauma centers. Following [9] we consider

six covariates (age, sex, injury time, systolic BP, GCS score, pupil reactivity). Since this is a real dataset, the true treatment effect is unknown. Results are displayed in Figure 3. All estimators (except one) indicate a positive treatment effect; however, the confidence intervals are wide and include the null value (zero or one), preventing any definitive conclusions about the treatment's effectiveness.

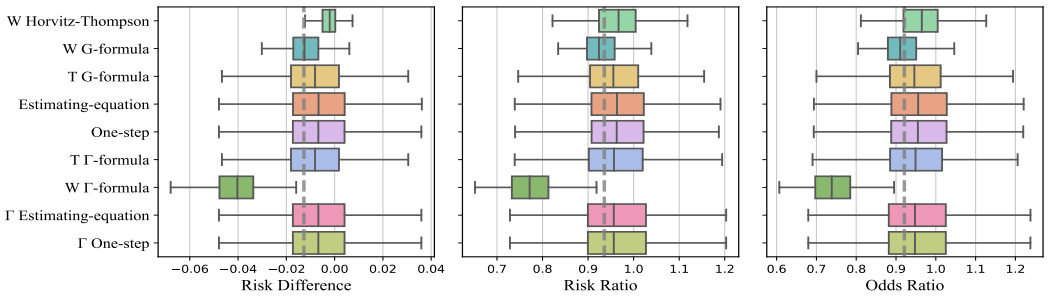

Figure 3: Comparison of estimators across different causal measures on the combined CRASH-3 and Traumabase dataset. Confidence intervals were estimated using stratified bootstrap resampling.

6 Conclusion

This article introduces a general framework for the generalization of first-moment, population-level causal estimands from RCTs to broader target populations. We propose different estimators (weighting based on density ratios, outcome regression methods, and approaches based on EIF) and analyze their statistical properties. A central source of complexity lies in the inherent nonlinearity of many causal estimands, which disrupts the alignment between two classical EIF techniques: one-step and estimating equations. This divergence gives rise to a diverse landscape of possible estimators, which differ depending on the underlying identification assumptions. While assuming exchangeability in effect measure is less restrictive than exchangeability in mean, few estimating methods are available in the former case, as most of them are based on CATE estimation, which is difficult for non-linear measure. In practice, we recommend to resort to EE estimators, as they are provably doubly robust and behave well in our controlled experimental setting. Further layers of complexity arise from the selection of appropriate nuisance estimators, whether parametric or nonparametric, each with its own trade-offs. Nevertheless, this work represents a step forward in enabling the computation of both absolute and relative causal measures in new target populations, thereby contributing to more informed clinical decision-making and external validity in causal research.

7 Acknowledgments

This work has been done in the frame of the PEPR SN SMATCH project and has benefited from a governmental grant managed by the Agence Nationale de la Recherche un- der the France 2030 programme, reference ANR-22-PESN- 0003.

References

[1] Boughdiri, A., J. Josse, and E. Scornet (2024). Quantifying treatment effects: Estimating risk ratios in causal inference.

[2] Campbell, H. and A. Remiro-Azócar (2025). Doubly robust augmented weighting estimators for the analysis of externally controlled single-arm trials and unanchored indirect treatment comparisons.

[3] Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2017). Double/debiased machine learning for treatment and causal parameters.

[4] Cole, P. and B. MacMAHON (1971). Attributable risk percent in case-control studies. *British journal of preventive & social medicine 25*(4), 242.

[5] Cole, S. R. and E. A. Stuart (2010). Generalizing evidence from randomized clinical trials to target populations: the actg 320 trial. *American journal of epidemiology 172*(1), 107–115.

[6] Colnet, B., J. Josse, G. Varoquaux, and E. Scornet (2022, January). Causal effect on a target population: A sensitivity analysis to handle missing covariates. *Journal of Causal Inference 10*(1), 372–414.

[7] Colnet, B., J. Josse, G. Varoquaux, and E. Scornet (2023). Risk ratio, odds ratio, risk difference... which causal measure is easier to generalize? *arXiv preprint arXiv:2303.16008*.

[8] Colnet, B., J. Josse, G. Varoquaux, and E. Scornet (2024). Reweighting the rct for generalization: finite sample error and variable selection.

[9] Colnet, B., I. Mayer, G. Chen, A. Dieng, R. Li, G. Varoquaux, J.-P. Vert, J. Josse, and S. Yang (2023). Causal inference methods for combining randomized trials and observational studies: a review.

[10] Colnet, B., I. Mayer, G. Chen, A. Dieng, R. Li, G. Varoquaux, J.-P. Vert, J. Josse, and S. Yang (2024). Causal inference methods for combining randomized trials and observational studies: a review. *Statistical science 39*(1), 165–191.

[11] Dahabreh, I. J., S. E. Robertson, L. C. Petito, M. A. Hernán, and J. A. Steingrimsson (2023). Efficient and robust methods for causally interpretable meta-analysis: Transporting inferences from multiple randomized trials to a target population. *Biometrics 79*(2), 1057–1072.

[12] Dahabreh, I. J., S. E. Robertson, J. A. Steingrimsson, E. A. Stuart, and M. A. Hernan (2020). Extending inferences from a randomized trial to a new target population. *Statistics in medicine 39*(14), 1999–2014.

[13] Dahabreh, I. J., J. M. Robins, S. J. Haneuse, and M. A. Hernán (2019). Generalizing causal inferences from randomized trials: counterfactual and graphical identification. *arXiv preprint arXiv:1906.10792*.

[14] Degtiar, I. and S. Rose (2023). A review of generalizability and transportability. *Annual Review of Statistics and Its Application 10*(1), 501–524.

[15] Demirel, I., A. Alaa, A. Philippakis, and D. Sontag (2024). Prediction-powered generalization of causal inferences. *arXiv preprint arXiv:2406.02873*.

[16] Fay, M. P. and F. Li (2024, Oct). Causal interpretation of the hazard ratio in randomized clinical trials. *Clinical Trials 21*(5), 623–635. Epub 2024 Apr 28.

[17] French Health Authority (2024). Pricing & reimbursement of drugs and hta policies in france.

[18] Greenland, S. (1987, 05). Interpretation and choice of effect measures in epidemiologic analysies. *American Journal of Epidemiology 125*(5), 761–768.

[19] Greenland, S., J. M. Robbins, and J. Pearl (1999, 01). Confounding and collapsibility in causal inference. *Statistical Science 14*, 29–46.

[20] Hong, H., L. Liu, and E. A. Stuart (2025, Feb). Estimating target population treatment effects in meta-analysis with individual participant-level data. *Statistical Methods in Medical Research 34*(2), 355–368. Epub 2025 Jan 19.

[21] Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association 47*(260), 663–685.

[22] Huitfeldt, A., M. P. Fox, E. J. Murray, A. Hróbjartsson, and R. M. Daniel (2022). Shall we count the living or the dead?

[23] Huitfeldt, A., M. Stensrud, and E. Suzuki (2019, 01). On the collapsibility of measures of effect in the counterfactual causal framework. *Emerging Themes in Epidemiology 16*.

[24] Kanamori, T., T. Suzuki, and M. Sugiyama (2010). Theoretical analysis of density ratio estimation. *IEICE transactions on fundamentals of electronics, communications and computer sciences 93*(4), 787–798.

[25] Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*.

[26] Kennedy, E. H. (2024). Semiparametric doubly robust targeted double machine learning: a review. *Handbook of Statistical Methods for Precision Medicine*, 207–236.

[27] Kern, H. L., E. A. Stuart, J. Hill, and D. P. Green (2016). Assessing methods for generalizing experimental impact estimates to target populations. *Journal of research on educational effectiveness 9*(1), 103–127.

[28] King, N. B., S. Harper, and M. E. Young (2012). Use of relative and absolute effect measures in reporting health inequalities: structured review. *Bmj 345*.

[29] Laupacis, A., D. L. Sackett, and R. S. Roberts (1988). An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine 318*(26), 1728–1733. PMID: 3374545.

[30] Liu, J., K. Zhu, S. Yang, and X. Wang (2025). Robust estimation and inference in hybrid controlled trials for binary outcomes: A case study on non-small cell lung cancer.

[31] Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gøtzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman (2010). Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ 340*.

[32] Pearl, J. (2009). *Causality* (2 ed.). Cambridge University Press.

[33] Pearl, J. and E. Bareinboim (2011). Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, AAAI'11, pp. 247–254. AAAI Press.

[34] Richardson, T. S., J. M. Robins, and L. Wang (2017). On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association 112*(519), 1121–1130.

[35] Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling 7*(9), 1393–1512.

[36] Rothwell, P. M. (2005). External validity of randomised controlled trials: "to whom do the results of this trial apply?". *The Lancet 365*, 82–93.

[37] Rott, K. W., G. Bronfort, H. Chu, J. D. Huling, B. Leininger, M. H. Murad, Z. Wang, and J. S. Hodges (2024). Causally interpretable meta-analysis: Clearly defined causal effects and two case studies. *Research Synthesis Methods 15*(1), 61–72.

[38] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational psychology 66*(5), 688–701.

[39] Schulz, K. F., D. G. Altman, D. Moher, and C. Group* (2010). Consort 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of internal medicine 152*(11), 726–732.

[40] Shirvaikar, V. and C. Holmes (2023). Targeting relative risk heterogeneity with causal forests.

[41] Splawa-Neyman, J., D. M. Dabrowska, and T. P. Speed (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science 5*(4), 465 – 472.

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

[43] Stuart, E. A., S. R. Cole, C. P. Bradshaw, and P. J. Leaf (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society Series A: Statistics in Society 174*(2), 369–386.

[44] Yadlowsky, S., F. Pellegrini, F. Lionetto, S. Braune, and L. Tian (2021). Estimation and validation of ratio-based conditional average treatment effects using observational data. *Journal of the American Statistical Association 116*(533), 335–352.

Technical Appendices and Supplementary Material

Contents

A First Moment Causal Measures

Measure	Effect Measure $\Phi(\psi_1, \psi_0)$	Effect Function $\Gamma(\tau, \psi_0)$
Risk Difference (RD)	$\Phi(\psi_1, \psi_0) = \psi_1 - \psi_0$	$\Gamma(\tau, \psi_0) = \psi_0 + \tau$
Risk Ratio (RR)	$\Phi(\psi_1, \psi_0) = \frac{\psi_1}{\psi_0}$	$\Gamma(\tau, \psi_0) = \tau \cdot \psi_0$
Odds Ratio (OR)	$\Phi(\psi_1, \psi_0) = \frac{\psi_1}{1-\psi_1} \cdot \frac{1-\psi_0}{\psi_0}$	$\Gamma(\tau, \psi_0) = \frac{\tau \cdot \psi_0}{1+\tau \cdot \psi_0 - \psi_0}$
Number Needed to Treat (NNT)	$\Phi(\psi_1, \psi_0) = \frac{1}{\psi_1 - \psi_0}$	$\Gamma(\tau, \psi_0) = \frac{1}{\tau} + \psi_0$
Switch Relative Risk (GRRR)	$\Phi(\psi_1, \psi_0) = \begin{cases} 1 - \frac{1-\psi_1}{1-\psi_0} & \text{if } \psi_1 > \psi_0 \\ 0 & \text{if } \psi_1 = \psi_0 \\ -1 + \frac{\psi_1}{\psi_0} & \text{if } \psi_1 < \psi_0 \end{cases}$	$\Gamma(\tau, \psi_0) = \begin{cases} 1 - (1-\tau)(1-\psi_0) & \text{if } \psi_0 > 0 \\ \tau & \text{if } \psi_0 = 0 \\ \tau(1+\psi_0) & \text{if } \psi_0 < 0 \end{cases}$
Excess Risk Ratio (ERR)	$\Phi(\psi_1, \psi_0) = \frac{\psi_1 - \psi_0}{\psi_0}$	$\Gamma(\tau, \psi_0) = \tau(1 + \psi_0)$
Survival Ratio (SR)	$\Phi(\psi_1, \psi_0) = \frac{1-\psi_1}{1-\psi_0}$	$\Gamma(\tau, \psi_0) = 1 - \tau(1 - \psi_0)$
Relative Susceptibility (RS)	$\Phi(\psi_1, \psi_0) = \frac{1-\psi_0}{1-\psi_1}$	$\Gamma(\tau, \psi_0) = 1 - \frac{1-\tau}{\psi_0}$
Log Odds Ratio (log-OR)	$\Phi(\psi_1, \psi_0) = \log\left(\frac{\psi_1(1-\psi_0)}{\psi_0(1-\psi_1)}\right)$	$\Gamma(\tau, \psi_0) = \exp(\psi_0) \cdot \frac{\tau}{1-\tau+\exp(\psi_0)\cdot\tau}$
Odds Product	$\Phi(\psi_1, \psi_0) = \frac{\psi_1}{1-\psi_1} \cdot \frac{\psi_0}{1-\psi_0}$	$\Gamma(\tau, \psi_0) = \frac{\sqrt{\tau \cdot \frac{\psi_0}{1-\psi_0}}}{1+\sqrt{\tau \cdot \frac{\psi_0}{1-\psi_0}}}$
Arcsine Difference	$\Phi(\psi_1, \psi_0) = \arcsin(\sqrt{\psi_1}) - \arcsin(\sqrt{\psi_0})$	$\Gamma(\tau, \psi_0) = \sin(\psi_0 + \arcsin(\sqrt{\tau}))^2$
Relative Risk Reduction (RRR)	$\Phi(\psi_1, \psi_0) = 1 - \frac{\psi_1}{\psi_0}$	$\Gamma(\tau, \psi_0) = \tau(1 - \psi_0)$

B Transporting/Reweighting a causal effect under exchangeability of conditional outcome

B.1 Identification under exchangeability of conditional outcome

$$\begin{aligned}
\mathbb{E}_T\left[Y^{(a)}\right] &:= \mathbb{E}_T\left[\mathbb{E}_T\left[Y^{(a)} \mid X\right]\right] \\
&= \mathbb{E}_T\left[\mathbb{E}_S\left[Y^{(a)} \mid X\right]\right] \quad \text{(Transportability)} \\
&= \mathbb{E}_S\left[\mathbb{E}_S\left[Y^{(a)} \mid X\right] \cdot \frac{P_T(X)}{P_S(X)}\right] \quad \text{(Overlap)} \\
&= \mathbb{E}_S\left[Y^{(a)} \cdot \frac{P_T(X)}{P_S(X)}\right]
\end{aligned}$$

B.2 Oracle weighted Horvitz-Thompson

Proposition 7. *For ease of notation, we let $\pi_a := \mathbb{P}_S(A = a) = \pi^a(1-\pi)^{1-a}$. Let $\psi_a = \mathbb{E}_T[Y^{(a)}]$ denote the target population mean potential outcome under treatment $a \in \{0, 1\}$. Define the oracle estimator*

$$\psi^*_{a,wHT} = \frac{1}{n}\sum_{i=1}^{N} S_i \cdot r(X_i) \cdot \frac{\mathbb{1}\{A_i = a\}Y_i}{\pi_a},$$

where $r(X)$ denotes the density ratio between the target and source covariate distributions and $n = \sum_{i=1}^{N} S_i$ is the number of units in the source sample. Then, under Assumption 1 to 4 we have,

$$\sqrt{N}\left(\psi^*_{a,wHT} - \psi_a\right) \xrightarrow{d} \mathcal{N}(0, V^*_{a,wHT}),$$

where the asymptotic variance $V^*_{a,wHT}$ is given by

$$V^*_{a,wHT} = \frac{1}{\alpha}\left(\frac{1}{\pi_a}\mathbb{E}_T\left[r(X)(Y^{(a)})^2\right] - (\psi_a)^2\right), \tag{15}$$

