The Annals of Applied Statistics
2020, Vol. 14, No. 4, 1806–1833
https://doi.org/10.1214/20-AOAS1359

THE STATISTICAL PERFORMANCE OF MATCHING-ADJUSTED INDIRECT COMPARISONS: ESTIMATING TREATMENT EFFECTS WITH AGGREGATE EXTERNAL CONTROL DATA

BY DAVID CHENG[1], RAJEEV AYYAGARI[2,*] AND JAMES SIGNOROVITCH[2,†]

[1]*Biostatistics Center, Massachusetts General Hospital, dcheng@mgh.harvard.edu*
[2]*Analysis Group, Inc., *rajeev.ayyagari@analysisgroup.com; †james.signorovitch@analysisgroup.com*

Indirect comparisons of treatment-specific outcomes across separate studies often inform decision making in the absence of head-to-head randomized comparisons. Differences in baseline characteristics between study populations may introduce confounding bias in such comparisons. Matching-adjusted indirect comparison (MAIC) (*Pharmacoeconomics* **28** (2010) 935–945) has been used to adjust for differences in observed baseline covariates when the individual patient-level data (IPD) are available for only one study and aggregate data (AGD) are available for the other study. The approach weights outcomes from the IPD using estimates of trial selection odds that balance baseline covariates between the IPD and AGD. With the increasing use of MAIC, there is a need for formal assessments of its statistical properties. In this paper we formulate identification assumptions for causal estimands that justify MAIC estimators. We then examine large sample properties and evaluate strategies for estimating standard errors without the full IPD from both studies. The finite-sample bias of MAIC and the performance of confidence intervals based on different standard error estimators are evaluated through simulations. The method is illustrated through an example comparing placebo arm and natural history outcomes in Duchenne muscular dystrophy.

1. Introduction. Indirect comparisons refer to comparisons of treatment-specific outcomes across different studies, as opposed to direct comparisons of treatments-specific outcomes within a single study. This frequently arises in clinical, economic and regulatory evaluations of new treatments (Altman et al. (2005), Sutton et al. (2008), Wells et al. (2009), Jansen et al. (2011)). Even when a new treatment has been compared with a standard of care in a randomized trial, for example, it is often important for decision-makers to contextualize expected outcomes with other treatments and external data sources. The need for indirect comparisons is also especially acute in settings where direct randomized comparisons are unethical or infeasible, such as in late-stage oncology trials (Adjei, Christian and Ivy (2009)), rare disease trials and evaluations of long-term outcomes in open-label extension trials. Indirect comparisons also inform the design of trials by providing reasonable prior estimates of effect sizes for power calculations and noninferiority margins (Snapinn and Jiang (2011)).

Since indirect comparisons involve comparisons between studies for which assignment to different studies is not randomized, the resulting treatment effect estimates may be confounded by cross-study differences in baseline characteristics. Such differences may exist even when treatment groups *within* each study have been randomized. In practice, a host of differences in the design and setting of each study can bias indirect comparisons. It is essential to consider, with clinical input, definitions and assessment methodologies for the outcome measure, patient selection criteri and background care, along with other issues (Pocock (1976), FDA (2001), Phillippo et al. (2016)). A nonexhaustive list of considerations is provided in Table 1. Although a wide range of issues may introduce bias, there are many cases

Received August 2019; revised April 2020.
Key words and phrases. Indirect comparison, matching-adjusted indirect comparison, causal inference, health technology assessment.

TABLE 1

Study design features that may vary across studies and bias indirect comparisons. Differences in observed baseline characteristics, the final category, are addressed by the adjustment methods described in this paper

Aspect of study design
Outcome assessment:
Clinical, imaging or laboratory methods used to process and measure outcomes.
Outcome definitions, including timing of assessments.
Event ascertainment or adjudication procedures.
Completeness of follow-up, reasons for drop-out.
Patient selection:
Inclusion and exclusion criteria.
Recruitment process (e.g., motivation, consent process, distance of site to patient).
Disease diagnostic criteria.
Background treatments and care settings:
Time period and geography, and associated standards of care.
Range of nonstudy treatment options available.
Concomitant medications.
Baseline patient characteristics:
Demographics.
Comorbidities.
Treatment history.
Severity and duration of medical condition.
Biomarkers.

in which separate studies are found, after careful evaluation, to be sufficiently similar for an indirect comparison. Registrational trials conducted for different treatments within the same indication during similar time periods, for instance, often share a high degree of similarity. In this paper we assume that trials are sufficiently similar in design such that attention can be focused on addressing bias that stems from differences in baseline characteristics.

When individual patient-level data (IPD) are available for both studies, the pooled data can essentially be regarded as observational data with a nonrandomized treatment assignment. In this case methods for estimating treatment effects with nonrandomized treatment based on outcome regression and propensity score approaches are well established (Lunceford and Davidian (2004), Kang and Schafer (2007)). Conducting indirect comparisons with full IPD can also be viewed as generalizing estimates from one study to a different target study population (Stuart et al. (2011), Hartman et al. (2015), Nie et al. (2013), Zhang et al. (2016)) which has a long history of application in regulatory settings in the form of studies with historical controls (FDA (2001), EMA CHMP (2006)). Access to full IPD should always be the preferred approach to comparative analyses, when possible.

In practice, however, many indirect comparisons are conducted in settings where the full IPD are not available for both study populations. When only aggregate data (AGD) consisting of summary statistics (e.g., means and standard errors from publications) for outcomes and baseline covariates are available, Bucher et al. (1997) introduced a widely used method to compare treatment effects on an appropriate scale of contrast (e.g., log odds ratio, risk difference, etc.) relative to a common comparator group. But, as we discuss in Section 3, this method requires a common comparator arm between studies and does not eliminate the risk of bias. If AGD from a large number of studies are available, mixed treatment comparisons generalize methods from meta-analysis and Bucher et al. (1997) to allow for comparisons of treatments within a network of randomized trials that are linked by common comparators (Lu and Ades (2004), Nixon, Bansback and Brennan (2007)). These approaches rely on "consistency" or "exchangeability" assumptions, requiring treatment effects to be constant across

trials on a specified scale of contrast. These assumptions are typically violated when study populations differ in baseline characteristics, particularly those that modify treatment effects on the chosen contrast scale.