Proof. Define

$$Z_i = S_i \cdot r(X_i) \cdot \frac{\mathbb{1}\{A_i = a\}Y_i}{\pi_a},$$

so that

$$\psi^*_{a,\text{wHT}} = \frac{1}{N}\sum_{i=1}^{N} Z_i.$$

We first compute the expectations of Z_i and S_i. By the definition of Z_i and using that $S_i \sim$ Bernoulli(α),

$$\begin{aligned}
\mathbb{E}[Z_i] &= \mathbb{E}\left[S \cdot r(X) \cdot \frac{\mathbb{1}\{A = a\}Y}{\pi_a}\right] \\
&= \alpha\,\mathbb{E}_S\left[r(X) \cdot \frac{\mathbb{1}\{A = a\}Y}{\pi_a}\right] \\
&= \alpha\,\mathbb{E}_S\left[r(X) \cdot Y^{(a)}\right] \\
&= \alpha\,\mathbb{E}_T[Y^{(a)}] \\
&= \alpha\psi_a,
\end{aligned}$$

where we used the consistency assumption $Y = Y^{(A)}$, the randomization of A, and the density ratio property. Since (Z_i, S_i) for $i = 1, \ldots, N$ are i.i.d., one can apply the multivariate central limit theorem to

$$\left(\frac{1}{N}\sum_{i=1}^{N} Z_i, \quad \frac{1}{N}\sum_{i=1}^{N} S_i\right),$$

which leads to

$$\sqrt{N}\left(\begin{pmatrix}\frac{1}{N}\sum_{i=1}^{N} Z_i \\ \frac{1}{N}\sum_{i=1}^{N} S_i\end{pmatrix} - \begin{pmatrix}\mathbb{E}[Z_i] \\ \mathbb{E}[S_i]\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix}\mathbb{V}[Z_i] & \text{Cov}(Z_i, S_i) \\ \text{Cov}(S_i, Z_i) & \mathbb{V}[S_i]\end{pmatrix}.$$

Noting that $\psi^*_{a,\text{wHT}}$ can be written as

$$\psi^*_{a,\text{wHT}} = \frac{\frac{1}{N}\sum_{i=1}^{N} Z_i}{\frac{1}{N}\sum_{i=1}^{N} S_i},$$

one can apply the Delta method to the map $h : (x, y) \mapsto x/y$, whose gradient evaluated at $(\alpha\psi_a, \alpha)$ is

$$\nabla h(\alpha\psi_a, \alpha) = \left(\frac{1}{\alpha},\ -\frac{\psi_a}{\alpha}\right).$$

Thus,

$$\sqrt{N}\left(\psi^*_{a,\text{wHT}} - \psi_a\right) \xrightarrow{d} \mathcal{N}(0, V^*_{a,\text{wHT}}),$$

where

$$\begin{aligned}
V^*_{a,\text{wHT}} &= \nabla h(\alpha\psi_a, \alpha)^\top \Sigma \nabla h(\alpha\psi_a, \alpha) &(16)\\
&= \frac{1}{\alpha^2}\mathbb{V}[Z_i] + \frac{(\psi_a)^2}{\alpha^2}\mathbb{V}[S_i] - \frac{2\psi_a}{\alpha^2}\text{Cov}(Z_i, S_i). &(17)
\end{aligned}$$

It remains to compute each term. Since $S_i \sim$ Bernoulli(α),

$$\mathbb{V}[S_i] = \alpha(1 - \alpha).$$

By direct computation,

$$\mathbb{V}[Z_i] = \alpha\, \mathbb{E}_T\left[\frac{r(X)(Y^{(a)})^2}{\pi_a}\right] - (\alpha\psi_a)^2.$$

Moreover, since $S_i \times Z_i = Z_i$, we have

$$\mathrm{Cov}(Z_i, S_i) = (1-\alpha)\alpha\psi_a.$$

Substituting into the expression of $V^*_{a,\text{wHT}}$, we find

$$V^*_{a,\text{wHT}} = \frac{1}{\alpha}\left(\frac{1}{\pi_a}\mathbb{E}_T\left[r(X)(Y^{(a)})^2\right] - (\psi_a)^2\right), \tag{18}$$

as claimed. □

Proposition 8 (Asymptotic Normality Oracle weighted Horvitz-Thompson)**.** *Let $\tau^*_{\Phi,wHT}$ denote the oracle weighted Horvitz-Thompson estimator. Then under Assumption 1 to 4, we have:*

$$\sqrt{N}\left(\tau^*_{\Phi,wHT} - \tau^T_\Phi\right) \xrightarrow{d} \mathcal{N}\left(0, V^*_{\Phi,wHT}\right).$$

Proof. Noting that $\tau^*_{\Phi,\text{wHT}} = \Phi(\psi^*_{1,\text{wHT}}, \psi^*_{0,\text{wHT}})$ and using Proposition 7, we know that $(\psi^*_{1,\text{wHT}}, \psi^*_{0,\text{wHT}})$ are jointly asymptotically normal. Specifically,

$$\sqrt{N}\begin{pmatrix}\psi^*_{1,\text{wHT}} - \psi_1 \\ \psi^*_{0,\text{wHT}} - \psi_0\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma^*_\pi\right),$$

where Σ^*_π is the asymptotic covariance matrice. Applying the delta method to the smooth function $\Phi : \mathbb{R}^2 \to \mathbb{R}$, we have

$$\sqrt{N}(\tau^*_{\Phi,\text{wHT}} - \tau^T_\Phi) \xrightarrow{d} \mathcal{N}\left(0, \nabla\Phi^\top \Sigma^*_\pi \nabla\Phi\right),$$

where $\nabla\Phi$ denotes the gradient of Φ evaluated at (ψ_1, ψ_0). Moreover, because the treatment assignment A is binary ($A \in \{0,1\}$), we have that for each unit, $A_i(1-A_i) = 0$, so the covariance between ψ^*_1 and ψ^*_0 is

$$\mathrm{Cov}(\psi^*_{1,\text{wHT}}, \psi^*_{0,\text{wHT}}) = -\psi_1\psi_0.$$

Expanding the delta method variances and fully factorizing we get:

$$V^*_{\Phi,\text{wHT}} = \frac{1}{\alpha}\Bigg(\left(\frac{\partial\Phi}{\partial\psi_1}\right)^2 \mathbb{E}_T\left[\frac{r(X)(Y^{(1)})^2}{\mathbb{P}_S(A=1)}\right] + \left(\frac{\partial\Phi}{\partial\psi_0}\right)^2 \mathbb{E}_T\left[\frac{r(X)(Y^{(0)})^2}{\mathbb{P}_S(A=0)}\right] - \left(\frac{\partial\Phi}{\partial\psi_1}\mathbb{E}_T\left[Y^{(1)}\right] + \frac{\partial\Phi}{\partial\psi_0}\mathbb{E}_T\left[Y^{(0)}\right]\right)^2\Bigg).$$

□

Example 2.

Measure	Variance
Risk Difference (RD)	$\frac{1}{\alpha}\left(\frac{\mathbb{E}_T\left[r(X)(Y^{(1)})^2\right]}{\pi} + \frac{\mathbb{E}_T\left[r(X)(Y^{(0)})^2\right]}{1-\pi} - (\tau^T_{RD})^2\right)$
Risk Ratio (RR)	$\frac{(\tau^T_{RR})^2}{\alpha}\left(\frac{\mathbb{E}_T\left[r(X)(Y^{(1)})^2\right]}{\pi\mathbb{E}_T\left[Y^{(1)}\right]^2} + \frac{\mathbb{E}_T\left[r(X)(Y^{(0)})^2\right]}{(1-\pi)\mathbb{E}_T\left[Y^{(0)}\right]^2}\right)$
Odds Ratio (OR)	$\frac{(\tau^T_{OR})^2}{\alpha}\left(\frac{\mathbb{E}_T\left[r(X)(Y^{(1)})^2\right]}{\pi(\mathbb{E}_T[Y^{(1)}])^2} + \frac{\mathbb{E}_T\left[r(X)(Y^{(0)})^2\right]}{(1-\pi)(\mathbb{E}_T[Y^{(0)}])^2} - 1\right)$

B.3 Logistic weighted Horvitz-Thomson Estimator

Proposition 9. *Let $\psi_a = \mathbb{E}_T[Y^{(a)}]$ denote the target population mean potential outcome under treatment $a \in \{0,1\}$. Define the estimator*

$$\hat{\psi}_{a,wHT} = \frac{1}{n}\sum_{i=1}^{N} S_i \cdot r(X_i, \hat{\beta}_N) \cdot \frac{\mathbb{1}\{A_i = a\}Y_i}{\pi_a}, \quad \textit{with} \quad r(x, \hat{\beta}_N) = \frac{n}{N-n} \cdot \frac{1-\sigma(x,\hat{\beta}_N)}{\sigma(x,\hat{\beta}_N)}$$

where $n = \sum_{i=1}^{N} S_i$ and $\hat{\beta}_N$ the maximum likelihood estimate (MLE) from logistic regression of the selection indicator S on covariates X, and $\sigma(x,\beta) = (1+e^{-x^\top\beta})^{-1}$ the logistic function. Then, under Assumption 1 to 3.1,

$$\sqrt{N}\begin{pmatrix}\hat{\psi}_{1,wHT} - \psi_1 \\ \hat{\psi}_{0,wHT} - \psi_0\end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma_{wHT}), \quad \textit{with} \quad \Sigma_{wHT} = \begin{pmatrix} V_{1,wHT} & C_{wHT} \\ C_{wHT} & V_{0,wHT}\end{pmatrix},$$

where the asymptotic variance $V_{a,wHT}$ is given by

$$V_{a,wHT} = \frac{\mathbb{E}_T[r(X)(Y^{(a)})^2]}{\alpha\pi_a} - \frac{\mathbb{E}_T[Y^{(a)}]^2}{1-\alpha} \tag{19}$$

$$- \mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[Y^{(a)}V] + 2\,\mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[\sigma(X)Y^{(a)}V], \tag{20}$$

$$\textit{and} \quad C_{wHT} = -\frac{\mathbb{E}_T[Y^{(1)}]\mathbb{E}_T[Y^{(0)}]}{1-\alpha} \tag{21}$$

$$+ \mathbb{E}_T[Y^{(1)}V^\top]Q^{-1}\mathbb{E}_T[\sigma(X)Y^{(0)}V] \tag{22}$$

$$+ \mathbb{E}_T[Y^{(0)}V^\top]Q^{-1}\mathbb{E}_T[\sigma(X)Y^{(1)}V] \tag{23}$$

$$- \mathbb{E}_T[Y^{(1)}V^\top]Q^{-1}\mathbb{E}_T[Y^{(0)}V] \tag{24}$$

with $V = (1, X)$ and $Q = (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)VV^\top\right]$.

Proof. Let $Z = (S, X, S\times A, S\times Y)$ and define the parameter vector $\theta = (\theta_0, \theta_1, \theta_2, \beta)$. We define the estimating function $\lambda(Z,\theta)$ and the estimator $\hat{\theta}_N$ as follows:

$$\lambda(Z,\theta) = \begin{pmatrix} \frac{1}{1-\theta_2}\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\mathbb{1}\{A=0\}Y}{\mathbb{P}_S(A=0)} - \theta_0 \\ \frac{1}{1-\theta_2}\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\mathbb{1}\{A=1\}Y}{\mathbb{P}_S(A=1)} - \theta_1 \\ S - \theta_2 \\ V(S-\sigma(X,\beta))\end{pmatrix}, \quad \hat{\theta}_N = \begin{pmatrix} \hat{\psi}_{0,\text{wHT}} \\ \hat{\psi}_{1,\text{wHT}} \\ \hat{\alpha} := \frac{1}{N}\sum_{i=1}^{N} S_i \\ \hat{\beta}_N\end{pmatrix}.$$

Note that the reweighting function can be expressed as:

$$r(X_i, \hat{\beta}_N) = \frac{1-\sigma(X_i,\hat{\beta}_N)}{\sigma(X_i,\hat{\beta}_N)} \cdot \frac{\hat{\alpha}}{1-\hat{\alpha}}.$$

Thus, we can rewrite the estimator as:

$$\hat{\psi}_{a,\text{wHT}} = \frac{1}{N\hat{\alpha}}\sum_{j=1}^{N} S_j r(X_j,\hat{\beta}_N)\frac{\mathbb{1}\{A_j=a\}Y_j}{\pi_a} \tag{25}$$

$$= \frac{1}{N}\sum_{j=1}^{N} S_j \frac{1-\sigma(X_j,\hat{\beta}_N)}{\sigma(X_j,\hat{\beta}_N)(1-\hat{\alpha})}\frac{\mathbb{1}\{A_j=a\}Y_j}{\pi_a}. \tag{26}$$

Furthermore, the log-likelihood function of β is:

$$-\ln L_N(\beta) = -\sum_{i=1}^{N} s_i \log\sigma(X_i;\beta) + (1-s_i)\log(1-\sigma(X_i;\beta))$$

where $\sigma(X;\beta) = (1+\exp(-X^\top\beta_1 - \beta_0))^{-1}$. Simple calculations show that

$$\frac{\partial \ln L_N}{\partial \beta_0}(\beta) = -\sum_{i=1}^N (S_i - \sigma(X_i;\beta)) \quad \text{and} \quad \frac{\partial \ln L_N}{\partial \beta_1}(\beta) = -\sum_{i=1}^N X_i(S_i - \sigma(X_i;\beta)).$$

Recalling that $V = (1, X)$, by definition of the MLE $\hat{\boldsymbol{\beta}}_N$, we get

$$\sum_{i=1}^N \lambda_3(Z_i, \hat{\theta}_N) = \sum_{i=1}^N V_i(S_i - \sigma(X_i; \hat{\boldsymbol{\beta}}_N)) = 0.$$

Gathering the previous equality and Equation (26), we obtain

$$\sum_{i=1}^N \lambda(Z_i, \hat{\boldsymbol{\theta}}_n) = 0, \tag{27}$$

which proves that $\hat{\theta}_N$ is an M-estimator of type λ. Furthermore, letting $\theta_\infty = (\mathbb{E}_T\left[Y^{(0)}\right], \mathbb{E}_T\left[Y^{(1)}\right], \alpha, \beta_\infty)$, we can compute the following quantities:

$$\begin{aligned}
\mathbb{E}\left[\frac{1-\sigma(X)}{\sigma(X)(1-\alpha)}\frac{S\mathbb{1}\{A=a\}Y}{\pi_a}\right] &= \mathbb{E}\left[\frac{r(X)}{\alpha}\frac{S\mathbb{1}\{A=a\}Y^{(a)}}{\pi_a}\right] \\
&= \frac{1}{\pi_a\alpha}\mathbb{P}(S=1)\mathbb{E}\left[r(X)\mathbb{1}\{A=a\}Y^{(a)}|S=1\right] \\
&= \frac{1}{\pi_a}\mathbb{E}_S\left[r(X)\mathbb{1}\{A=a\}Y^{(a)}\right] \\
&= \mathbb{E}_S\left[r(X)Y^{(a)}\right] \text{ since } \quad r(X) = P_T(X)/P_S(X) \\
&= \mathbb{E}_T\left[Y^{(a)}\right].
\end{aligned}$$

Thus, we have $\mathbb{E}\left[\lambda_0(Z,\theta_\infty)\right] = \mathbb{E}\left[\lambda_1(Z,\theta_\infty)\right] = 0$. Besides,

$$\begin{aligned}
\mathbb{E}\left[\lambda_3(Z,\theta_\infty)\right] &= \mathbb{E}\left[V(S-\sigma(X))\right] \\
&= \mathbb{E}\left[V\cdot\mathbb{E}\left[S-\sigma(X)\mid X\right]\right] && \text{(Law of Total Probability)} \\
&= \mathbb{E}\left[V\cdot(\mathbb{E}\left[S\mid X\right]-\sigma(X))\right] && (\sigma(X) \text{ is a function of } X) \\
&= 0 && \text{(Definition of } \sigma(X))
\end{aligned}$$

Therefore, we have

$$\mathbb{E}\left[\lambda(Z,\theta_\infty)\right] = 0. \tag{28}$$

Now, we show that θ_∞ defined above is the unique value that satisfies (28). We directly have that $\theta_2 = \alpha$. Let

$$L(\beta) = -\,\mathbb{E}\Big[S\,\ln\big(\sigma(X,\beta)\big)\;+\;\big(1-S\big)\,\ln\big(1-\sigma(X,\beta)\big)\Big].$$

A direct calculation shows that

$$\nabla_\beta L(\beta) = \mathbb{E}\Big[V\big(\sigma(X,\beta)-S\big)\Big] \quad \text{and} \quad \nabla^2_\beta L(\beta) = \mathbb{E}\Big[VV^\top\sigma(X,\beta)\big(1-\sigma(X,\beta)\big)\Big].$$