It is often the case that IPD are available for one study whereas only AGD are available for others. This can occur, for example, when researchers can access IPD from a study they conducted but cannot directly access IPD underlying published AGD from other studies. When IPD are available for only one study and AGD are available for another, two general approaches have been used. The simulated treatment comparison (STC) method estimates an adjusted mean outcome under the IPD treatment by fitting an outcome regression model to the IPD and plugging in baseline covariates from the AGD (Caro and Ishak (2010), Ishak, Proskorovsky and Benedict (2015)). Indirect comparisons can be obtained by contrasting the predicted value with observed mean outcomes in the AGD. If a nonlinear regression model is used, such as a logit link in a logistic regression model or a log link in a proportional hazards model, this method incurs bias because expectations do not commute with nonlinear functions. Addressing this requires parametric assumptions about the full joint distribution of covariates from the AGD. STC could also be biased when the postulated outcome regression model is misspecified. An additional method is matching-adjusted indirect comparison (MAIC) (Signorovitch et al. (2010)), which estimates mean outcomes for the IPD in the population represented by the AGD by estimating trial selection odds through a method of moments approach and reweighting the IPD. This avoids issues that arise from regression modeling, though it still generally assumes correctly specified models for the trial selection odds. MAIC is similar in spirit to a number of other reweighting methods that seek to balance covariates between treatment groups to estimate causal effects (Zubizarreta (2015), Li, Morgan and Zaslavsky (2018)) and identical to the covariate balancing propensity score used to estimate average treatment effects on the treated (Imai and Ratkovic (2014)) and entropy-balancing (Hainmueller (2012), Zhao and Percival (2017)) when a logistic regression model is assumed for the trial assignment model (see Section 2.2). However, these other methods are largely motivated by concerns about the balancing performance of propensity scores estimated from parametric models and focus on estimation and inference in other settings where IPD are fully observed.

Although MAIC has been successfully applied in health technology assessments, in published outcomes research studies (Signorovitch et al. (2012, 2013), Phillippo et al. (2016), Swallow et al. (2016)) and in clinical regulatory evaluations (EMA CHMP (2018)), few formal evaluations of its underlying assumptions and statistical properties exist. We aim to partly address this gap. Specifically, we formalize the problem in the framework of counterfactual outcomes and formulate identification assumptions for causal estimands in the setting where IPD is available for only one study in Sections 2.1 and 2.2. We then study the large sample properties of the MAIC estimator in Section 2.3 and discuss strategies to estimate the standard error in the absence of IPD from the AGD trial in Section 2.4. An investigation of the finite sample performance of the MAIC estimator and standard error estimators is reported in Section 3. We illustrate the method through an application to Duchenne muscular dystrophy in Section 4 and conclude with some further discussion in Section 5. Proofs are deferred to Appendix D.

2. Method.

2.1. *Problem setup and notation.* For each individual i in either the IPD or AGD study, let Y_i denote a continuous or binary outcome and $\mathbf{X}_i$ a p-dimensional vector of baseline covariates belonging to covariate space $\mathcal{X}$. Let $Z_i \in \{0, 1, 2\}$ denote treatment assignment to either a common comparator ($Z_i = 0$) or a treatment studied only in the IPD or AGD trial

($Z_i = 1$ or $Z_i = 2$, respectively) which is randomized within each trial. Let $T_i \in \{1, 2\}$ denote trial assignment to the IPD trial ($T_i = 1$) or AGD trial ($T_i = 2$). A common comparator is assumed to be available here to facilitate some parts of the subsequent exposition, but it is not strictly needed for MAIC. In cases with single arm studies, the common comparator data is omitted and $Z_i = T_i$. The IPD, in general, thus consists of independent and identically distributed (i.i.d.) observations $\mathscr{D}_{\text{IPD}} = \{(Y_i, Z_i, \mathbf{X}_i) : T_i = 1\}$. The AGD consists of data summaries $\mathscr{D}_{\text{AGD}} = \{\bar{Y}_{2z}, \bar{\mathbf{X}}_{2z}, S^2_{Y,2z}, \mathbf{S}^2_{\mathbf{X},2z}, N_{2z} : z = 0, 2\}$, where $N_{tz} = \sum_{T_i=t} I(Z_i = z)$, $\bar{\mathbf{X}}_{tz} = \sum_{T_i=t} \mathbf{X}_i I(Z_i = z)/N_{tz}$ and $\bar{Y}_{tz} = \sum_{T_i=t} Y_i I(Z_i = z)/N_{tz}$. Among arm z of trial t, the sample variance of the outcome is $S^2_{Y,tz} = \sum_{T_i=t}(Y_i - \bar{Y}_{tz})^2 I(Z_i = z)/(N_{tz} - 1)$ and of all covariates is $\mathbf{S}^2_{\mathbf{X},tz} = \sum_{T_i=t}(\mathbf{X}_i - \bar{\mathbf{X}}_{tz})^{\odot 2} I(Z_i = z)/(N_{tz} - 1)$, with $\odot$ denoting elementwise exponentiation. The sample covariance matrix for $\mathbf{X}$ is not fully available since the covariance between covariates is, typically, not reported in publications. Let the total size between trials be $N = \sum_{t,z} N_{tz}$.

The goal of an indirect comparison is to conduct a "fair" comparison of the mean outcomes under treatment 1 to treatment 2, ideally accounting for differences in outcomes due to discrepancies in the distribution of $\mathbf{X}$ between studies. When IPD on both baseline covariates $\mathbf{X}$ and outcomes Y are available for a group of patients, it is potentially possible to reestimate their mean outcomes while adjusting the distribution of $\mathbf{X}$ to more closely match that of another population with sufficiently overlapping support. As IPD is available for patients treated with treatment 1, we can adjust their mean outcomes to more closely match the distribution of $\mathbf{X}$ of those who received treatment 2, that is, the $T = 2$ population, but not vice versa, since only AGD is available for patients who received treatment 2. Let $Y_i(z)$ denote the counterfactual outcome had patient i been treated with treatment z. With these considerations in mind, we take the target estimand to be

$$\Delta = E\{Y(1)|T = 2\} - E\{Y(2)|T = 2\}. \tag{1}$$

This is the average treatment effect on the treated (ATT) among those assigned to treatment 2. When a different scale for the treatment effect contrast is of interest, we can consider a generalization,

$$\Delta_g = g(E\{Y(1)|T = 2\}) - g(E\{Y(2)|T = 2\}),$$

where $g(\cdot)$ is a given link function. For example, when Y is binary, taking $g(u) = \log\{u/(1 - u)\}$ specifies that Δ_g is on the log odds ratio scale (LOR). Regardless of the scale of contrast, the main challenge will be to identify and estimate $E\{Y(1)|T = 2\}$. We will focus on Δ as the target parameter for conciseness and note that a transformation can be applied to obtain Δ_g.

2.2. *Identification.* As Δ is defined in terms of unobserved counterfactual outcomes, we consider the following assumptions required for identification:

ASSUMPTION 2.1. Random treatment within trial: $Z \perp\!\!\!\perp \{\mathbf{X}, Y(1), Y(0)\}|T = 1$ and $Z \perp\!\!\!\perp \{\mathbf{X}, Y(2), Y(0)\}|T = 2$.

ASSUMPTION 2.2. Consistency: $Y = Y(Z)$ with probability 1.

ASSUMPTION 2.3. Positivity of trial assignment: $P(T = 1|\mathbf{X} = \mathbf{x}) \geq \epsilon_{T|\mathbf{X}}$ for all $\mathbf{x} \in \mathcal{X}$, for some $\epsilon_{T|\mathbf{X}} > 0$.

ASSUMPTION 2.4. Ignorability of trial assignment for the outcome under treatment 1: $T \perp\!\!\!\perp Y(1)|\mathbf{X}$.

The first assumption refers to the independence of treatment assignment with covariates and counterfactual outcomes within each trial due to randomization. This assumption would be omitted in the case of single-arm studies. The consistency assumption states that the observed outcome coincides with the counterfactual outcome under the treatment received, which excludes settings with interference and different versions of a treatment. The positivity assumption indicates patients are not assigned to the AGD trial using a deterministic or nearly deterministic rule in $\mathbf{X}$. This is a nontrivial assumption that could be violated, for instance, when inclusion/exclusion criteria are such that the AGD trial includes patients who are excluded from the IPD trial. In such cases the underlying trial populations are not similar enough to conduct an indirect comparison with any method without extrapolating beyond the population of the available IPD. Conversely, the assumption allows for $P(T=2|\mathbf{X}=\mathbf{x})=0$ for some $\mathbf{x}\in\mathcal{X}$ so that there may be patients enrolled in the IPD trial who would be excluded from the AGD trial. In theory, such patients can simply be excluded to achieve balance in the distribution of $\mathbf{X}$ between studies. We offer some more discussion on the effects of violating this assumption in practice in Section 5. The trial assignment ignorability assumption states that we observe sufficient covariates $\mathbf{X}$ such that trial assignment is unrelated to the counterfactual outcome under treatment 1 within strata of $\mathbf{X}$. This is plausible when the patient populations between the two trials are similar enough such that conditioning on the observed covariates $\mathbf{X}$ is enough to control for the differences between study populations that could lead to confounding bias. As adjustments for the the distribution of $\mathbf{X}$ is applied to only the IPD estimates, ignorability is required only for treatment 1 and not for treatment 2.

Based on these assumptions, the second term of Δ can be identified as $E\{Y(2)|T=2\}=E(Y|Z=2,T=2)$ using Assumptions 2.1 and 2.2. The first term of Δ can be identified as

$$
\begin{aligned}
E\big\{Y(1)|T=2\big\} &= E\big[E\big\{Y(1)|\mathbf{X},T=2\big\}|T=2\big] \\
&= E\big[E\{Y|\mathbf{X},Z=1,T=1\}|T=2\big] \\
&= E\left\{\frac{P(T=1)}{P(T=2)}\frac{P(T=2|\mathbf{X})}{P(T=1|\mathbf{X})}E(Y|T=1,\mathbf{X},Z=1)|T=1\right\} \\
&= E\left\{\frac{I(Z=1)}{P(Z=1|T=1)}\frac{I(T=1)}{P(T=2)}\omega(\mathbf{X})Y\right\} \\
&= E\big\{I(Z=1,T=1)\omega(\mathbf{X})Y\big\}/E\big\{I(Z=1,T=1)\omega(\mathbf{X})\big\},
\end{aligned}
\tag{2}
$$

where $\omega(\mathbf{X})=P(T=2|\mathbf{X})/P(T=1|\mathbf{X})$ denotes the odds of trial assignment given $\mathbf{X}$. We used Assumptions 2.1, 2.2, 2.4 in the second equality, 2.3 in the third and fourth equalities and 2.1 in the fourth and final equalities. This suggests that Δ can be identified if $\omega(\mathbf{X})$ can be identified. Even if $\omega(\mathbf{X})$ could not be identified exactly, the last equality shows it would be sufficient if $\omega(\mathbf{X})$ could be identified up to a constant due to the ratio of terms involving $\omega(\mathbf{X})$.

Before proceeding, we first discuss an alternative to Assumption 2.4 that leverages the common comparator arm when data on it is available:

ASSUMPTION 2.5. Ignorability of trial assignment for the counterfactual difference: $T \perp\!\!\!\perp \{Y(1)-Y(0)\}|\mathbf{X}$.