Since $\mathbb{E}[S\mid X] = \sigma(X,\beta_\infty)$, and $V=(1,X)$,

$$\nabla_\beta L(\beta_\infty) = \mathbb{E}\Big[V\left(e(X,\beta_\infty)-S\right)\Big] = 0$$

making β_∞ a stationary point. Furthermore, using overlap we have $\sigma(X,\beta)\big(1-\sigma(X,\beta)\big) \geq \eta^2$ therefore $\forall v \in \mathbb{R}^{p+1}$:

$$\begin{aligned}
v^\top\nabla^2_\beta L(\beta)v &= \mathbb{E}\Big[||V^\top v||_2^2\,\sigma(X,\beta)\left(1-\sigma(X,\beta)\right)\Big] \\
&\geq \eta^2\mathbb{E}\Big[||V^\top v||_2^2\Big] \\
&\geq \eta^2 v^\top\mathbb{E}\Big[V\,V^\top\Big]v.
\end{aligned}$$

Since we assumed that $\mathbb{E}[X\,X^\top]$ is positive-definite, the Hessian $\nabla^2_\beta L(\beta)$ is positive-definite, so $L(\beta)$ is strictly convex. Hence there is a unique global minimizer of $L(\beta)$; since β_∞ is a critical point, it must be that unique minimizer. Consequently, any solution to

$$\mathbb{E}\Big[V\left(\sigma(X,\beta)-S\right)\Big]=0$$

must equal β_∞. Since β, θ_2 are now fixed and since the first two components of ψ are linear with respect to θ_0 and θ_1, θ_∞ is the only value satisfying (28). We want to show that for every θ in a neighborhood of θ_∞, all the components of the second derivatives are integrable for all $k \in \{0,\dots,3\}$:

$$\left|\frac{\partial^2}{\partial^2\theta}\lambda_k(z,\theta)\right|$$

While this holds trivially for most components, the integrability of the following specific terms requires closer attention:

$$\left|\frac{\partial^2}{\partial\theta_2\partial\beta}\lambda_a(z,\theta)\right|,\qquad \left|\frac{\partial^2}{\partial\beta\partial\beta}\lambda_a(z,\theta)\right|,\qquad \left|\frac{\partial^2}{\partial\beta\partial\beta}\lambda_3(z,\theta)\right|\quad \text{for all } a\in\{0,1\}.$$

First, consider the mixed partial derivative with respect to θ_2 and β:

$$\left|\frac{\partial^2}{\partial\theta_2\partial\beta}\lambda_a(z,\theta)\right| = \frac{1}{(1-\theta_2)^2}\cdot\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\cdot\frac{S\mathbb{1}\{A=a\}Y}{\pi_a}V^\top.$$

For each coordinate $i\in\{1,\dots,d\}$, the expectation is bounded as follows:

$$\begin{aligned}\mathbb{E}\left[\left|\frac{\partial^2}{\partial\theta_2\partial\beta}\lambda_a(z,\theta)\right|_i\right] &= \frac{1}{(1-\theta_2)^2}\,\mathbb{E}\left[\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\cdot\frac{S\mathbb{1}\{A=a\}Y}{\pi_a}V_i\right]\\ &\le \frac{1}{(1-\theta_2)^2}\left(\mathbb{E}\left[\left(\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\right)^2\cdot\frac{S\mathbb{1}\{A=a\}Y^2}{\pi_a^2}\right]\cdot\mathbb{E}[V_i^2]\right)^{1/2},\end{aligned}$$

using the Cauchy–Schwarz inequality. The first expectation is finite under the assumption that Y is square-integrable, and due to the exponential tail behavior of the logistic function. The sub-Gaussianity of X implies that all moments of V are finite, ensuring integrability of $\mathbb{E}[V_i^2]$.

Second, consider the pure second derivative with respect to β:

$$\left|\frac{\partial^2}{\partial\beta\partial\beta}\lambda_a(z,\theta)\right| = \frac{1}{1-\theta_2}\cdot\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\cdot\frac{S\mathbb{1}\{A=a\}Y}{\pi_a}VV^\top.$$

Each entry of this matrix takes the form $C\cdot V_iV_j$ for some random coefficient C, and the integrability of these entries follows from the same reasoning as above.

Third, the second derivative of λ_3 is given by

$$\left|\frac{\partial^2}{\partial\beta\partial\beta}\lambda_3(z,\theta)\right| = |V_kV_lV_m\cdot\sigma(X,\beta)(1-\sigma(X,\beta))(1-2\sigma(X,\beta))| \le |V_kV_lV_m|.$$

By applying Hölder's inequality, the following bound is obtained:

$$\mathbb{E}\left[|V_kV_lV_m|\right]\le\mathbb{E}[V_k^2]^{1/2}\cdot\mathbb{E}[V_l^4]^{1/4}\cdot\mathbb{E}[V_m^4]^{1/4},$$

which is finite due to the sub-Gaussianity of X. Consequently, each second derivative component

$$\left|\frac{\partial^2}{\partial^2\theta}\lambda_k(z,\theta)\right|$$

is integrable for all $k\in\{0,\dots,3\}$, in the neighborhood of θ_∞. Define

$$A\left(\theta_\infty\right)=\mathbb{E}\left[\left.\frac{\partial\lambda}{\partial\theta}\right|_{\theta=\theta_\infty}\right]\quad\text{and}\quad B(\theta_\infty)=\mathbb{E}\left[\lambda(Z,\theta_\infty)\lambda(Z,\theta_\infty)^T\right].$$

Next, we verify the conditions of Theorem 7.2 in [42]. To do so, we compute $A\left(\theta_{\infty}\right)$ and $B\left(\theta_{\infty}\right)$. Since

$$\frac{\partial \lambda}{\partial \theta}(Z, \theta)=\begin{pmatrix} -1 & 0 & \frac{1}{(1-\theta_2)^2}\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\mathbb{1}\{A=0\}Y}{\mathbb{P}_S(A=0)} & \frac{-1}{1-\theta_2}\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\mathbb{1}\{A=0\}Y}{\mathbb{P}_S(A=0)}V^\top \\ 0 & -1 & \frac{1}{(1-\theta_2)^2}\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\mathbb{1}\{A=1\}Y}{\mathbb{P}_S(A=1)} & \frac{-1}{1-\theta_2}\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\mathbb{1}\{A=1\}Y}{\mathbb{P}_S(A=1)}V^\top \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -\sigma(X,\beta)(1-\sigma(X,\beta))VV^\top \end{pmatrix}, \tag{29}$$

We obtain with $\theta_{\infty}=\left(\mathbb{E}_T\left[Y^{(0)}\right], \mathbb{E}_T\left[Y^{(1)}\right], \alpha, \beta_{\infty}\right)$

$$A\left(\theta_{\infty}\right)=\begin{pmatrix} -1 & 0 & \frac{\mathbb{E}_T\left[Y^{(0)}\right]}{1-\alpha} & -\mathbb{E}_T\left[Y^{(0)}V^\top\right] \\ 0 & -1 & \frac{\mathbb{E}_T\left[Y^{(1)}\right]}{1-\alpha} & -\mathbb{E}_T\left[Y^{(1)}V^\top\right] \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -Q \end{pmatrix},$$

where $Q=\mathbb{E}\left[\sigma(X, \beta_{\infty})\left(1-\sigma(X, \beta_{\infty})\right) VV^{\top}\right]$, which using Schur complement leads to:

$$A(\theta_{\infty})^{-1}=\begin{pmatrix} -1 & 0 & -\frac{\mathbb{E}_T[Y^{(0)}]}{1-\alpha} & \mathbb{E}_T[Y^{(0)}V^\top]Q^{-1} \\ 0 & -1 & -\frac{\mathbb{E}_T[Y^{(1)}]}{1-\alpha} & \mathbb{E}_T[Y^{(1)}V^\top]Q^{-1} \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -Q^{-1} \end{pmatrix}$$

Regarding $B\left(\theta_{\infty}\right)$, elementary calculations show that

$$B(\theta_{\infty})=\begin{pmatrix} \frac{\mathbb{E}_T\left[r(X)(Y^{(0)})^2\right]}{\alpha\mathbb{P}_S(A=0)}-\mathbb{E}_T\left[Y^{(0)}\right]^2 & -\mathbb{E}_T\left[Y^{(0)}\right]\mathbb{E}_T\left[Y^{(1)}\right] & (1-\alpha)\mathbb{E}_T\left[Y^{(0)}\right] & \mathbb{E}_T\left[(1-\sigma(X))V^\top Y^{(\alpha)}\right] \\ -\mathbb{E}_T\left[Y^{(0)}\right]\mathbb{E}_T\left[Y^{(1)}\right] & \frac{\mathbb{E}_T\left[r(X)(Y^{(1)})^2\right]}{\alpha\mathbb{P}_S(A=1)}-\mathbb{E}_T\left[Y^{(1)}\right]^2 & (1-\alpha)\mathbb{E}_T\left[Y^{(1)}\right] & \mathbb{E}_T\left[(1-\sigma(X))V^\top Y^{(1)}\right] \\ (1-\alpha)\mathbb{E}_T\left[Y^{(0)}\right] & (1-\alpha)\mathbb{E}_T\left[Y^{(1)}\right] & \alpha(1-\alpha) & (1-\alpha)\mathbb{E}_T\left[\sigma(X)V^\top\right] \\ \mathbb{E}_T\left[(1-\sigma(X))VY^{(0)}\right] & \mathbb{E}_T\left[(1-\sigma(X))VY^{(1)}\right] & (1-\alpha)\mathbb{E}_T\left[\sigma(X)V\right] & Q \end{pmatrix}.$$

Based on the previous calculations, we have:

- $\lambda(z, \theta)$ and its first two partial derivatives with respect to θ exist for all z and for all θ in the neighborhood of θ_{∞}.
- For each θ in the neighborhood of θ_{∞}, we have for all $k \in\{0,3\}\left|\frac{\partial^2}{\partial^2 \theta} \lambda_k(z, \theta)\right|$ is integrable.
- $A(\theta_{\infty})$ exists and is nonsingular.
- $B(\theta_{\infty})$ exists and is finite.

We also have

$$\sum_{i=1}^{n} \lambda(Z_i, \hat{\theta}_N)=0 \quad \text{and} \quad \hat{\theta}_N \xrightarrow{p} \theta_{\infty}.$$

Then the conditions of Theorem 7.2 in Stefanski and Boos [42] are satisfied, and we can conclude that

$$\sqrt{n}\left(\hat{\theta}_N-\theta_{\infty}\right) \xrightarrow{d} \mathcal{N}\left(0, A(\theta_{\infty})^{-1} B(\theta_{\infty})(A(\theta_{\infty})^{-1})^{\top}\right),$$

Letting $\nu_0^{\top}$ and $\nu_1^{\top}$ be respectively the first and second row of $A(\theta_{\infty})^{-1}$:

$$\nu_0^{\top}=\left(-1,\, 0,\, -\frac{\mathbb{E}_T[Y^{(0)}]}{1-\alpha},\, \mathbb{E}_T[Y^{(0)}V^\top]Q^{-1}\right),$$

and

$$\nu_1^{\top}=\left(0,\, -1,\, -\frac{\mathbb{E}_T[Y^{(1)}]}{1-\alpha},\, \mathbb{E}_T[Y^{(1)}V^\top]Q^{-1}\right).$$

Expanding the quadratic form explicitly, and using Lemma 1, we get:

$$V_{a,\text{wHT}} = \nu_a^\top B(\theta_\infty)\nu_a = \frac{\mathbb{E}_T[r(X)(Y^{(a)})^2]}{\alpha\pi_a} - \frac{\mathbb{E}_T[Y^{(a)}]^2}{1-\alpha} - \mathbb{E}_T[Y^{(a)}V]^\top Q^{-1}\mathbb{E}_T[Y^{(a)}V] + 2\mathbb{E}_T[Y^{(a)}V]^\top Q^{-1}\mathbb{E}_T[\sigma(X)VY^{(a)}].$$

and:

$$\begin{aligned} C_{\text{wHT}} = \nu_a^\top B(\theta_\infty)\nu_{1-a} = &-\frac{\mathbb{E}_T[Y^{(1)}]\mathbb{E}_T[Y^{(0)}]}{1-\alpha} + \mathbb{E}_T[Y^{(1)}V^\top]Q^{-1}\mathbb{E}_T[Y^{(0)}V] \\ &- \mathbb{E}_T[Y^{(1)}V^\top]Q^{-1}\mathbb{E}_T[(1-\sigma(X))Y^{(0)}V] \\ &- \mathbb{E}_T[Y^{(0)}V^\top]Q^{-1}\mathbb{E}_T[(1-\sigma(X))Y^{(1)}V] \end{aligned}$$

□

Lemma 1. *We have* $\mathbb{E}_T[\sigma(X)V]^\top Q^{-1} = \frac{u_1(d+1)^\top}{1-\alpha}$ *and* $Q = (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)VV^\top\right]$.

Proof.

$$\begin{aligned} Q &= \mathbb{E}\left[\sigma(X,\beta)\left(1-\sigma(X,\beta)\right)VV^\top\right] \\ &= \mathbb{P}(S=1)\mathbb{E}_S\left[\sigma(X,\beta)\left(1-\sigma(X,\beta)\right)VV^\top\right] + \mathbb{P}(S=0)\mathbb{E}_T\left[\sigma(X,\beta)\left(1-\sigma(X,\beta)\right)VV^\top\right] \\ &= (1-\alpha)\mathbb{E}_S\left[\sigma(X,\beta)^2 r(X)VV^\top\right] + (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)\left(1-\sigma(X,\beta)\right)VV^\top\right] \\ &= (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)^2 VV^\top\right] + (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)\left(1-\sigma(X,\beta)\right)VV^\top\right] \\ &= (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)VV^\top\right] \end{aligned}$$

Therefore using using the block inverse matrix formula:

$$Q^{-1} = \frac{1}{1-\alpha}\begin{pmatrix} S^{-1} & -S^{-1}\mathbb{E}_T[\sigma(X)X]^\top P^{-1} \\ -S^{-1}P^{-1}\mathbb{E}_T[\sigma(X)X] & P^{-1} + P^{-1}\mathbb{E}_T[\sigma(X)X]\mathbb{E}_T[\sigma(X)X]^\top P^{-1}S^{-1} \end{pmatrix},$$

where $P = \mathbb{E}_T[\sigma(X)XX^\top]$ and $S = \mathbb{E}_T[\sigma(X)] - \mathbb{E}_T[\sigma(X)X]^\top P^{-1}\mathbb{E}_T[\sigma(X)X]$.

Expanding $\begin{bmatrix}\mathbb{E}_T[\sigma(X)] & \mathbb{E}_T[\sigma(X)X]^\top\end{bmatrix}(1-\alpha)Q^{-1}$, we get:

$$\mathbb{E}_T[\sigma(X)]S^{-1} - S^{-1}\mathbb{E}_T[\sigma(X)X]^\top P^{-1}\mathbb{E}_T[\sigma(X)X] = S^{-1}(\underbrace{\mathbb{E}_T[\sigma(X)] - \mathbb{E}_T[\sigma(X)X]^\top P^{-1}\mathbb{E}_T[\sigma(X)X]}_{S}) = 1$$

and

$$-\mathbb{E}_T[\sigma(X)]S^{-1}\mathbb{E}_T[\sigma(X)X]^\top P^{-1} + \mathbb{E}_T[\sigma(X)X]^\top P^{-1} + \mathbb{E}_T[\sigma(X)X]^\top P^{-1}\mathbb{E}_T[\sigma(X)X]\mathbb{E}_T[\sigma(X)X]^\top P^{-1}S^{-1} = 0$$

Hence:

$$\mathbb{E}_T[\sigma(X)V]^\top Q^{-1} = \frac{u_1(d+1)^\top}{1-\alpha}$$

□

Proposition 10 (asymptotic normality of weighted Horvitz-Thompson estimator)**.** *Let* $\sigma(x,\beta) = \left(1+\exp(-x^\top\beta_1 - \beta_0)\right)^{-1}$ *denote the logistic function, where* $\hat{\beta}_N$ *is the maximum likelihood estimate (MLE) obtained from logistic regression of the selection indicator* S *on covariates* X*. Define the estimated density ratio as:*

$$r(x,\hat{\beta}_N) = \frac{n}{N-n}\cdot\frac{1-\sigma(x,\hat{\beta}_N)}{\sigma(x,\hat{\beta}_N)}, \qquad \textit{with} \quad n = \sum_{i=1}^N S_i.$$

Let $\hat{\tau}_{\Phi,wHT}$ *denote the weighted Horvitz-Thompson estimator constructed using the estimated ratio* $r(x,\hat{\beta}_N)$*. Then, under Assumption 1 to 3.1:*

$$\sqrt{N}\left(\hat{\tau}_{\Phi,wHT} - \tau_\Phi^T\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\Phi,wHT}\right).$$

Proof. We begin by observing that the estimator of interest can be written as smooth transformations of its vector-valued estimator:

$$\hat{\tau}_{\Phi,\text{wHT}} = \Phi(\hat{\psi}_{1,\text{wHT}}, \hat{\psi}_{0,\text{wHT}}).$$

By Propositions 9, the pair $(\hat{\psi}_{1,\text{wHT}}, \hat{\psi}_{0,\text{wHT}})$ is jointly asymptotically normal. Specifically,

$$\sqrt{N}\begin{pmatrix}\hat{\psi}_{1,\text{wHT}} - \psi_1 \\ \hat{\psi}_{0,\text{wHT}} - \psi_0\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\text{wHT}}\right),$$

where Σ_{wHT} is the asymptotic covariance matrices, with entries determined by the variances and covariances of the components $\hat{\psi}_{1,\text{wHT}}$ and $\hat{\psi}_{0,\text{wHT}}$. Since $\Phi : \mathbb{R}^2 \to \mathbb{R}$ is assumed to be a smooth function, we can apply the delta method to each estimator. Let $\nabla\Phi$ denote the gradient of Φ, evaluated at the true parameter vector (ψ_1, ψ_0). Then:

$$\sqrt{N}(\hat{\tau}_{\Phi,\text{wHT}} - \tau_\Phi^T) \xrightarrow{d} \mathcal{N}(0, V_{\Phi,\text{wHT}}),$$

where the asymptotic variance is given by the quadratic form:

$$V_{\Phi,\text{wHT}} = \nabla\Phi^\top \Sigma_{\text{wHT}} \nabla\Phi.$$

□

Proposition 11. *Let $\psi_a = \mathbb{E}_T[Y(a)]$ denote the target population mean potential outcome under treatment $a \in \{0, 1\}$. Define the estimator*

$$\hat{\psi}_{a,N} = \frac{1}{n}\sum_{i=1}^{N} S_i \cdot r(X_i, \hat{\beta}_N) \cdot \frac{\mathbb{1}\{A_i = a\}Y_i}{\hat{\pi}_a}, \quad \textit{with} \quad r(x, \hat{\beta}_N) = \frac{n}{N-n} \cdot \frac{1 - \sigma(x, \hat{\beta}_N)}{\sigma(x, \hat{\beta}_N)},$$

where

$$\hat{\pi}_a = \frac{1}{n}\sum_{S_i=1} \mathbb{1}\{A_i = a\}, \quad n = \sum_{i=1}^{N} S_i,$$

$\hat{\beta}_N$ is the MLE from logistic regression of S on X, and $\sigma(x, \beta) = (1 + e^{-x^\top\beta})^{-1}$ is the logistic function. Then, under regularity conditions,

$$\sqrt{N}\begin{bmatrix}\hat{\psi}_{1,N} - \psi_1 \\ \hat{\psi}_{0,N} - \psi_0\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma_N\right), \quad \textit{with} \quad \Sigma_N = \begin{bmatrix}V_{1,N} & C_N \\ C_N & V_{0,N}\end{bmatrix},$$

where

$$\begin{aligned}
V_{a,N} = {} & \frac{\mathbb{E}_T\left[r(X)\,(Y^{(a)})^2\right]}{\alpha\,\pi_a} + \mathbb{E}_T[Y^{(a)}]^2\left(\frac{2}{\alpha} - \frac{1}{\alpha(1-\alpha)} - \frac{1}{\alpha\pi_a}\right) \\
& - \mathbb{E}_T[Y^{(a)}V]^\top Q^{-1}\mathbb{E}_T[Y^{(a)}V] + 2\,\mathbb{E}_T[Y^{(a)}V]^\top Q^{-1}\mathbb{E}_T[\sigma(X)VY^{(a)}], \\
C_N = {} & \mathbb{E}_T[Y^{(a)}]\mathbb{E}_T[Y^{(1-a)}] \cdot \frac{1-2\alpha}{\alpha(1-\alpha)} \\
& - \mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[(1-\sigma(X))Y^{(1-a)}V] \\
& - \mathbb{E}_T[Y^{(1-a)}V^\top]Q^{-1}\mathbb{E}_T[(1-\sigma(X))Y^{(a)}V] \\
& + \mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[Y^{(1-a)}V],
\end{aligned}$$

with $V = (1, X) \in \mathbb{R}^{p+1}$ and

$$Q = (1-\alpha)\,\mathbb{E}_T[\sigma(X, \beta)\,VV^\top].$$

Proof. We follow the same initial setup and notation as in the proof of the Re-weighted Horvitz-Thomson estimator. The key difference is the empirical estimation of π_a, introducing an additional

estimating equation for θ_3. The remainder of the proof—existence, uniqueness, M-estimation structure, and regularity conditions—follows identically. Define here:

$$\lambda(Z,\boldsymbol{\theta}) = \begin{bmatrix} \frac{\theta_2}{1-\theta_2}\,\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\,\frac{S\,\mathbb{1}\{A=0\}\,Y}{\theta_3} - \theta_0 \\ \frac{\theta_2}{1-\theta_2}\,\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\,\frac{S\,\mathbb{1}\{A=1\}\,Y}{\theta_4} - \theta_1 \\ S - \theta_2 \\ S\mathbb{1}\{A=0\} - \theta_3 \\ S\mathbb{1}\{A=1\} - \theta_4 \\ V(S-\sigma(X,\beta)) \end{bmatrix} \qquad \hat{\boldsymbol{\theta}}_N = \begin{bmatrix} \hat{\psi}_{0,\mathrm{N}} \\ \hat{\psi}_{1,\mathrm{N}} \\ \hat{\alpha} := \frac{1}{N}\sum_{i=1}^N S_i \\ \frac{1}{N}\sum_{i=1}^N S_i\mathbb{1}\{A_i=0\} \\ \frac{1}{N}\sum_{i=1}^N S_i\mathbb{1}\{A_i=1\} \\ \hat{\beta}_N \end{bmatrix}$$

We now rewrite the estimator using this notation. Recall

$$r(X_j,\hat{\beta}_N) = \frac{1-\sigma(X_j,\hat{\beta}_N)}{\sigma(X_j,\hat{\beta}_N)} \cdot \frac{\hat{\alpha}}{1-\hat{\alpha}}.$$

Then:

$$\hat{\psi}_{a,\hat{\pi}} = \frac{1}{N\hat{\alpha}} \sum_{j=1}^N S_j r(X_j,\hat{\beta}_N) \cdot \frac{\mathbb{1}\{A_j=a\}Y_j}{\hat{\pi}_a},$$

and

$$\hat{\pi}_a = \frac{1}{N\hat{\alpha}} \sum_{i=1}^N S_i \mathbb{1}\{A_i=a\}.$$

Substituting in gives:

$$\hat{\psi}_{a,\hat{\pi}} = \frac{1}{N} \sum_{j=1}^N S_j \cdot \frac{1-\sigma(X_j,\hat{\beta}_N)}{\sigma(X_j,\hat{\beta}_N)} \cdot \frac{\hat{\alpha}}{1-\hat{\alpha}} \cdot \frac{\mathbb{1}\{A_j=a\}Y_j}{\frac{1}{N}\sum_{i=1}^N S_i\mathbb{1}\{A_i=a\}}.$$

Let $A(\theta_\infty)$ be the Jacobian of $\lambda(Z,\theta)$ at the population limit θ_∞, and $B(\theta_\infty)$ the corresponding covariance matrix where

$$\theta_\infty = [\mathbb{E}_T[Y^{(0)}], \mathbb{E}_T[Y^{(1)}], \alpha, \alpha\pi_0, \alpha\pi_1, \beta_\infty]^T$$

Let

$$Q = \mathbb{E}_T[\sigma(X,\beta_\infty)(1-\sigma(X,\beta_\infty))VV^\top].$$

The inverse Jacobian block $A(\theta_\infty)^{-1}$ is block lower-triangular, with expressions for rows $\nu_0^\top$, $\nu_1^\top$ given by:

$$\nu_a^\top = \begin{bmatrix} -\mathbb{1}\{a=0\} & -\mathbb{1}\{a=1\} & -\frac{\mathbb{E}_T[Y^{(a)}]}{(1-\alpha)\alpha} & \frac{\mathbb{E}_T[Y^{(a)}]}{\alpha\pi_a} & 0 & \mathbb{E}_T[VY^{(a)}]^\top Q^{-1} \end{bmatrix}.$$

We then compute the asymptotic variance as

$$V_{a,\hat{\pi}} = \nu_a^\top B(\theta_\infty)\nu_a, \quad C_{\hat{\pi}} = \nu_0^\top B(\theta_\infty)\nu_1,$$

where the expansion of $\nu_a^\top B(\theta_\infty)\nu_b$ yields the desired closed-form expressions for $V_{a,\mathrm{N}}$ and C_{N} stated in the proposition. □

B.4 Weighted and transported G-formula

Proposition 12. *Let $\psi_a = \mathbb{E}_T[Y(a)]$ denote the target population mean potential outcome under treatment $a \in \{0,1\}$. Define the oracle estimators*

$$\psi^*_{a,\mathrm{wG}} = \frac{1}{n}\sum_{i=1}^N S_i \cdot r(X_i) \cdot \mu^S_{(a)}(X_i), \quad \textit{and} \quad \psi^*_{a,\mathrm{tG}} = \frac{1}{N-n}\sum_{i=1}^N (1-S_i) \cdot \mu^S_{(a)}(X_i)$$

where $r(X)$ denotes the density ratio between the target and source covariate distributions and $n = \sum_{i=1}^N S_i$ is the number of units in the source sample. Then, under Assumption 1 to 3,

$$\sqrt{N}\left(\psi^*_{a,\mathrm{tG}} - \psi_a\right) \xrightarrow{d} \mathcal{N}(0, V^*_{a,\mathrm{tG}}),$$

Furthermore under Section 3.1:

$$\sqrt{N}\left(\psi^*_{a,\mathrm{wG}} - \psi_a\right) \xrightarrow{d} \mathcal{N}(0, V^*_{a,\mathrm{wG}}),$$

where the asymptotic variances are given by

$$V^*_{a,\mathrm{wG}} = \frac{1}{\alpha}\left(\mathbb{E}_T\left[r(X)(\mu^S_{(a)}(X))^2\right] - \mathbb{E}_T[Y(a)]^2\right)$$

and

$$V^*_{a,\mathrm{tG}} = \frac{1}{1-\alpha}\left(\mathbb{E}_T\left[\left(\mu^S_{(a)}(X)\right)^2\right] - \mathbb{E}_T[Y^{(a)}]^2\right).$$

Proof. **Transported G-formula.** Define

$$Z_i = (1 - S_i) \cdot \mu^S_{(a)}(X),$$

so that

$$\psi^*_{a,\mathrm{tG}} = \frac{\frac{1}{N}\sum_{i=1}^N Z_i}{\frac{1}{N}\sum_{i=1}^N 1 - S_i}.$$

We first compute the expectations of Z_i and S_i. By the definition of Z_i and using that $S_i \sim$ Bernoulli(α),

$$\begin{aligned}\mathbb{E}[Z_i] &= \mathbb{E}\left[(1 - S) \cdot \mu^S_{(a)}(X)\right] \\ &= (1-\alpha)\,\mathbb{E}_T\left[\mu^S_{(a)}(X)\right] \\ &= (1-\alpha)\psi_a,\end{aligned}$$

where we used the consistency assumption $Y = Y^{(A)}$, the randomization of A, and the density ratio property. Since the (Z_i, S_i) for all $i = 1, \ldots, N$ are i.i.d., we can apply the multivariate central limit theorem to

$$\left(\frac{1}{N}\sum_{i=1}^N Z_i, \quad \frac{1}{N}\sum_{i=1}^N 1 - S_i\right),$$

to obtain

$$\sqrt{N}\left(\begin{pmatrix}\frac{1}{N}\sum_{i=1}^N Z_i \\ \frac{1}{N}\sum_{i=1}^N 1 - S_i\end{pmatrix} - \begin{pmatrix}\mathbb{E}[Z_i] \\ \mathbb{E}[1 - S_i]\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix}\mathbb{V}[Z_i] & \mathrm{Cov}(Z_i, S_i) \\ \mathrm{Cov}(S_i, Z_i) & \mathbb{V}[S_i]\end{pmatrix}.$$

since $\psi^*_{a,\mathrm{tG}}$ can be written as

$$\psi^*_{a,\mathrm{tG}} = \frac{\frac{1}{N}\sum_{i=1}^N Z_i}{\frac{1}{N}\sum_{i=1}^N 1 - S_i},$$

we apply the Delta method to the map $h : (x, y) \mapsto x/y$, whose gradient evaluated at $((1-\alpha)\psi_a, 1-\alpha)$ is

$$u = \nabla h((1-\alpha)\psi_a, 1-\alpha) = \left(\frac{1}{1-\alpha}, -\frac{\psi_a}{1-\alpha}\right).$$

Thus,

$$\sqrt{N}\left(\psi^*_{a,\mathrm{tG}} - \psi_a\right) \xrightarrow{d} \mathcal{N}(0, V^*_{a,\mathrm{tG}}),$$

where

$$V^*_{a,\mathrm{tG}} = u^\top \Sigma u.$$

Expanding, we obtain

$$V^*_{a,\mathrm{tG}} = \frac{1}{(1-\alpha)^2}\mathbb{V}[Z_i] + \frac{(\psi_a)^2}{(1-\alpha)^2}\mathbb{V}[S_i] - \frac{2\psi_a}{(1-\alpha)^2}\mathrm{Cov}(Z_i, S_i).$$

It remains to compute each term. Since $S_i \sim \text{Bernoulli}(\alpha)$,

$$\mathbb{V}[S_i] = \alpha(1-\alpha).$$

By direct computation,

$$\mathbb{V}[Z_i] = (1-\alpha)\,\mathbb{E}_T\left[\left(\mu^S_{(a)}(X)\right)^2\right] - ((1-\alpha)\psi_a)^2.$$

Moreover, since $(1-S_i)\times Z_i = Z_i$, we have

$$\text{Cov}(Z_i, S_i) = (1-\alpha)\alpha\psi_a.$$

Substituting into the expression for $V^*_{a,\text{tG}}$, we find

$$V^*_{a,\text{tG}} = \frac{1}{1-\alpha}\left(\mathbb{E}_T\left[\left(\mu^S_{(a)}(X)\right)^2\right] - (\psi_a)^2\right).$$

weighted G-formula. Define

$$Z_i = S_i \cdot r(X_i) \cdot \mu^S_{(a)}(X_i),$$

so that

$$\psi^*_{a,\text{wG}} = \frac{\frac{1}{N}\sum_{i=1}^N Z_i}{\frac{1}{N}\sum_{i=1}^N S_i} = \frac{\bar{Z}_N}{\bar{S}_N}.$$

We first compute the expectations of Z_i and S_i. Using the definition of Z_i and that S is binary,

$$\begin{aligned}\mathbb{E}[Z_i] &= \mathbb{E}\left[S_i \cdot r(X) \cdot \mu^S_{(a)}(X)\right] \\ &= \alpha\,\mathbb{E}_S\left[r(X)\cdot\mu^S_{(a)}(X)\right] \\ &= \alpha\,\mathbb{E}_T\left[\mu^S_{(a)}(X)\right] = \alpha\psi_a,\end{aligned}$$

where we used the importance sampling identity $\mathbb{E}_S[r(X)f(X)] = \mathbb{E}_T[f(X)]$. By the multivariate central limit theorem,

$$\sqrt{N}\left(\begin{pmatrix}\bar{Z}_N\\ \bar{S}_N\end{pmatrix} - \begin{pmatrix}\alpha\psi_a\\ \alpha\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0,\Sigma),$$

where

$$\Sigma = \begin{pmatrix}\mathbb{V}[Z_i] & \text{Cov}(Z_i,S_i)\\ \text{Cov}(Z_i,S_i) & \mathbb{V}[S_i]\end{pmatrix}.$$