To make use of this assumption, the target parameter can be written $\Delta = E\{Y(1)-Y(0)|T=2\}-E\{Y(2)-Y(0)|T=2\}$, where $E\{Y(2)-Y(0)|T=2\}$ is straightforward to

identify due to randomization of treatment within trial. The first term can be identified by

$$
\begin{aligned}
E\{Y(1)-Y(0)|T=2\} &= E\big[E\{Y(1)-Y(0)|\mathbf{X},T=2\}|T=2\big]\\
&= E\big[E\{Y(1)-Y(0)|\mathbf{X},T=1\}|T=2\big]\\
&= E\big[E\{Y|\mathbf{X},Z=1,T=1\}-E\{Y|\mathbf{X},Z=0,T=1\}|T=2\big]\\
&= E\left\{\frac{I(Z=1)}{P(Z=1|T=1)}\frac{I(T=1)}{P(T=2)}\frac{P(T=2|\mathbf{X})}{P(T=1|\mathbf{X})}Y\right\}\\
&\quad - E\left\{\frac{I(Z=0)}{P(Z=0|T=1)}\frac{I(T=1)}{P(T=2)}\frac{P(T=2|\mathbf{X})}{P(T=1|\mathbf{X})}Y\right\},
\end{aligned}
\tag{3}
$$

where we use Assumption 2.5 in the second equality and Assumptions 2.1 and 2.2 in the third equality. The final steps proceed, as in (2), which uses Assumptions 2.3 and 2.1. This alternative assumption could be easier to justify in some cases. For example, suppose the outcome model can be specified by

$$E(Y|Z,\mathbf{X}) = \beta_0 + \beta_1 I(Z\neq 0) + \boldsymbol{\beta}_2^{\mathsf{T}}\mathbf{X} + \boldsymbol{\beta}_3^{\mathsf{T}} I(Z\neq 0)\mathbf{X}_M + \beta_4 I(Z=2),$$

where $\mathbf{X}_M \subseteq \mathbf{X}$ is a subset of the prognostic covariates that modify the treatment effect of the active treatments and the components of $\boldsymbol{\beta}_3$ are not exactly the negative of the components of $\boldsymbol{\beta}_2$ corresponding to the same covariate in $\mathbf{X}_M$. In such cases it suffices to adjust only for the treatment effect modifiers $\mathbf{X}_M$ rather than the full set of $\mathbf{X}$ that may differ in distribution between trials (Phillippo et al. (2016)). This approach can also be used to identify Δ_g for a nonlinear $g(\cdot)$ by writing $E\{Y(1)|T=2\} = E\{Y(1)-Y(0)|T=2\} + E\{Y(0)|T=2\}$. The second term can be easily identified from randomization within trial. This approach offers an alternative identification strategy but still relies on strong assumptions about the form of the conditional mean of Y and assumes covariates are known not to be effect modifiers on the additive scale. Despite its appeal, Assumption 2.5 is not necessarily weaker than 2.4 and should be evaluated based on clinical input to the extent possible in practice.

It now remains for us to identify $\omega(\mathbf{X})$ from the observed data. Had IPD been available for both trials, then this would be straightforward. In the absence of the full IPD, under Assumption 2.3, $\omega(\mathbf{X})$ is a solution in $h(\mathbf{X})$ to the integral equation

$$E\{h(\mathbf{X})I(T=1)\overrightarrow{\mathbf{X}}\} - \boldsymbol{\mu}^*_{\overrightarrow{\mathbf{X}}_2} P(T=2) = \mathbf{0}, \tag{4}$$

where $\overrightarrow{\mathbf{X}} = (1,\mathbf{X}^{\mathsf{T}})^{\mathsf{T}}$ and $\boldsymbol{\mu}^*_{\overrightarrow{\mathbf{X}}_2} = E(\overrightarrow{\mathbf{X}}|T=2)$. If $h^*(\mathbf{X})$ is a solution to this equation and the solution is *unique*, then it must be that $h^*(\mathbf{X}) = \omega(\mathbf{X})$. In general, without any restrictions on $h(\mathbf{X})$, there may not be a unique solution. However, $h(\mathbf{X})$ can be reasonably parameterized such that there exists a unique solution. For example, suppose the trial assignment follows a logistic regression model,

$$\operatorname{logit} P(T=2|\mathbf{X}) = \boldsymbol{\alpha}^{\mathsf{T}}\overrightarrow{\mathbf{X}}, \tag{5}$$

for some $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_1^{\mathsf{T}})^{\mathsf{T}}$. Then under this model, $\omega(\mathbf{X}) = \exp(\boldsymbol{\alpha}^{*\mathsf{T}}\overrightarrow{\mathbf{X}})$, where $\boldsymbol{\alpha}^*$ is the true value of $\boldsymbol{\alpha}$. Restricting $h(\mathbf{X}) = \omega(\mathbf{X};\boldsymbol{\alpha}) = \exp(\boldsymbol{\alpha}^{\mathsf{T}}\overrightarrow{\mathbf{X}})$, (4) admits a unique solution because $Q(\boldsymbol{\alpha}) = E\{\omega(\mathbf{X};\boldsymbol{\alpha})I(T=1)\} - \boldsymbol{\alpha}^{\mathsf{T}}\boldsymbol{\mu}^*_{\overrightarrow{\mathbf{X}}_2} P(T=2)$ is strictly convex in $\boldsymbol{\alpha}$ (Signorovitch et al. (2010)).

Other moment conditions can also be used. Since $\omega(\mathbf{X})$ needs only to be identified up to a constant, an alternative condition is

$$E\{h(\mathbf{X}-\boldsymbol{\mu}^*_{\mathbf{X}_2})I(T=1)(\mathbf{X}-\boldsymbol{\mu}^*_{\mathbf{X}_2})\} = \mathbf{0}, \tag{6}$$

where $\boldsymbol{\mu}^*_{\mathbf{X}_2} = E(\mathbf{X}|T=2)$. Under model (5), $\boldsymbol{\alpha}_1$ can be identified by similar arguments by restricting $h(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}_2}) = \omega(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}_2};\boldsymbol{\alpha}_1) = \exp\{\boldsymbol{\alpha}_1^\mathsf{T}(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}_2})\}$. This approach identifies $\omega(\mathbf{X})$ up to a scalar constant and avoids estimation of an additional intercept parameter that is not needed for estimating Δ. Another potentially useful condition is

$$E\big\{h(\mathbf{X})I(T=1)\mathbf{t}(\mathbf{X})\big\} - E\big\{\mathbf{t}(\mathbf{X})|T=2\big\}P(T=2) = \mathbf{0}, \tag{7}$$

where we take $\mathbf{t}(\mathbf{X}) = (1, X_1, \ldots, X_p, X_1^2, \ldots, X_p^2)^\mathsf{T}$ to be elements of $\mathbf{X}$ and their squares. In contrast to conditions (4) and (6), using this condition in practice would require the availability of the sample variances of the covariates in the AGD to estimate $E\{\mathbf{t}(\mathbf{X})|T=2\}$. If we again restrict $h(\mathbf{X}) = \omega(\mathbf{X};\boldsymbol{\alpha}^\dagger) = \exp\{\boldsymbol{\alpha}^{\dagger\mathsf{T}}\mathbf{t}(\mathbf{X})\}$, where $\boldsymbol{\alpha}^\dagger = (\boldsymbol{\alpha}^\mathsf{T}, \boldsymbol{\alpha}_2^\mathsf{T})^\mathsf{T}$, then $\omega(\mathbf{X})$ is again identified under the more general trial assignment model

$$\operatorname{logit} P(T=2|\mathbf{X}) = \boldsymbol{\alpha}^{\dagger\mathsf{T}}\mathbf{t}(\mathbf{X}).$$

The model from (5) can be viewed as a submodel of this model that restricts $\boldsymbol{\alpha}_2 = \mathbf{0}$. When (5) is correct, fitting this larger model comes at the cost of finite-sample efficiency loss. However, if this expanded model is correct, then using (4) for estimation would yield biased estimates of $\omega(\mathbf{X})$ and Δ. Balancing the first and second moments with this expanded model thus involves trade-off between robustness and efficiency. We next discuss estimation of $\omega(\mathbf{X})$ and Δ based on these identification conditions.

2.3. *Estimation.* For the rest of the paper, except for Section 2.4, we will assume a model where trial assignment is correctly specified by (5). We first estimate $\omega(\mathbf{X})$ by solving empirical versions of the proposed moment conditions. For instance, let $\widehat{\boldsymbol{\alpha}}_1$ be the solution to the equation

$$N^{-1}\sum_{i=1}^{N}\omega(\mathbf{X}_i - \bar{\mathbf{X}}_2;\boldsymbol{\alpha}_1)I(T_i=1)(\mathbf{X}_i - \bar{\mathbf{X}}_2) = \mathbf{0}, \tag{8}$$

where $\bar{\mathbf{X}}_2 = (\bar{\mathbf{X}}_{22}N_{22} + \bar{\mathbf{X}}_{20}N_{20})/(N_{22}+N_{20})$. We use the moment condition from (6) to avoid estimating the additional intercept parameter. The following result states that $\widehat{\boldsymbol{\alpha}}_1$ is consistent and asymptotically linear and thus also asymptotically normal.

THEOREM 2.1. *If the trial assignment model is correctly specified by* (5)*, then* $\widehat{\boldsymbol{\alpha}}_1 \xrightarrow{p} \boldsymbol{\alpha}_1^*$*, where* $\boldsymbol{\alpha}_1^*$ *is the true coefficient in* (5)*. Furthermore,* $\widehat{\boldsymbol{\alpha}}_1$ *is asymptotically linear such that*

$$N^{1/2}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*) = N^{-1/2}\sum_{i=1}^{N}\boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}^*_{\mathbf{X}_2}) + o_p(1), \tag{9}$$

where $\boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) = \mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\{\mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) + \mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\mu}_{\mathbf{X}_2}, \boldsymbol{\alpha}_1)\}$ *is a mean zero random vector with finite variance and*

$$\begin{aligned}
\mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) &= (\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\exp\{\boldsymbol{\alpha}_1^\mathsf{T}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\}I(T_i=1),\\
\mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\mu}_{\mathbf{X}_2}, \boldsymbol{\alpha}_1) &= -E\big[\exp\{\boldsymbol{\alpha}_1^\mathsf{T}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\}I(T_i=1)\big](\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\frac{I(T_i=2)}{P(T_i=2)},\\
\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) &= -E\big[(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})^\mathsf{T}\exp\{\tilde{\boldsymbol{\alpha}}_1^\mathsf{T}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\}I(T_i=1)\big].
\end{aligned}$$

This expansion reveals two sources that contribute to the asymptotic variance. Suppressing implicit arguments for the parameters, the $\mathbf{U}_i^{\boldsymbol{\alpha}_1}$ term is contributed from estimating $\boldsymbol{\alpha}_1$ when $\boldsymbol{\mu}_{\mathbf{X}_2}$ is known. The $\mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}$ term is the additional contribution when $\boldsymbol{\mu}_{\mathbf{X}_2}$ is considered to be

estimated by $\bar{\mathbf{X}}_2$. Though this expansion clarifies the sources of variability, the influence function cannot be directly used to compute the asymptotic variance since $\mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}$ involves IPD from the AGD trial.

Following estimation of $\boldsymbol{\alpha}_1$, Δ can subsequently be estimated by an empirical version of (2), plugging in $\omega(\mathbf{X}; \widehat{\boldsymbol{\alpha}}_1)$ for $\omega(\mathbf{X})$. We consider identification based on (2) instead of (3) for conciseness of presentation, but similar results will hold if (3) is used. Specifically, the estimator can be expressed as

$$\widehat{\Delta} = \sum_{i=1}^{N} I(Z_i = 1, T_i = 1)\omega(\mathbf{X}_i; \widehat{\boldsymbol{\alpha}}_1)Y_i / \sum_{i=1}^{N} I(Z_i = 1, T_i = 1)\omega(\mathbf{X}_i; \widehat{\boldsymbol{\alpha}}_1) - \bar{Y}_{22}. \tag{10}$$

The following result states that $\widehat{\Delta}$ is consistent and asymptotically linear and thus also asymptotically normal.