The gradient of the map $h:(x,y)\mapsto x/y$ evaluated at $(\alpha\psi_a,\alpha)$ is

$$u = \nabla h(\alpha\psi_a,\alpha) = \left(\frac{1}{\alpha},\, -\frac{\psi_a}{\alpha}\right).$$

By the Delta method,

$$\sqrt{N}\left(\psi^*_{a,\text{wG}} - \psi_a\right) \xrightarrow{d} \mathcal{N}(0, V^*_{a,\text{wG}}),$$

where

$$V^*_{a,\text{wG}} = u^\top\Sigma u = \frac{1}{\alpha^2}\mathbb{V}[Z_i] + \frac{\psi_a^2}{\alpha^2}\mathbb{V}[S_i] - \frac{2\psi_a}{\alpha^2}\text{Cov}(Z_i,S_i).$$

We compute each term:

$$\begin{aligned}\mathbb{V}[Z_i] &= \mathbb{E}[Z_i^2] - (\mathbb{E}[Z_i])^2 = \alpha\,\mathbb{E}_T\left[r(X)\left(\mu^S_{(a)}(X)\right)^2\right] - \alpha^2\psi_a^2,\\ \mathbb{V}[S_i] &= \alpha(1-\alpha),\\ \text{Cov}(Z_i,S_i) &= \mathbb{E}[Z_iS_i] - \mathbb{E}[Z_i]\mathbb{E}[S_i] = \alpha\,\mathbb{E}_T\left[\mu^S_{(a)}(X)\right] - \alpha^2\psi_a = \alpha(1-\alpha)\psi_a.\end{aligned}$$

Substituting into the expression for $V^*_{a,\mathrm{wG}}$, we get:

$$\begin{aligned}
V^*_{a,\mathrm{wG}} &= \frac{1}{\alpha^2}\left(\alpha\,\mathbb{E}_T\left[r(X)(\mu^S_{(a)}(X))^2\right] - \alpha^2\psi_a^2\right) + \frac{\psi_a^2}{\alpha^2}\alpha(1-\alpha) - \frac{2\psi_a}{\alpha^2}\alpha(1-\alpha)\psi_a \\
&= \frac{1}{\alpha}\mathbb{E}_T\left[r(X)(\mu^S_{(a)}(X))^2\right] - \psi_a^2 + \frac{\psi_a^2(1-\alpha)}{\alpha} - \frac{2\psi_a^2(1-\alpha)}{\alpha} \\
&= \frac{1}{\alpha}\mathbb{E}_T\left[r(X)(\mu^S_{(a)}(X))^2\right] - \frac{\psi_a^2}{\alpha}.
\end{aligned}$$

□

Proposition 13. *Grant Assumption 1 to 1 defining $\beta^{(a)} := [c^{(a)}, \gamma^{(a)}]$, $V := [1, X]$. We rearrange the source Y_i and V_i so that the first n_1 observations of correspond to $A = 1$. We then define $\mathbf{Y}_1 = (Y_1, \ldots, Y_{n_1})^\top$ and $\mathbf{Y}_0 = (Y_{n_1+1}, \ldots, Y_n)^\top$, as well as $\mathbf{V}_1 = (V_1, \ldots, V_{n_1})^\top$ and $\mathbf{V}_0 = (V_{n_1+1}, \ldots, V_n)^\top$. Letting $\hat{\alpha} = (\sum_{i=1}^n S_i)/N$ and for all $a \in \{0, 1\}$,*

$$\bar{V}^{(0)} = \frac{1}{\sum_{i=1}^n \mathbb{1}_{S_i=0}} \sum_{i=1}^n \mathbb{1}_{S_i=0} V_i \quad \textit{and} \quad \hat{\beta}^{(a)} = \left(\frac{1}{n_a}\mathbf{V}_a^\top \mathbf{V}_a\right)^{-1} \frac{1}{n_a}\mathbf{V}_a^\top \mathbf{Y}_a.$$

Defining $\nu = \mathbb{E}_S[X]$ and $\Sigma = Var(X|S = 1)$, we have

$$\sqrt{N}(\hat{\theta}_N - \theta_\infty) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where

$$\hat{\theta}_N = \begin{pmatrix} \bar{V}_{(0)} \\ \hat{\beta}^{(0)} \\ \hat{\beta}^{(1)} \end{pmatrix}, \quad \theta_\infty = \begin{pmatrix} \mathbb{E}_T[V] \\ \beta^{(0)} \\ \beta^{(1)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \frac{\mathrm{Var}[V|S=0]}{(1-\alpha)} & 0 & 0 \\ 0 & \frac{\sigma^2 M^{-1}}{\alpha(1-\pi)} & 0 \\ 0 & 0 & \frac{\sigma^2 M^{-1}}{\alpha\pi} \end{pmatrix},$$

with $M^{-1} = \begin{bmatrix} 1 + \nu^T\Sigma^{-1}\nu & -\nu^T\Sigma^{-1} \\ -\Sigma^{-1}\nu & \Sigma^{-1} \end{bmatrix}$.

Proof. Using M-estimation theory to prove asymptotic normality of the θ_N, we first define the following:

$$\lambda(Z, \theta) = \begin{pmatrix} \lambda_1(Z,\theta) \\ \lambda_2(Z,\theta) \\ \lambda_3(Z,\theta) \end{pmatrix} := \begin{pmatrix} (1-S)(V - \theta_0) \\ S(1-A)\left(V\epsilon^{(0)} - VV^\top\left(\theta_1 - \beta^{(0)}\right)\right) \\ SA\left(V\epsilon^{(1)} - VV^\top\left(\theta_2 - \beta^{(1)}\right)\right) \end{pmatrix}$$

where $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)$. We still have that $\hat{\theta}_N$ is an M-estimator of type λ [see 42] since

$$\sum_{i=1}^N \lambda(Z_i, \hat{\theta}_N) = 0.$$

Note that

$$\begin{aligned}
\mathbb{E}\left[\lambda_1(Z, \theta_\infty)\right] &= \mathbb{E}\left[(1-S)\left(V - \mathbb{E}_T[V]\right)\right] \\
&= (1-\alpha)\mathbb{E}_T\left[(V - \mathbb{E}_T[V])\right] \\
&= 0.
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E}\left[\lambda_2(Z, \theta_\infty)\right] &= \mathbb{E}\left[S(1-A)V\epsilon^{(0)}\right] \\
&= \alpha\mathbb{E}_S\left[(1-A)V\epsilon^{(0)}\right] \\
&= \alpha(1-\pi)\mathbb{E}_S\left[V\epsilon^{(0)}\right] \\
&= \alpha(1-\pi)\mathbb{E}_S\left[V\mathbb{E}_S\left[\epsilon^{(0)}|V\right]\right] \\
&= 0.
\end{aligned}$$

Similarly, we can show that $\mathbb{E}\left[\lambda_3(Z, \theta_\infty)\right] = 0$. Since $\lambda(Z, \theta)$ is a linear function of θ, θ_∞ is the only value of θ such that $\mathbb{E}\left[\lambda(Z, \theta)\right] = 0$ Define

$$A\left(\theta_\infty\right) = \mathbb{E}\left[\frac{\partial \lambda}{\partial \theta}\Big|_{\theta=\theta_\infty}\right] \quad \text{and} \quad B(\theta_\infty) = \mathbb{E}\left[\lambda(Z, \theta_\infty)\lambda(Z, \theta_\infty)^T\right].$$

Next, we check the conditions of Theorem 7.2 in Stefanski and Boos [42]. First, we compute $A\left(\theta_\infty\right)$ and $B\left(\theta_\infty\right)$. Since

$$\frac{\partial \lambda}{\partial \theta}(Z, \theta) = \begin{pmatrix} -(1-S) & 0 & 0 \\ 0 & -S(1-A)VV^\top & 0 \\ 0 & 0 & -SAVV^\top \end{pmatrix},$$

we obtain

$$A\left(\theta_\infty\right) = \begin{pmatrix} -(1-\alpha) & 0 & 0 \\ 0 & -\alpha(1-\pi)M & 0 \\ 0 & 0 & -\alpha\pi M \end{pmatrix},$$

where $M = \mathbb{E}_S\left[VV^\top\right]$, which leads to

$$A^{-1}\left(\theta_\infty\right) = \begin{pmatrix} -\frac{1}{1-\alpha} & 0 & 0 \\ 0 & -\frac{M^{-1}}{\alpha(1-\pi)} & 0 \\ 0 & 0 & -\frac{M^{-1}}{\alpha\pi} \end{pmatrix}.$$

Regarding $B(\theta_\infty)$, since we have $A(1-A) = 0$ and $S(1-S) = 0$, elementary calculations show that:

$$B(\theta_\infty)_{2,3} = B(\theta_\infty)_{3,2} = 0 \quad \text{and} \quad \begin{matrix} B(\theta_\infty)_{1,2} = B(\theta_\infty)_{2,1} = 0 \\ B(\theta_\infty)_{1,3} = B(\theta_\infty)_{3,1} = 0. \end{matrix}$$

Besides

$$\begin{aligned} B(\theta_\infty)_{1,1} &= \mathbb{E}\left[(1-S)^2(V - \mathbb{E}_T\left[V\right])(V - \mathbb{E}_T\left[V\right])^\top\right] \\ &= (1-\alpha)\operatorname{Var}\left[V|S=0\right], \end{aligned}$$

We can also note that:

$$\begin{aligned} B(\theta_\infty)_{3,3} &= \mathbb{E}\left[S^2A^2VV^\top(\epsilon^{(1)})^2\right] \\ &= \alpha\pi\mathbb{E}_S\left[VV^\top(\epsilon^{(1)})^2|A=1\right] \\ &= \alpha\pi\sigma^2M \end{aligned}$$

and similarly,

$$B(\theta_\infty)_{2,2} = \alpha(1-\pi)\sigma^2M.$$

Gathering all calculations, we have

$$B(\theta_\infty) = \begin{pmatrix} (1-\alpha)\operatorname{Var}\left[V|S=0\right] & 0 & 0 \\ 0 & \alpha(1-\pi)\sigma^2M & 0 \\ 0 & 0 & \alpha\pi\sigma^2M \end{pmatrix},$$

Based on the previous calculations, we have:

- $\lambda(z, \theta)$ and its first two partial derivatives with respect to θ exist for all z and for all θ in the neighborhood of θ_∞.
- For each θ in the neighborhood of θ_∞, we have for all $i, j, k \in \{1, 3\}$:

$$\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\lambda_k(z, \theta)\right| \leq 1.$$

- $A(\theta_\infty)$ exists and is nonsingular.
- $B(\theta_\infty)$ exists and is finite.

Since we have:

$$\sum_{i=1}^{n} \lambda(T_i, Z_i, \hat{\theta}_N) = 0 \quad \text{and} \quad \hat{\theta}_N \xrightarrow{p} \theta_\infty.$$

Then, the conditions of Theorem 7.2 in Stefanski and Boos [42] are satisfied, we have:

$$\sqrt{n}\left(\hat{\theta}_N - \theta_\infty\right) \xrightarrow{d} \mathcal{N}\left(0, A(\theta_\infty)^{-1} B(\theta_\infty)(A(\theta_\infty)^{-1})^\top\right),$$

where:

$$A(\theta_\infty)^{-1} B(\theta_\infty)(A(\theta_\infty)^{-1})^\top = \begin{pmatrix} \frac{\text{Var}[V|S=0]}{(1-\alpha)} & 0 & 0 \\ 0 & \frac{\sigma^2 M^{-1}}{\alpha(1-\pi)} & 0 \\ 0 & 0 & \frac{\sigma^2 M^{-1}}{\alpha\pi} \end{pmatrix}.$$

□

Corollary 1. *For all $a \in \{0,1\}$, let $\hat{\tau}^{OLS}_{a,\text{tG}}$ denote the transported G-formula estimator where linear regressions are used to estimate $\mu^S_{(a)}$. Then, under Assumption 1 to 3 and 1:*

$$\sqrt{N}\begin{pmatrix} \hat{\tau}^{OLS}_{1,\text{tG}} - \psi_1 \\ \hat{\tau}^{OLS}_{0,\text{tG}} - \psi_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \Sigma^{OLS}_{\text{tG}}\right),$$

with

$$\Sigma^{OLS}_{\text{tG}} = \begin{pmatrix} (\beta^{(1)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(1)} + \mathbb{E}_T[V]^\top \frac{\sigma^2 M^{-1}}{\alpha\pi}\mathbb{E}_T[V] & (\beta^{(1)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(0)} \\ (\beta^{(1)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(0)} & (\beta^{(0)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(0)} + \mathbb{E}_T[V]^\top \frac{\sigma^2 M^{-1}}{\alpha(1-\pi)}\mathbb{E}_T[V] \end{pmatrix},$$

Proof. Recall that for $a \in \{0,1\}$, we have:

$$\hat{\tau}^{\text{OLS}}_{a,\text{tG}} = (\hat{\beta}^{(a)})^\top \bar{V}_{(0)}$$

From Proposition 13, we know that:

$$\sqrt{N}(\hat{\theta}_N - \theta_\infty) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

with $\hat{\theta}_N = (\bar{V}_{(0)}^\top, (\hat{\beta}^{(0)})^\top, (\hat{\beta}^{(1)})^\top)^\top \in \mathbb{R}^{3p+3}$, and where Σ is the block-diagonal covariance matrix. By applying the delta method to the map

$$g(\bar{V}_{(0)}, \hat{\beta}^{(0)}, \hat{\beta}^{(1)}) = \begin{pmatrix} (\hat{\beta}^{(1)})^\top \bar{V}_{(0)} \\ (\hat{\beta}^{(0)})^\top \bar{V}_{(0)} \end{pmatrix},$$

the asymptotic distribution of $\sqrt{N}(\hat{\tau}^{\text{OLS}}_{a,\text{tG}} - \psi_a)$ is multivariate normal with covariance matrix:

$$\Sigma^{\text{OLS}}_{\text{tG}} = J\Sigma J^\top,$$

where J is the Jacobian of g evaluated at $(\mathbb{E}_T[V], \beta^{(0)}, \beta^{(1)})$:

$$J = \begin{pmatrix} (\beta_{(1)})^\top & 0 & \mathbb{E}_T[V]^\top \\ (\beta_{(0)})^\top & \mathbb{E}_T[V]^\top & 0 \end{pmatrix}.$$

we obtain:

$$\Sigma^{\text{OLS}}_{\text{tG}} = \begin{pmatrix} (\beta^{(1)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(1)} + \mathbb{E}_T[V]^\top \frac{\sigma^2 M^{-1}}{\alpha\pi}\mathbb{E}_T[V] & (\beta^{(1)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(0)} \\ (\beta^{(1)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(0)} & (\beta^{(0)})^\top \frac{\text{Var}[V|S=0]}{1-\alpha}\beta^{(0)} + \mathbb{E}_T[V]^\top \frac{\sigma^2 M^{-1}}{\alpha(1-\pi)}\mathbb{E}_T[V] \end{pmatrix}, \tag{30}$$

□

Proposition 14. *For all $a \in \{0,1\}$, let $\hat{\tau}^{OLS}_{a,\mathrm{wG}}$ denote the weighted G-formula estimators where the density ratio is estimated using a logistic regression and linear regressions are used to estimate $\mu^S_{(a)}$. Then, under Assumption 1 to 1:*

$$\sqrt{N}\left(\begin{pmatrix}\hat{\psi}^{OLS}_{0,\mathrm{wG}}\\ \hat{\psi}^{OLS}_{1,\mathrm{wG}}\end{pmatrix} - \begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma^{OLS}_{\mathrm{wG}}).$$

Proof. Let $Z = (S, X, S \times A, S \times Y)$ and define $\theta = (\theta_1, \theta_2, \theta_3, \beta, \theta_4, \theta_5)$. Consider the estimating function $\lambda(Z, \theta)$ defined as:

$$\lambda(Z,\theta) = \begin{pmatrix} \frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\theta_4^\top V}{1-\theta_3} - \theta_1, \\ \frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}\frac{S\theta_5^\top V}{1-\theta_3} - \theta_2, \\ S - \theta_3, \\ V(S - \sigma(X,\beta)), \\ S(1-A)\left(V\epsilon(0) - VV^\top(\theta_4 - \beta^{(0)})\right), \\ SA\left(V\epsilon(1) - VV^\top(\theta_5 - \beta^{(1)})\right) \end{pmatrix}$$

Define $\hat{\theta}_N = (\hat{\psi}^{\mathrm{OLS}}_{0,\mathrm{wG}}, \hat{\psi}^{\mathrm{OLS}}_{1,\mathrm{wG}}, \hat{\alpha}, \hat{\beta}_N, \hat{\beta}^{(0)}, \hat{\beta}^{(1)})$, where $\hat{\beta}_N$ is the MLE from the logistic regression of S on X, and $r(X_i, \hat{\beta}_N) = \frac{1-\sigma(X_i,\hat{\beta}_N)}{\sigma(X_i,\hat{\beta}_N)} \cdot \frac{\hat{\alpha}}{1-\hat{\alpha}}$. Then, the estimators $\hat{\psi}^{\mathrm{OLS}}_{a,\mathrm{wG}}$ take the form:

$$\hat{\psi}^{\mathrm{OLS}}_{a,\mathrm{wG}} = \frac{1}{N}\sum_{i=1}^{N} S_i \cdot \frac{1-\sigma(X_i,\hat{\beta}_N)}{\sigma(X_i,\hat{\beta}_N)(1-\hat{\alpha})} \cdot (\hat{\beta}^{(a)})^\top V_i.$$

The estimator $\hat{\theta}_N$ solves the estimating equation:

$$\sum_{i=1}^{N} \lambda(Z_i, \hat{\theta}_N) = 0.$$

This setup is structurally identical to that of proposition 9 and 13, and the M-estimation theory in Stefanski and Boos [42] applies directly. Specifically, the regularity conditions (e.g., smoothness of λ, identifiability, uniqueness of root) are satisfied.