THEOREM 2.2. *Suppose that the identification assumptions* (2.1), (2.2), (2.3) *and* (2.4) *hold, and the trial assignment model is correctly specified by* (5). *Then,* $\widehat{\Delta} \xrightarrow{p} \Delta^*$, *where* Δ^* *is the true target parameter* Δ. *Furthermore,* $\widehat{\Delta}$ *is asymptotically linear such that*

$$N^{1/2}\big(\widehat{\Delta} - \Delta^*\big) = N^{-1/2} \sum_{i=1}^{N} \varphi_i\big(\Delta^*, \mu_1^*, \boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*\big) + o_p(1), \tag{11}$$

where $\mu_1^* = E\{Y(1)|T = 2\}$ *is the true counterfactual mean for the IPD treatment in the AGD population and* $\varphi_i(\Delta, \mu_1, \boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) = \varphi_i^{\mu_2}(\Delta, \mu_1) + \varphi_i^{\mu_1}(\mu_1, \boldsymbol{\alpha}_1) + J^{\mu_1}(\boldsymbol{\alpha}_1)^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1, \boldsymbol{\alpha}_1)^{\mathsf{T}} \times \boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})$ *is a mean zero random variable with finite variance and*

$$\begin{aligned}
\varphi_i^{\mu_2}(\Delta, \mu_1) &= (\mu_1 - Y_i - \Delta)\frac{I(Z_i = 2, T_i = 2)}{P(Z_i = 2, T_i = 2)},\\
\varphi_i^{\mu_1}(\mu_1, \boldsymbol{\alpha}_1) &= J^{\mu_1}(\boldsymbol{\alpha}_1)^{-1}(Y_i - \mu_1)e^{\boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{X}_i} I(Z_i = 1, T_i = 1),\\
J^{\mu_1}(\boldsymbol{\alpha}_1) &= E\big\{e^{\boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{X}_i} I(Z_i = 1, T_i = 1)\big\},\\
\widetilde{\boldsymbol{C}}_1(\mu_1, \boldsymbol{\alpha}_1) &= E\big\{\mathbf{X}_i(Y_i - \mu_1)e^{\boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{X}_i} I(Z_i = 1, T_i = 1)\big\}.
\end{aligned}$$

The first $\varphi_i^{\mu_2}$ term accounts for estimation of $\mu_2^* = E\{Y(2)|T = 2\}$ from $\bar{Y}_{22}$. The subsequent terms account for estimating μ_1^* through weighting, which can be further decomposed into a term contributed for estimating μ_1^* when $\boldsymbol{\alpha}_1$ is known and another term for estimating $\boldsymbol{\alpha}_1$. Again, the asymptotic variance cannot be directly computed from this influence function because $\varphi_i^{\mu_2}$ and $\boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}$ involve the IPD from the AGD trial. We argue in Section 2.4 that it is often sufficient to compensate for ignoring this term by simply incorporating the marginal variance of $\bar{Y}_{22}$ from the AGD trial and consider other potential strategies.

2.4. *Estimation of asymptotic variance.* Estimating the asymptotic variance of $\widehat{\Delta}$ is complicated by the fact that $\widehat{\Delta}$ depends on $\bar{\mathbf{X}}_2$ and $\bar{Y}_{22}$, and the IPD is not available to estimate contributions that account for their variability. As we discuss in Appendix C, though one can regard $\boldsymbol{\mu}_{\mathbf{X}_2}^* = \bar{\mathbf{X}}_2$ and $\mu_2^* = \bar{Y}_{22}$ to be fixed in the sampling scheme, this may not always be justifiable. In the following we consider strategies to estimate the full asymptotic variance, regarding $\bar{\mathbf{X}}_2$ and $\bar{Y}_{22}$ as random, in the absence of the full IPD. To facilitate the subsequent

considerations, we first define the following contributions to the influence function for $\widehat{\Delta}$ for estimating $\boldsymbol{\alpha}_1$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$:

$$\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \mu_1, \boldsymbol{\mu}_{\mathbf{X}_2}) = J^{\mu_1}(\boldsymbol{\alpha}_1)^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1, \boldsymbol{\alpha}_1)^{\mathsf{T}}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}),$$
$$\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\mu}_{\mathbf{X}_2}, \mu_1, \boldsymbol{\alpha}_1) = J^{\mu_1}(\boldsymbol{\alpha}_1)^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1, \boldsymbol{\alpha}_1)^{\mathsf{T}}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}).$$

By calculating the variance of $\varphi_i(\Delta^*, \mu_1^*, \boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)$, the full asymptotic variance of $\widehat{\Delta}$ can now be expressed as

$$\begin{aligned}\sigma^2 = {} & \{\operatorname{Var}(\varphi_i^{\mu_1}) + \operatorname{Var}(\varphi_i^{\mu_2})\} + \{\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}) + 2\operatorname{Cov}(\varphi_i^{\mu_1}, \widetilde{\varphi}_i^{\boldsymbol{\alpha}_1})\} \\ & + \{\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}) + 2\operatorname{Cov}(\varphi_i^{\mu_2}, \widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}})\},\end{aligned} \tag{12}$$

where the arguments of the components of the influence function are suppressed but implicitly evaluated at their respective truth. Decomposing the variance this way, the first two terms constitutes the asymptotic variance had $\boldsymbol{\alpha}_1$ been known. The second two terms are contributed from estimating $\boldsymbol{\alpha}_1$ had $\boldsymbol{\mu}_{\mathbf{X}_2}$ been known, and the final two terms account for estimating $\boldsymbol{\mu}_{\mathbf{X}_2}$.

It is generally not possible to fully estimate σ^2 without further assumptions, as $\varphi_i^{\mu_2}$ and $\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}$ involve IPD from the AGD trial. However, it may still be possible to obtain reasonable approximations. The following lemma further clarifies the form of the additional contributions from estimating $\boldsymbol{\alpha}_1$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$ under correct identification and modeling assumptions:

LEMMA 2.1. *Let identification assumptions* (2.1), (2.2), (2.3) *and* (2.4) *be satisfied, and the trial assignment model be correctly specified by* (5). *The terms contributed from estimating* $\boldsymbol{\alpha}_1$ *from* (12) *can be simplified as*

$$\begin{aligned}&\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}) = P(T=2)^{-1}\boldsymbol{C}_1^{\mathsf{T}}\operatorname{Var}(\mathbf{X}|T=2)^{-1}, \\ &E\{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}_2}^*)(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}_2}^*)^{\mathsf{T}}\omega(\mathbf{X})|T=2\}\operatorname{Var}(\mathbf{X}|T=2)^{-1}\boldsymbol{C}_1, \\ &\operatorname{Cov}(\varphi_i^{\mu_1}, \widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}) = -P(T=2)^{-1}\boldsymbol{C}_1^{\mathsf{T}}\operatorname{Var}(\mathbf{X}|T=2)^{-1}, \\ &E\{(Y(1)-\mu_1^*)(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}_2}^*)\omega(\mathbf{X})|T=2\},\end{aligned}$$

where $\boldsymbol{C}_1 = \operatorname{Cov}\{Y(1), \mathbf{X}|T=2\}$. *Moreover, the terms contributed for estimating* $\boldsymbol{\mu}_{\mathbf{X}_2}$ *from* (12) *can be simplified as*

$$\begin{aligned}\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}) &= P(T=2)^{-1}\boldsymbol{C}_1^{\mathsf{T}}\operatorname{Var}(\mathbf{X}|T=2)^{-1}\boldsymbol{C}_1, \\ \operatorname{Cov}(\varphi_i^{\mu_2}, \widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}) &= -P(T=2)^{-1}\boldsymbol{C}_1^{\mathsf{T}}\operatorname{Var}(\mathbf{X}|T=2)^{-1}\boldsymbol{C}_2,\end{aligned}$$

where $\boldsymbol{C}_2 = \operatorname{Cov}\{Y(2), \mathbf{X}|T=2\}$.

As long as there are no strong interactions between $\mathbf{X}$ and the treatment, it can be expected that $\boldsymbol{C}_1 \approx \boldsymbol{C}_2$. In this case, if, additionally, the trial assignment model is at least approximately correctly specified, then $\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}) + 2\operatorname{Cov}(\varphi_i^{\mu_2}, \widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}) \approx -\boldsymbol{C}_1^{\mathsf{T}}\operatorname{Var}(\mathbf{X}|T=2)\boldsymbol{C}_1 < 0$. Omitting the contributions from estimating $\boldsymbol{\mu}_{\mathbf{X}_2}$ when estimating σ^2 thus tends to produce conservative standard errors. By bounding $\omega(\mathbf{X})$, a similar phenomenon occurs for terms contributed from estimating $\boldsymbol{\alpha}_1$ so that $\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}) + 2\operatorname{Cov}(\varphi_i^{\mu_1}, \widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}) < 0$ when the trial assignment model is correct and the lower and upper bounds for $\omega(\mathbf{X})$ are not too extreme. These results suggest that omitting contributions for estimating both $\boldsymbol{\alpha}_1$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$ can yield conservative standard errors in scenarios where the trial assignment model is correctly specified.

A simple approach to estimating σ^2 is thus to estimate only $\mathrm{Var}(\varphi_i^{\mu_1})$ and $\mathrm{Var}(\varphi_i^{\mu_2})$, fully omitting contributions for $\boldsymbol{\alpha}_1$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$, as in

$$\widehat{\sigma}^2_{fo} = \widehat{\mathrm{Var}}\{\varphi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1)\} + \widehat{\mathrm{Var}}\{\varphi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}, \tag{13}$$

where $\widehat{\mu}_1$ is the weighted average from the IPD as in the first term of (10), $\widehat{\mathrm{Var}}\{\varphi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1)\}$ is the sample variance of $\varphi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1)$ and $\widehat{\mathrm{Var}}\{\varphi_i^{\widetilde{\mu}_2}(\widehat{\Delta}, \widehat{\mu}_1)\} = S^2_{Y,22}/(N_{22}/N)$. Previous approaches to estimating standard errors for MAIC using robust sandwich estimators (Signorovitch et al. (2010), Phillippo et al. (2016)), which are sometimes utilized in practice, are similar to this approach in that they ignore the variability from estimating $\boldsymbol{\alpha}_1$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$. Instead of fully ignoring the contributions for both $\boldsymbol{\alpha}_1$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$, another approach is to partially omit only the contribution for $\boldsymbol{\mu}_{\mathbf{X}_2}$, as in

$$\widehat{\sigma}^2_{po} = \widehat{\mathrm{Var}}\{\varphi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1) + \widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\mu}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2})\} + \widehat{\mathrm{Var}}\{\varphi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}, \tag{14}$$

where $\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} = \bar{\mathbf{X}}_2$. This is still feasible, as $\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}$ does not involve IPD from the AGD trial. A final approach is to attempt to fully estimate σ^2. Without any further assumptions, $\mathrm{Var}(\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}})$ can be approximated by

$$V^{\boldsymbol{\mu}_{\mathbf{X}_2}} = -\frac{E\{e^{\boldsymbol{\alpha}_1^{*\mathsf{T}}(\mathbf{X}_i - \boldsymbol{\mu}^*_{\mathbf{X}_2})} I(T_i = 1)\}}{J^{\mu_1}(\boldsymbol{\alpha}_1^*)^2 P(T_i = 2)} \widetilde{\boldsymbol{C}}_1(\boldsymbol{\alpha}_1^*)^{\mathsf{T}} \mathbf{J}^{\boldsymbol{\alpha}_1^*}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}^*_{\mathbf{X}_2})^{-1} \widetilde{\boldsymbol{C}}_1(\boldsymbol{\alpha}_1^*),$$

which partially simplifies $\mathrm{Var}(\varphi_i^{\widetilde{\boldsymbol{\mu}}_{\mathbf{X}_2}})$ under correct trial selection model to obviate the need for IPD from the AGD trial. The covariance term $\mathrm{Cov}(\varphi_i^{\mu_2}, \widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}})$ can be bounded by the Cauchy–Schwartz inequality. This suggests estimating σ^2 by

$$\begin{aligned}\widehat{\sigma}^2_{cs} = {} & \widehat{\mathrm{Var}}\{\varphi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1) + \widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\mu}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2})\} + \widehat{\mathrm{Var}}\{\varphi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\} \\ & + \widehat{V}^{\boldsymbol{\mu}_{\mathbf{X}_2}} + 2\big[\widehat{\mathrm{Var}}\{\varphi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}\widehat{V}^{\boldsymbol{\mu}_{\mathbf{X}_2}}\big]^{1/2},\end{aligned} \tag{15}$$