As before, the asymptotic distribution of the M-estimator is:

$$\sqrt{N}(\hat{\theta}_N - \theta_\infty) \xrightarrow{d} \mathcal{N}(0, A^{-1}B(A^{-1})^\top),$$

where A and B are the Jacobian of the estimating function and its variance, respectively, evaluated at $\theta_\infty = (\psi_0, \psi_1, \alpha, \beta_\infty, \beta^{(0)}, \beta^{(1)})$. Focusing on the top-left 2×2 block of the sandwich covariance matrix—corresponding to $(\hat{\psi}^{\mathrm{OLS}}_{0,\mathrm{wG}}, \hat{\psi}^{\mathrm{OLS}}_{1,\mathrm{wG}})$—we denote this block by $\Sigma^{\mathrm{OLS}}_{\mathrm{wG}}$, and conclude:

$$\sqrt{N}\begin{pmatrix}\hat{\psi}^{\mathrm{OLS}}_{0,\mathrm{wG}} - \psi_0\\ \hat{\psi}^{\mathrm{OLS}}_{1,\mathrm{wG}} - \psi_1\end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma^{\mathrm{OLS}}_{\mathrm{wG}}) \quad \text{where} \quad \Sigma^{\mathrm{OLS}}_{\mathrm{wG}} = \begin{pmatrix} V^{\mathrm{OLS}}_{0,\mathrm{wG}} & C^{\mathrm{OLS}}_{\mathrm{wG}} \\ C^{\mathrm{OLS}}_{\mathrm{wG}} & V^{\mathrm{OLS}}_{1,\mathrm{wG}}\end{pmatrix},$$

where using Lemma 1, we simplify the expression to:

$$V^{\mathrm{OLS}}_{a,\mathrm{wG}} = \frac{(\beta^{(a)})^\top \Delta \beta^{(a)}}{\alpha} - \frac{\psi_a^2}{1-\alpha} + 2\frac{(\beta^{(a)})^\top \mathbb{E}_T[VV^\top]\beta^{(a)}}{1-\alpha} - (\beta^{(a)})^\top \mathbb{E}_T[VV^\top]Q^{-1}\mathbb{E}_T[VV^\top]\beta^{(a)} \tag{31}$$

$$+ \sigma^2 \cdot \frac{\mathbb{E}_T[V]^\top M^{-1}\mathbb{E}_T[V]}{\alpha\pi_a}. \tag{32}$$

For the covariance terms we get:

$$C^{\mathrm{OLS}}_{\mathrm{wG}} = \frac{(\beta^{(0)})^\top \Delta \beta^{(1)}}{\alpha} - \frac{\psi_1\psi_0}{1-\alpha} + 2\frac{(\beta^{(1)})^\top \mathbb{E}_T[VV^\top]\beta^{(0)}}{1-\alpha} - (\beta^{(0)})^\top \mathbb{E}_T[VV^\top]Q^{-1}\mathbb{E}_T[VV^\top]\beta^{(1)}, \tag{33}$$

where $\Delta = \mathbb{E}_T[r(X)VV^\top]$. □

Lemma 2. *Grant Assumption 1 to Linear model 1, then:*

- *the matrix* $\Sigma^{OLS}_{\text{wG}} - \Sigma^{OLS}_{\text{tG}}$ *is semi-definite positive,*
- *the matrix* $\Sigma_{wHT} - \Sigma^{OLS}_{\text{wG}}$ *is semi-definite positive.*

Consequently, $V^{OLS}_{\Phi,\text{tG}} \leq V^{OLS}_{\Phi,\text{wG}} \leq V_{\Phi,wHT}$.

Proof. **First statement** We first start with $\Sigma_{\text{wHT}} - \Sigma^{\text{OLS}}_{\text{wG}}$. Under Linear model 1

$$Y^{(a)} = (\beta^{(a)})^\top V + \epsilon^{(a)}, \quad \text{where} \quad \mathbb{E}[\epsilon^{(a)} \mid X] = 0, \quad \text{Var}(\epsilon^{(a)} \mid X) = \sigma^2,$$

we can express the variances $V_{a,\text{wHT}}$ and the covariance term C_{wHT} defined in 19 of the weighted Horvitz-thompson in terms of the model parameters and the distribution of X. Starting with $V_{a,\text{wHT}}$, we expand each expectation by substituting the linear form of $Y^{(a)}$. The first term of $V_{a,\text{wHT}}$ becomes

$$\mathbb{E}_T[r(X)(Y^{(a)})^2] = \mathbb{E}_T\left[r(X)\left(((\beta^{(a)})^\top V)^2 + 2(\beta^{(a)})^\top V\epsilon^{(a)} + (\epsilon^{(a)})^2\right)\right].$$

Taking expectations and applying the assumptions $\mathbb{E}[\epsilon^{(a)} \mid X] = 0$ and $\text{Var}(\epsilon^{(a)} \mid X) = \sigma^2$, this simplifies to

$$\mathbb{E}_T[r(X)(Y^{(a)})^2] = \mathbb{E}_T[r(X)((\beta^{(a)})^\top V)^2] + \sigma^2\mathbb{E}_T[r(X)].$$

The second term, becomes $(\mathbb{E}_T[(\beta^{(a)})^\top X])^2$, since the error has mean zero. Moving to the third term, we use linearity to write

$$\mathbb{E}_T[Y^{(a)}V^\top] = (\beta^{(a)})^\top\mathbb{E}_T[VV^\top],$$

and thus the quadratic form becomes

$$\mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[Y^{(a)}V] = (\beta^{(a)})^\top\mathbb{E}_T[VV^\top]Q^{-1}\mathbb{E}_T[VV^\top]\beta^{(a)}.$$

For the fourth term, we compute

$$\mathbb{E}_T[\sigma(X)Y^{(a)}V] = \mathbb{E}_T[\sigma(X)VV^\top]\beta^{(a)},$$

which leads to

$$2\,\mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[\sigma(X)Y^{(a)}V] = 2\,(\beta^{(a)})^\top\mathbb{E}_T[VV^\top]Q^{-1}\mathbb{E}_T[\sigma(X)VV^\top]\beta^{(a)}.$$

Noting that $Q = (1-\alpha)\mathbb{E}_T\left[\sigma(X,\beta)VV^\top\right]$, this further simplifies to:

$$2\,\mathbb{E}_T[Y^{(a)}V^\top]Q^{-1}\mathbb{E}_T[\sigma(X)Y^{(a)}V] = \frac{2\,(\beta^{(a)})^\top\mathbb{E}_T[VV^\top]\beta^{(a)}}{1-\alpha}.$$

Combining all components, using linearity and the definition of Δ, we have

$$\begin{aligned} V_{a,\text{wHT}} = \frac{1}{\alpha\pi_a}\left((\beta^{(a)})^\top\Delta\beta^{(a)} + \sigma^2\mathbb{E}_T[r(X)]\right) - \frac{(\mathbb{E}_T[(\beta^{(a)})^\top V])^2}{1-\alpha} \\ - (\beta^{(a)})^\top\mathbb{E}_T[VV^\top]Q^{-1}\mathbb{E}_T[VV^\top]\beta^{(a)} + 2\,\frac{(\beta^{(a)})^\top\mathbb{E}_T[VV^\top]\beta^{(a)}}{1-\alpha}. \end{aligned}$$

Turning to the cross-covariance term C_{wHT}, we proceed similarly. Noting that

$$\begin{aligned} \mathbb{E}_T[Y^{(a)}] &= \mathbb{E}_T[(\beta^{(a)})^\top V], \\ \mathbb{E}_T[Y^{(a)}V^\top] &= (\beta^{(a)})^\top\mathbb{E}_T[VV^\top], \\ \mathbb{E}_T[\sigma(X)Y^{(a)}V] &= \mathbb{E}_T[\sigma(X)VV^\top]\beta^{(a)}, \end{aligned}$$

we substitute these into the original definition to obtain

$$\begin{aligned} C_{\text{wHT}} = &-\frac{\mathbb{E}_T[(\beta^{(1)})^\top V]\cdot\mathbb{E}_T[(\beta^{(0)})^\top V]}{1-\alpha} + 2\frac{(\beta^{(1)})^\top\mathbb{E}_T[VV^\top]\beta^{(0)}}{1-\alpha} \\ &- (\beta^{(1)})^\top\mathbb{E}_T[VV^\top]Q^{-1}\mathbb{E}_T[VV^\top]\beta^{(0)}. \end{aligned}$$

Now, we can compute the following quantities:

$$V_{a,\text{wHT}} - V_{a,\text{wG}}^{\text{OLS}} = \frac{1-\pi_a}{\alpha\pi_a}(\beta^{(a)})^\top \Delta\beta^{(a)} + \frac{\sigma^2}{\alpha\pi_a}\left(\mathbb{E}_T[r(X)] - \mathbb{E}_T[V]^\top M^{-1}\mathbb{E}_T[V]\right)$$

and

$$C_{\text{wHT}} - C_{\text{wG}}^{\text{OLS}} = -\frac{(\beta^{(0)})^\top \Delta\beta^{(1)}}{\alpha}.$$

Since $0 < \pi_a < 1$, we immediately have that

$$\frac{1-\pi_a}{\alpha\pi_a}(\beta^{(a)})^\top \Delta\beta^{(a)} \geq 0.$$

Now, turning to the second term in the expression for $V_{a,\text{wHT}} - V_{a,\text{wG}}^{\text{OLS}}$, observe that

$$\mathbb{E}_T[r(X)] - \mathbb{E}_T[V]^\top M^{-1}\mathbb{E}_T[V]$$

can be rewritten using the fact that the target distribution T is defined via reweighting from the source distribution S, with $r(X)$. This yields:

$$\mathbb{E}_T[r(X)] - \mathbb{E}_T[V]^\top M^{-1}\mathbb{E}_T[V] = \mathbb{E}_S[r(X)^2] - \mathbb{E}_S[r(X)V]^\top \mathbb{E}_S[VV^\top]^{-1}\mathbb{E}_S[r(X)V].$$

To interpret this expression, we recognize it as a variance-type quantity. In particular, we can rewrite it as:

$$\begin{aligned}&\mathbb{E}_S[r(X)^2] - 2\mathbb{E}_S[r(X)V]^\top \mathbb{E}_S[VV^\top]^{-1}\mathbb{E}_S[r(X)V] \\ &\quad + \mathbb{E}_S[r(X)V]^\top \mathbb{E}_S[VV^\top]^{-1}\mathbb{E}_S[VV^\top]\mathbb{E}_S[VV^\top]^{-1}\mathbb{E}_S[r(X)V],\end{aligned}$$

which simplifies to:

$$\mathbb{E}_S\left[\left(r(X) - \mathbb{E}_S[r(X)V]^\top \mathbb{E}_S[VV^\top]^{-1}V\right)^2\right].$$

This is the expected squared residual from projecting $r(X)$ onto the linear span of V, and is therefore nonnegative. Hence,

$$\mathbb{E}_T[r(X)] - \mathbb{E}_T[V]^\top M^{-1}\mathbb{E}_T[V] \geq 0.$$

Combining this result with the earlier inequality, we conclude that

$$V_{a,\text{wHT}} - V_{a,\text{wG}}^{\text{OLS}} \geq 0.$$

Since this holds for both $a = 0$ and $a = 1$, and noting that

$$C_{\text{wHT}} - C_{\text{wG}}^{\text{OLS}} = -\frac{(\beta^{(0)})^\top \Delta\beta^{(1)}}{\alpha},$$

we now analyze the overall matrix difference. The trace satisfies:

$$\text{tr}(\Sigma_{\text{wHT}} - \Sigma_{\text{wG}}^{\text{OLS}}) = (V_{1,\text{wHT}} - V_{1,\text{wG}}^{\text{OLS}}) + (V_{0,\text{wHT}} - V_{0,\text{wG}}^{\text{OLS}}) \geq 0,$$

and the determinant becomes:

$$\begin{aligned}\det(\Sigma_{\text{wHT}} - \Sigma_{\text{wG}}^{\text{OLS}}) &= (V_{1,\text{wHT}} - V_{1,\text{wG}}^{\text{OLS}})(V_{0,\text{wHT}} - V_{0,\text{wG}}^{\text{OLS}}) - (C_{\text{wHT}} - C_{\text{wG}}^{\text{OLS}})^2 \\ &\geq \frac{1}{\alpha^2}\left((\beta^{(1)})^\top \Delta\beta^{(1)} \cdot (\beta^{(0)})^\top \Delta\beta^{(0)} - ((\beta^{(1)})^\top \Delta\beta^{(0)})^2\right) \geq 0,\end{aligned}$$

where the final inequality follows from Cauchy–Schwarz. Since both the trace and determinant of $\Sigma_{\text{wHT}} - \Sigma_{\text{wG}}^{\text{OLS}}$ are nonnegative, we conclude that the matrix $\Sigma_{\text{wHT}} - \Sigma_{\text{wG}}^{\text{OLS}}$ is positive semi-definite.

Second statement Now, we want to prove that $\Sigma_{\text{wG}}^{\text{OLS}} - \Sigma_{\text{tG}}^{\text{OLS}}$ is semi-definite positive. Observe that using the expression defined in 19, 31 and 30 we have

$$V_{a,\text{wG}}^{\text{OLS}} - V_{a,\text{tG}}^{\text{OLS}} = (\beta^{(a)})^\top H\beta^{(a)} \quad \text{and} \quad C_{\text{wG}}^{\text{OLS}} - C_{\text{tG}}^{\text{OLS}} = (\beta^{(1)})^\top H\beta^{(0)},$$

where the matrix H is defined as

$$H := \frac{1}{1-\alpha}\left(\mathbb{E}_T\left[\frac{1-\sigma(X,\beta)}{\sigma(X,\beta)}VV^\top\right] + \mathbb{E}_T[VV^\top] - \mathbb{E}_T[VV^\top]\left(\mathbb{E}_T[\sigma(X,\beta)VV^\top]\right)^{-1}\mathbb{E}_T[VV^\top]\right).$$

This can be rewritten more compactly as:

$$H = \frac{1}{1-\alpha}(C - AB^{-1}A),$$

where we define:

$$C = \mathbb{E}_T\left[\frac{1}{\sigma(X,\beta)}VV^\top\right], \quad A = \mathbb{E}_T[VV^\top], \quad B = \mathbb{E}_T[\sigma(X,\beta)VV^\top].$$

First, we define the vector:

$$\tilde{V} := \begin{bmatrix} \sqrt{\sigma(X,\beta)}V \\ \frac{1}{\sqrt{\sigma(X,\beta)}}V \end{bmatrix} \in \mathbb{R}^{2d}.$$

(Note that here we changed the definition of $\tilde{V}$ to simplify later steps.)

Then the outer product $\tilde{V}\tilde{V}^\top$ is:

$$\tilde{V}\tilde{V}^\top = \begin{bmatrix} \sigma(X,\beta)VV^\top & VV^\top \\ VV^\top & \frac{1}{\sigma(X,\beta)}VV^\top \end{bmatrix}.$$

Taking expectation, we define the matrix:

$$M := \mathbb{E}_T[\tilde{V}\tilde{V}^\top] = \begin{bmatrix} B & A \\ A^\top & C \end{bmatrix}.$$

We now show that M is PSD. For any $z \in \mathbb{R}^{2d}$, we have:

$$z^\top M z = \mathbb{E}_T\left[z^\top \tilde{V}\tilde{V}^\top z\right] = \mathbb{E}_T\left[(z^\top \tilde{V})^2\right] \geq 0,$$

since each term in the expectation is a square. Therefore, $M \succeq 0$. By the Schur Complement Lemma for block matrices, provided that B is invertible and square and $B \succeq 0$ and $M \succeq 0$, we get that

$$H = \frac{1}{1-\alpha}(C - AB^{-1}A) \succeq 0.$$

Now, by definition:

$$\Sigma^{\text{OLS}}_{\text{wG}} - \Sigma^{\text{OLS}}_{\text{tG}} = \begin{bmatrix} (\beta^{(0)})^\top H\beta^{(0)} & (\beta^{(1)})^\top H\beta^{(0)} \\ (\beta^{(1)})^\top H\beta^{(0)} & (\beta^{(1)})^\top H\beta^{(1)} \end{bmatrix}.$$

Since this matrix is a Gram matrix induced by $H \succeq 0$, it follows that $\Sigma^{\text{OLS}}_{\text{wG}} - \Sigma^{\text{OLS}}_{\text{tG}} \succeq 0$. □

Proposition 15. *Grant Assumption 1 to 1, and let $\hat{\tau}^{OLS}_{\Phi,\text{tG}}$ denote the transported G-formula estimators where linear regressions are used to estimate $\mu^S_{(a)}$. Then, $\hat{\tau}^{OLS}_{\Phi,\text{tG}}$ is asymptotically normal:*

$$\sqrt{N}\left(\hat{\tau}^{OLS}_{\Phi,\text{tG}} - \tau^T_\Phi\right) \xrightarrow{d} \mathcal{N}\left(0, V^{OLS}_{\Phi,\text{tG}}\right),$$

where

$$V^{OLS}_{\Phi,\text{tG}} = \frac{1}{1-\alpha}\left\|\frac{\partial\Phi}{\partial\psi_1}\beta^{(1)} + \frac{\partial\Phi}{\partial\psi_0}\beta^{(0)}\right\|_{Var[V|S=0]} + \frac{\sigma^2\mathbb{E}_T[V]^\top M^{-1}\mathbb{E}_T[V]}{\alpha}\left[\frac{1}{1-\pi}\left(\frac{\partial\Phi}{\partial\psi_0}\right)^2 + \frac{1}{\pi}\left(\frac{\partial\Phi}{\partial\psi_1}\right)^2\right].$$

Proof. We begin by analyzing the transported OLS estimator, defined as

$$\hat{\tau}^{\text{OLS}}_{\Phi,\text{tG}} = \Phi\left((\hat{\beta}^{(1)})^\top\bar{V}_{(0)},\ (\hat{\beta}^{(0)})^\top\bar{V}_{(0)}\right),$$

where $\bar{V}_{(0)}$ is the empirical mean of covariates in the target population. Under Linear model 1, the corresponding population estimand is

$$\tau^T_\Phi = \Phi\left((\beta^{(1)})^\top\mathbb{E}_T[V],\ (\beta^{(0)})^\top\mathbb{E}_T[V]\right),$$

where $\mathbb{E}_T[V]$ is the expectation of covariates under the target distribution.

To study the asymptotic behavior of the estimator, we perform a first-order Taylor expansion of Φ around the point $(\psi_1^*, \psi_0^*) := \left((\beta^{(1)})^\top \mathbb{E}_T[V],\ (\beta^{(0)})^\top \mathbb{E}_T[V]\right)$. This yields:

$$\begin{aligned}
\hat{\tau}^{\text{OLS}}_{\Phi,\text{tG}} - \tau^T_\Phi &= \Phi\left((\hat{\beta}^{(1)})^\top \bar{V}_{(0)},\ (\hat{\beta}^{(0)})^\top \bar{V}_{(0)}\right) - \Phi\left((\beta^{(1)})^\top \mathbb{E}_T[V],\ (\beta^{(0)})^\top \mathbb{E}_T[V]\right) \\
&= \frac{\partial \Phi}{\partial \psi_1}\Big|_{(\psi_1^*,\psi_0^*)} \left((\hat{\beta}^{(1)})^\top \bar{V}_{(0)} - (\beta^{(1)})^\top \mathbb{E}_T[V]\right) \\
&\quad + \frac{\partial \Phi}{\partial \psi_0}\Big|_{(\psi_1^*,\psi_0^*)} \left((\hat{\beta}^{(0)})^\top \bar{V}_{(0)} - (\beta^{(0)})^\top \mathbb{E}_T[V]\right) \\
&\quad + o_p\left(\left\| \begin{pmatrix} (\hat{\beta}^{(1)})^\top \bar{V}_{(0)} - (\beta^{(1)})^\top \mathbb{E}_T[V] \\ (\hat{\beta}^{(0)})^\top \bar{V}_{(0)} - (\beta^{(0)})^\top \mathbb{E}_T[V] \end{pmatrix} \right\|\right).
\end{aligned}$$

To further decompose the linear terms, note that for each $a \in \{0,1\}$, we can write:

$$(\hat{\beta}^{(a)})^\top \bar{V}_{(0)} - (\beta^{(a)})^\top \mathbb{E}_T[V] = (\hat{\beta}^{(a)} - \beta^{(a)})^\top \mathbb{E}_T[V] + (\hat{\beta}^{(a)})^\top \left(\bar{V}_{(0)} - \mathbb{E}_T[V]\right).$$

Combining these expressions, we obtain:

$$\begin{aligned}
\hat{\tau}^{\text{OLS}}_{\Phi,\text{tG}} - \tau^T_\Phi &= \frac{\partial \Phi}{\partial \psi_1}\Big|_{(\psi_1^*,\psi_0^*)} \left[(\hat{\beta}^{(1)} - \beta^{(1)})^\top \mathbb{E}_T[V] + (\hat{\beta}^{(1)})^\top \left(\bar{V}_{(0)} - \mathbb{E}_T[V]\right)\right] \\
&\quad + \frac{\partial \Phi}{\partial \psi_0}\Big|_{(\psi_1^*,\psi_0^*)} \left[(\hat{\beta}^{(0)} - \beta^{(0)})^\top \mathbb{E}_T[V] + (\hat{\beta}^{(0)})^\top \left(\bar{V}_{(0)} - \mathbb{E}_T[V]\right)\right] + o_p(N^{-1/2}).
\end{aligned}$$

By the multivariate Central Limit Theorem, the Law of Large Numbers, and Slutsky's theorem—along with 13—we conclude:

$$\sqrt{N}\left(\hat{\tau}^{\text{OLS}}_{\Phi,\text{tG}} - \tau^T_\Phi\right) \xrightarrow{d} \mathcal{N}\left(0,\ \alpha_\infty^\top \Sigma\, \alpha_\infty\right),$$

where Σ is defined in 30 and the influence vector α_∞ is given by

$$\alpha_\infty = \frac{\partial \Phi}{\partial \psi_1}\Big|_{(\psi_1^*,\psi_0^*)} \alpha_{1,\infty} + \frac{\partial \Phi}{\partial \psi_0}\Big|_{(\psi_1^*,\psi_0^*)} \alpha_{0,\infty},$$

and each $\alpha_{a,\infty} \in \mathbb{R}^p$ is defined as

$$\alpha_{a,\infty}^\top = \left((\beta^{(a)})^\top,\ \mathbb{E}_T[V]^\top \cdot \mathbb{1}_{\{a=0\}},\ \mathbb{E}_T[V]^\top \cdot \mathbb{1}_{\{a=1\}}\right) \in \mathbb{R}^{3p+3}.$$

Combining this with the block-diagonal form of Σ, we obtain the explicit asymptotic variance:

$$\begin{aligned}
\alpha_\infty^\top \Sigma \alpha_\infty &= \frac{1}{1-\alpha} \left\| \frac{\partial \Phi}{\partial \psi_1}\beta^{(1)} + \frac{\partial \Phi}{\partial \psi_0}\beta^{(0)} \right\|_{\text{Var}[V|S=0]} \\
&\quad + \frac{\sigma^2 \mathbb{E}_T[V]^\top M^{-1} \mathbb{E}_T[V]}{\alpha} \left[\frac{1}{1-\pi}\left(\frac{\partial \Phi}{\partial \psi_0}\right)^2 + \frac{1}{\pi}\left(\frac{\partial \Phi}{\partial \psi_1}\right)^2\right].
\end{aligned}$$

□

Proposition 16. *Let $\hat{\tau}^{OLS}_{\Phi,\text{wG}}$ denote the weighted G-formula estimators where the density ratio is estimated using a logistic regression and linear regressions are used to estimate $\mu^S_{(a)}$. Then, under Assumption 1 to 1, $\hat{\tau}^{OLS}_{\Phi,\text{wG}}$ is asymptotically normal:*

$$\sqrt{N}\left(\hat{\tau}^{OLS}_{\Phi,\text{wG}} - \tau^T_\Phi\right) \xrightarrow{d} \mathcal{N}\left(0, V^{OLS}_{\Phi,\text{wG}}\right),$$

where

$$V^{OLS}_{\Phi,\text{wG}} = \left(\frac{\partial \Phi}{\partial \psi_1}\right)^2 V^{OLS}_{1,\text{wG}} + \left(\frac{\partial \Phi}{\partial \psi_0}\right)^2 V^{OLS}_{0,\text{wG}} + 2\left(\frac{\partial \Phi}{\partial \psi_1}\right)\left(\frac{\partial \Phi}{\partial \psi_0}\right) C^{OLS}_{\text{wG}}.$$

Proof. Using 14, we apply the delta method to $\hat{\tau}^{\text{OLS}}_{\Phi,\text{wG}} = \Phi(\hat{\psi}^{\text{OLS}}_{1,\text{wG}}, \hat{\psi}^{\text{OLS}}_{0,\text{wG}})$ of the two-dimensional asymptotically normal vector. By the delta method:

$$\sqrt{N}\left(\hat{\tau}^{\text{OLS}}_{\Phi,\text{wG}} - \Phi(\psi_1,\psi_0)\right) \xrightarrow{d} \mathcal{N}\left(0, \nabla\Phi(\psi_1,\psi_0)^\top \Sigma^{\text{OLS}}_{\text{wG}} \nabla\Phi(\psi_1,\psi_0)\right),$$

where $\nabla\Phi(\psi_1,\psi_0) = \left(\frac{\partial\Phi}{\partial\psi_1}, \frac{\partial\Phi}{\partial\psi_0}\right)^\top$ is the gradient of Φ evaluated at the population means. Therefore we get:

$$V^{\text{OLS}}_{\Phi,\text{wG}} = \left(\frac{\partial\Phi}{\partial\psi_1}\right)^2 V^{\text{OLS}}_{1,\text{wG}} + \left(\frac{\partial\Phi}{\partial\psi_0}\right)^2 V^{\text{OLS}}_{0,\text{wG}} + 2\left(\frac{\partial\Phi}{\partial\psi_1}\right)\left(\frac{\partial\Phi}{\partial\psi_0}\right) C^{\text{OLS}}_{\text{wG}}.$$

□

B.5 Semiparametric Efficient Estimators under Exchangeability of conditional outcome

Proof of Proposition 3. In this proof, we use the usual machinery of influence function computation, as described for instance in [26]. In particular, we will for the sake of the computation, assume that X is a categorical variable taking value in a countable space. We first recall that if

$$\psi(P_{\text{obs}}) := \mathbb{E}_{P_{\text{obs}}}[h(Z)|\mathcal{A}] \tag{34}$$

for some measurable function h and some event of positive mass $\mathcal{A}$, then the influence function of ψ at P_{obs} is given by

$$\text{IF}(\psi)(Z) := \frac{\mathbb{1}\{\mathcal{A}\}}{P_{\text{obs}}(\mathcal{A})}(h(Z) - \psi). \tag{35}$$

We now rewrite ψ_a using functional of the form (34):

$$\begin{aligned}
\psi_a &:= \mathbb{E}_P[Y^{(a)}|S=0] \\
&= \mathbb{E}_{P_{\text{obs}}}[\mathbb{E}_P[Y^{(a)}|X,S=0]|S=0] \\
&= \mathbb{E}_{P_{\text{obs}}}[\mathbb{E}_P[Y^{(a)}|X,S=1]|S=0] \\
&= \mathbb{E}_{P_{\text{obs}}}[\mathbb{E}_{P_{\text{obs}}}[Y|X,S=1,A=a]|S=0] \\
&= \sum_{x\in\mathcal{X}} \mathbb{E}_{P_{\text{obs}}}[\mathbb{1}\{X=x\}|S=0] \times \mathbb{E}_{P_{\text{obs}}}[Y|X=x,S=1,A=a].
\end{aligned}$$

Using (35), and usual properties of the influence functions [26, Sec 3.4.3], we find

$$\begin{aligned}
\text{IF}(\psi_a)(Z) = &\sum_{x\in\mathcal{X}} \frac{1-S}{1-\alpha}(\mathbb{1}\{X=x\} - \mathbb{E}_{P_{\text{obs}}}[\mathbb{1}\{X=x\}|S=0]) \times \mathbb{E}_{P_{\text{obs}}}[Y|X=x,S=1,A=a] \\
&+ \sum_{x\in\mathcal{X}} \mathbb{E}_{P_{\text{obs}}}[\mathbb{1}\{X=x\}|S=0] \times \frac{\mathbb{1}\{X=x,S=1,A=a\}}{P_{\text{obs}}(X=x,S=1,A=a)}(Y - \mathbb{E}_{P_{\text{obs}}}[Y|X=x,S=1,A=a]).
\end{aligned}$$

The first sum simply rewrites

$$\frac{1-S}{1-\alpha}(\mu_{(a)}(X) - \Psi_a(P)),$$

and for the second, notice that, using that A and S are independent:

$$\frac{\mathbb{E}_{P_{\text{obs}}}[\mathbb{1}\{X=x\}|S=0]}{P(X=x,S=1,A=a)} = \frac{r(X)}{\alpha P(A=a)},$$

so that the sum rewrites

$$\frac{\mathbb{1}\{S=1,A=a\}}{\alpha P(A=a)} r(X)(Y - \mu_{(a)}(X)).$$

In the end, we indeed find

$$\text{IF}(\psi_a)(Z) = \frac{1-S}{1-\alpha}(\mu_{(a)}(X) - \psi_a) + \frac{S\mathbb{1}\{A=a\}}{\alpha P(A=a)} r(X)(Y - \mu_{(a)}(X)).$$

□

Proof of Proposition 5. By conditioning with respect to the randomness of $\widehat{\mu}_{(a)}$ and $\widehat{r}$, we can treat these functions as deterministic. By using the law of large number, we see that m/N goes to $1-\alpha$ and that $\widehat{\psi}_{(a)}$ converges towards

$$\frac{1}{1-\alpha}\mathbb{E}[(1-S)\widehat{\mu}_{(a)}(X)] + \frac{1}{\alpha}\mathbb{E}\left[S\widehat{r}(X)(Y^a - \widehat{\mu}_{(a)}(X)\right].$$

If $\widehat{\mu}_{(a)} = \mu_{(a)}$, then the second term in the above formula cancels and we are left with

$$\frac{1}{1-\alpha}\mathbb{E}[(1-S)\widehat{\mu}_{(a)}(X)] = \mathbb{E}[\widehat{\mu}_{(a)}(X)|S=0] = \psi_a.$$

If $\widehat{r} = r$, then the second term yields

$$\mathbb{E}\left[r(X)(Y - \widehat{\mu}_{(a)}(X)|S=1\right] = \mathbb{E}\left[Y^a - \widehat{\mu}_{(a)}(X)|S=0\right] = \psi_a - \mathbb{E}\left[\widehat{\mu}_{(a)}(X)|S=0\right],$$

which also yields the results. Using continuity of Φ allows to conclude. □

B.6 Computations of one-step estimators and estimating equation estimators for the OR.

Note that $\widehat{\psi}_a^{\mathrm{EE}}$ is solution to $\sum \varphi_a(Z_i, \widehat{\eta}, \widehat{\psi}_a^{\mathrm{EE}}) = 0$. Then, given the expression of φ_a in Proposition 3, it holds that for any other estimator $\widehat{\psi}_a$:

$$\sum_{i=1}^{N} \varphi_a(Z_i, \widehat{\eta}, \widehat{\psi}_a) = -\frac{m}{1-\alpha}(\widehat{\psi}_a - \widehat{\psi}_a^{\mathrm{EE}}).$$

We will use this observation in the subsequent computations.

(RD) For the risk difference, the estimating equation approach yields:

$$\widehat{\tau}_{\mathrm{RD}}^{\mathrm{EE}} = \widehat{\psi}_1^{\mathrm{EE}} - \widehat{\psi}_0^{\mathrm{EE}}.$$

The one step approach, based on initial estimators $\widehat{\psi}_1$ and $\widehat{\psi}_0$ yields an estimate of the form:

$$\begin{aligned}\widehat{\tau}_{\mathrm{RD}}^{\mathrm{OS}} &= \widehat{\psi}_1 - \widehat{\psi}_0 + \frac{m}{N(1-\alpha)}(\widehat{\psi}_1^{\mathrm{EE}} - \widehat{\psi}_1) - \frac{m}{N(1-\alpha)}(\widehat{\psi}_0^{\mathrm{EE}} - \widehat{\psi}_0) \\ &= \frac{m}{N(1-\alpha)}\widehat{\tau}_{\mathrm{RD}}^{\mathrm{EE}} + \left(1 - \frac{m}{N(1-\alpha)}\right)\widehat{\tau}_{\mathrm{RD}}\end{aligned}$$

In particular, we see that starting from estimators $\widehat{\psi}_a$ of the form $\sum_{S_i=0}\widehat{\mu}_{(a)}(X_i)$ yields a final estimator that has the same structure as the estimating equation estimator, up to scaling factors depending on α, N and m that are asymptotically close to 1.

(RR) For the risk ratio, the estimating equation approach yields:

$$\widehat{\tau}_{\mathrm{RR}}^{\mathrm{EE}} = \frac{\widehat{\psi}_1^{\mathrm{EE}}}{\widehat{\psi}_0^{\mathrm{EE}}}.$$

In contrast, the one step approach, based on initial estimators $\widehat{\psi}_1$ and $\widehat{\psi}_0$ yields an estimate of the form:

$$\widehat{\tau}_{\mathrm{RR}}^{\mathrm{OS}} = \frac{\widehat{\psi}_1}{\widehat{\psi}_0} + \frac{1}{\widehat{\psi}_0}\frac{m}{N(1-\alpha)}(\widehat{\psi}_1^{\mathrm{EE}} - \widehat{\psi}_1) - \frac{\widehat{\psi}_1}{\widehat{\psi}_0^2}\frac{m}{N(1-\alpha)}(\widehat{\psi}_0^{\mathrm{EE}} - \widehat{\psi}_0).$$

In particular, we see that in general, $\widehat{\tau}_{\mathrm{RR}}^{\mathrm{OS}} \neq \widehat{\tau}_{\mathrm{OR}}^{\mathrm{EE}}$, unless of course we initially picked $\widehat{\psi}_a^{\mathrm{EE}}$ to apply the one-step correction.

(OR) For the odds ratio, the estimating equation approach yields:

$$\widehat{\tau}_{\mathrm{OR}}^{\mathrm{EE}} = \frac{\widehat{\psi}_1^{\mathrm{EE}}}{1-\widehat{\psi}_1^{\mathrm{EE}}}\frac{1-\widehat{\psi}_0^{\mathrm{EE}}}{\widehat{\psi}_0^{\mathrm{EE}}}.$$

In contrast, the one step approach, based on initial estimators $\widehat{\psi}_1$ and $\widehat{\psi}_0$ yields an estimate of the form:

$$\begin{aligned}\widehat{\tau}^{\mathrm{OS}}_{\mathrm{OR}} =& \frac{\widehat{\psi}_1}{1-\widehat{\psi}_1}\frac{1-\widehat{\psi}_0}{\widehat{\psi}_0} + \frac{1}{(1-\widehat{\psi}_1)^2}\frac{1-\widehat{\psi}_0}{\widehat{\psi}_0}\frac{m}{N(1-\alpha)}(\widehat{\psi}_1 - \widehat{\psi}^{\mathrm{EE}}_1) \\ &- \frac{\widehat{\psi}_1}{1-\widehat{\psi}_1}\frac{1}{\widehat{\psi}_0^2}\frac{m}{N(1-\alpha)}(\widehat{\psi}_0 - \widehat{\psi}^{\mathrm{EE}}_0).\end{aligned}$$

In particular, we see that in general, $\widehat{\tau}^{\mathrm{OS}}_{\mathrm{OR}} \neq \widehat{\tau}^{\mathrm{EE}}_{\mathrm{OR}}$, unless of course we initially picked $\widehat{\psi}^{\mathrm{EE}}_a$ to apply the one-step correction.

C Transporting/Reweighting a causal effect under exchangeability of CATE

C.1 Identification under exchangeability of conditional outcome

Risk Difference (RD)

$$\tau^{\mathrm{T}}_{\mathrm{RD}} = \mathbb{E}_{\mathrm{T}}\left[\tau^{\mathrm{S}}_{\mathrm{RD}}(X)\right]$$

Risk Ratio (RR)

$$\tau^{\mathrm{T}}_{\mathrm{RR}} = \frac{\mathbb{E}_{\mathrm{T}}\left[\tau^{\mathrm{S}}_{\mathrm{RR}}(X)\cdot\mu^{\mathrm{T}}_{(0)}(X)\right]}{\mathbb{E}_{\mathrm{T}}\left[Y^{(0)}\right]}$$

Odds Ratio (OR)

$$\tau^{\mathrm{T}}_{\mathrm{OR}} = \left(\frac{\mathbb{E}_{\mathrm{T}}\left[Y^{(0)}\right]}{1-\mathbb{E}_{\mathrm{T}}\left[Y^{(0)}\right]}\right)^{-1}\cdot\frac{\mathbb{E}_{\mathrm{T}}\left[\frac{\tau^{\mathrm{S}}_{\mathrm{OR}}(X)\cdot\mu^{\mathrm{T}}_{(0)}(X)}{1+\tau^{\mathrm{S}}_{\mathrm{OR}}(X)\cdot\mu^{\mathrm{T}}_{(0)}(X)-\mu^{\mathrm{T}}_{(0)}(X)}\right]}{1-\mathbb{E}_{\mathrm{T}}\left[\frac{\tau^{\mathrm{S}}_{\mathrm{OR}}(X)\cdot\mu^{\mathrm{T}}_{(0)}(X)}{1+\tau^{\mathrm{S}}_{\mathrm{OR}}(X)\cdot\mu^{\mathrm{T}}_{(0)}(X)-\mu^{\mathrm{T}}_{(0)}(X)}\right]}$$

C.2 Semiparametric efficient estimators under exchangeability of treatment effect

Proof of Proposition 6. We use the same tricks as in the proof of Proposition 3. We first notice that

$$\psi^{\mathrm{T}}_1 = \sum_{x\in\mathcal{X}} P_{\mathrm{obs}}(X=x|S=0)\Gamma(\tau_\Phi(x),\mu^{\mathrm{T}}_{(0)}(x)),$$

so that, with a slight abuse of notation

$$\begin{aligned}\mathrm{IF}(\psi^{\mathrm{T}}_1) =& \sum_{x\in\mathcal{X}}\frac{1-S}{1-\alpha}(\mathbb{1}\{X=x\} - P_{\mathrm{obs}}(X=x|S=0))\Gamma(\tau_\Phi(x),\mu^{\mathrm{T}}_{(0)}(x)) \\ &+ \sum_{x\in\mathcal{X}} P_{\mathrm{obs}}(X=x|S=0)\mathrm{IF}(\tau_\Phi(x))\partial_1\Gamma(\tau_\Phi(x),\mu^{\mathrm{T}}_{(0)}(x)).\end{aligned}$$

Using that $\tau_\Phi(x) = \Phi(\mathbb{E}[Y|A=1,S=1,X=x],\mathbb{E}[Y|A=0,S=1,X=x])$, we further find that

$$\begin{aligned}\mathrm{IF}(\tau_\Phi(x)) =& \frac{\mathbb{1}\{A=1,S=1,X=x\}}{P(A=1,S=1,X=x)}(Y-\mu^{\mathrm{S}}_{(1)}(x))\partial_1\Phi(\mu^{\mathrm{S}}_{(1)}(x),\mu^{\mathrm{S}}_{(0)}(x)) \\ &+ \frac{\mathbb{1}\{A=0,S=1,X=x\}}{P(A=0,S=1,X=x)}(Y-\mu^{\mathrm{S}}_{(0)}(x))\partial_0\Phi(\mu^{\mathrm{S}}_{(1)}(x),\mu^{\mathrm{S}}_{(0)}(x))\end{aligned}$$

Again, we know that

$$P(A=a,X=x,S=1) = \alpha\pi P(X=x|S=1) = \alpha\pi r(X)P(X=x|S=0).$$

Furthermore, since $\Gamma(\Phi(a,b),b) = a$ for all a,b, differentiating with respect to a or b yields $\partial_0\Phi(a,b)\partial_1\Gamma(\Phi(a,b),b)) + \partial_0\Gamma(\Phi(a,b),b) = 0$ and $\partial_1\Phi(a,b)\partial_1\Gamma(\Phi(a,b),b) = 1$. Patching all of this together yields the result. □

C.3 Computations of one-step estimators and estimating equation estimators under exchangeability of CATE

Example 3 (Application to the usual causal measures.). We give the expression of $\widehat{\psi}_1^{\mathrm{EE}}$ for the most usual causal measures.

(RD) For the risk difference, we find

$$\widehat{\psi}_1^{\mathrm{EE}} = \frac{1}{m}\sum_{S_i=0} \mu_{(0)}^{\mathrm{T}}(X_i) + \widehat{\tau}_\Phi(X_i) + \frac{1-\alpha}{\alpha m}\sum_{S_i=1} \widehat{r}(X_i)\left(\frac{A_i}{\pi}(Y_i - \widehat{\tau}_\Phi(X_i) - \widehat{\mu}_{(0)}^{\mathrm{S}}(X_i)) - \frac{1-A_i}{1-\pi}(Y_i - \widehat{\mu}_{(0)}^{\mathrm{S}}(X_i))\right).$$

(RR) For the risk ratio, we find:

$$\widehat{\psi}_1^{\mathrm{EE}} = \frac{1}{m}\sum_{S_i=0} \mu_{(0)}^{\mathrm{T}}(X_i)\widehat{\tau}_\Phi(X_i) + \frac{1-\alpha}{\alpha m}\sum_{S_i=1} \widehat{r}(X_i)\left(\frac{A_i}{\pi}(Y_i - \widehat{\mu}_{(0)}^{\mathrm{S}}(X_i)\widehat{\tau}_\Phi(X_i)) - \frac{1-A_i}{1-\pi}(Y_i - \widehat{\mu}_{(0)}^{\mathrm{S}}(X_i))\widehat{\tau}_\Phi(X_i)\right).$$

(OR) For the odds ratio, we find:

$$\widehat{\psi}_1^{\mathrm{EE}} = \frac{1}{m}\sum_{S_i=0} \frac{\mu_{(0)}^{\mathrm{T}}(X_i)\widehat{\tau}_\Phi(X_i)}{1-\mu_{(0)}^{\mathrm{T}}(X_i)+\mu_{(0)}^{\mathrm{T}}(X_i)\widehat{\tau}_\Phi(X_i)} + \frac{1-\alpha}{\alpha m}\sum_{S_i=1} \widehat{r}(X_i)\left(\frac{A_i}{\pi}\left(Y_i - \frac{\mu_{(0)}^{\mathrm{S}}(X_i)\widehat{\tau}_\Phi(X_i)}{1-\mu_{(0)}^{\mathrm{S}}(X_i)+\mu_{(0)}^{\mathrm{S}}(X_i)\widehat{\tau}_\Phi(X_i)}\right) - \frac{1-A_i}{1-\pi}(Y_i - \widehat{\mu}_{(0)}^{\mathrm{S}}(X_i))\frac{\widehat{\tau}_\Phi(X_i)}{(1-\widehat{\mu}_{(0)}^{\mathrm{S}}(X_i)+\widehat{\mu}_{(0)}^{\mathrm{S}}(X_i)\widehat{\tau}_\Phi(X_i))^2}\right).$$

D Simulation

For the simulations we have implemented all estimators in Python using Scikit-Learn for our regression and classification models. All our experiments were run on a 8GB M1 Mac.

Linear setting under Assumption 3: we evaluate estimators under a linear response surface:

$$\mu_s^{(a)}(V) = \beta_a^\top V \quad \text{with} \quad \beta_1 = (0.5,\ 1.2,\ 1.1,\ 3.3,\ -0.6) \quad \text{and} \quad \beta_0 = (-0.2,\ -0.6,\ 0.6,\ 1.7,\ 0.3).$$

Since β_0 and β_1 remain unchanged across the source and target domains, Assumption 3 is satisfied, results are depicted in Figure 4. As expected from the linear generative process, all estimators perform well across all measures, with the transported and weighted G-formula exhibiting particularly low variance in this setting—outperforming the influence function–based estimators.

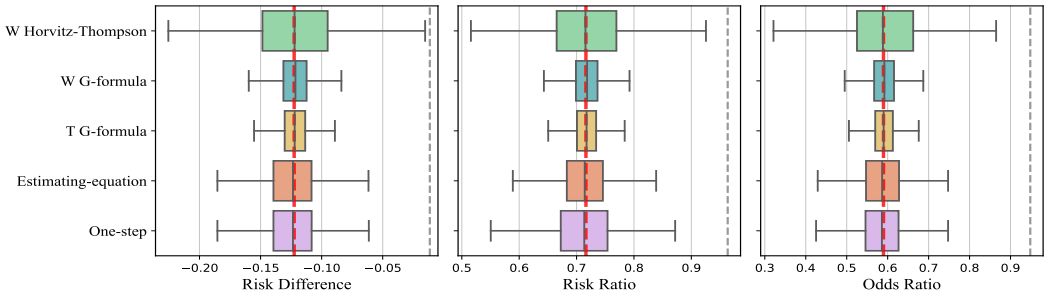

Figure 4: Comparison of estimators across different causal measures under a linear outcome model with a sample size of $N = 50{,}000$ and 3,000 repetitions.

NeurIPS Paper Checklist

A. **Claims**

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: All the claims presented in the abstract and introduction are formalized as propositions within the paper, based on classical causal inference assumptions.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

B. **Limitations**

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: We discuss (mainly in the conclusion) the limitation of our work.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.

- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

C. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Rigorous proofs are provided in the Appendix, with references to the theorems and lemmas upon which our proofs rely.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

D. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We described in the paper or appendix the settings of the model we use and parameters we used for our simulations.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example

(a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

E. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Experiments are considered simple enough to be reproduced without providing open acces to our code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

F. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Experimental Setting/Details are fully provided in the paper or in Appendix.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

G. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We used boxplots to report our estimations.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

H. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: Yes, the paper does indicate the type of compute used for the experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

I. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: We made sure to preserve anonymity in our paper.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

J. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discussed the impact of our findings to better estimate treatment effects.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

K. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

L. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All theorems as well as the data used for real world experiment are properly credited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

M. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

N. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

O. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer:[Yes]

Justification: The TraumaBase© obtained approval from the Institutional Review Board (Comité de Protection des Personnes, Paris VI,) from the Advisory Committee for Information Processing in Health Research (CCTIRS, 11.305bis) and from the National Commission on Informatics and Liberties (CNIL, 911461).

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

P. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs were not used for any relevant parts of this paper.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.