where $\widehat{V}^{\boldsymbol{\mu}_{\mathbf{X}_2}}$ is an empirical version of $V^{\boldsymbol{\mu}_{\mathbf{X}_2}}$. Among these proposed approaches, we expect $\widehat{\sigma}^2_{fo}$ to be more conservative than $\widehat{\sigma}^2_{po}$, as it potentially omits negative contributions to the asymptotic variance when underlying assumptions are satisfied. We will see in the simulation results of Section 3 that this conservativeness of $\widehat{\sigma}^2_{fo}$ tends to improve its accuracy in approximating the true standard error in small samples, when both $\widehat{\sigma}^2_{fo}$ and $\widehat{\sigma}^2_{po}$ tend to underestimate, without paying a large price in terms of overestimation in large samples. Such underestimation in small samples has also been observed for sandwich variance estimators in other settings (Kauermann and Carroll (2001), Fay and Graubard (2001)). $\widehat{\sigma}^2_{cs}$ can be expected to be the most conservative, as it uses a very conservative bound for covariance. This estimator could be potentially useful in situations when conservative confidence intervals and hypothesis tests are prioritized over efficiency considerations.

3. Simulations. We performed simulations to assess the finite sample bias of MAIC and alternative estimators. In particular, we sought to identify scenarios where proposed approaches fail to provide reliable inferences. We also assessed the coverage and relative length of CIs based on the proposed variance estimators. Besides the estimator $\widehat{\Delta}$ (MAIC-NAB), which is based on (2), we also consider an anchored MAIC approach (MAIC-ACB), based on (3), $\widehat{\Delta}^{\mathrm{ACB}} = \widehat{\Delta} - \{\sum_{i=1}^N I(Z_i = 0, T_i = 1)\omega(\mathbf{X}_i; \widehat{\boldsymbol{\alpha}}_1) Y_i / \sum_{i=1}^N I(Z_i = 0, T_i = 1)\omega(\mathbf{X}_i; \widehat{\boldsymbol{\alpha}}_1) - \bar{Y}_{20}\}$. Additionally, we consider the method of Bucher et al. (1997) (BUC) and the basic formulation of simulated treatment comparisons of Ishak, Proskorovsky and Benedict (2015) (STC). To be consistent in the scale of contrast with other methods, we implement BUC on the risk difference scale as $\widehat{\Delta}^{\mathrm{BUC}} = (\bar{Y}_{11} - \bar{Y}_{10}) - (\bar{Y}_{22} - \bar{Y}_{20})$. As such,

BUC would require that $E(Y|Z,\mathbf{X})$ be linear in $\mathbf{X}$ to be unbiased, where $\mathbf{X}$ are covariates that satisfy Assumption 2.5. We also considered simulations for contrasts on the log odds ratio scale for all methods in Appendix B. For STCs we assume a logistic regression model for $E(Y|Z=1,\mathbf{X})$,

$$m_1(\mathbf{X};\boldsymbol{\gamma}) = g(\boldsymbol{\gamma}^\mathsf{T}\overrightarrow{\mathbf{X}}),$$

where $g(\cdot)$ denotes the inverse-logit link function for binary outcomes in the simulations. This model is fit using data from active arm in the IPD trial with the estimator denoted by $\widehat{\boldsymbol{\gamma}}$. It is then used to extrapolate the mean outcome had individuals in the AGD trial received treatment $Z=1$ by $\widehat{\Delta}^{\mathrm{STC}} = m_1(\bar{\mathbf{X}}_2;\widehat{\boldsymbol{\gamma}}) - \bar{Y}_{22}$. This approach would be unbiased if the model for $E(Y|Z=1,\mathbf{X})$ is correctly specified with $g(\cdot)$ being linear. But even if $m_1(\mathbf{X};\boldsymbol{\gamma})$ is correctly specified and $\mathbf{X}$ satisfy the identification assumptions, still $E\{Y(1)|T=2\} = E\{E(Y|Z=1,\mathbf{X})|T=2\} \neq E\{Y|Z=1,\mathbf{X}=E(\mathbf{X}|T=2)\} = m(\boldsymbol{\mu}_{\mathbf{X}_2};\boldsymbol{\gamma})$ when $g(\cdot)$ is nonlinear which results in bias.

We simulated data jointly for both the IPD and AGD trials in the case with binary Y and continuous $\mathbf{X}$. In all scenarios, independent observations were simulated according to $\mathbf{X} \sim N(\mathbf{0}, 0.8\mathbf{I}_p + .2)$, $T|\mathbf{X} \sim \mathrm{Ber}\{P(T=2|\mathbf{X})\}+1$, $Z \sim \mathrm{Ber}(0.5)\cdot T$ and $Y|\mathbf{X},Z,T \sim \mathrm{Ber}\{E(Y|\mathbf{X},Z,T)\}$, where

$$\begin{aligned} \operatorname{logit} P(T=2|\mathbf{X}) &= \alpha_0 + \boldsymbol{\alpha}_1^\mathsf{T}\mathbf{X}, \\ \operatorname{logit} E(Y|\mathbf{X},Z,T) &= \beta_0 + \{\boldsymbol{\beta}_1^\mathsf{T} + \boldsymbol{\beta}_3^\mathsf{T} I(Z>0)\}\mathbf{X} + \beta_2 I(Z>0) + I(Z=2)\beta_4, \end{aligned} \tag{16}$$

with $\alpha_0 = 0$ $\beta_0 = -1$, $\beta_2 = 0.1$ and $\beta_4 = .5$. For $P(T=2|\mathbf{X})$, $\boldsymbol{\alpha}_1 \neq \mathbf{0}$ induces imbalance in the distribution of $\mathbf{X}$ in the IPD and AGD trial populations. Any imbalance in a subvector of $\mathbf{X}$ that also has a nonzero coefficient in $\boldsymbol{\beta}_3$ subsequently induces confounding when comparing outcomes between trials. $\boldsymbol{\beta}_3 \neq \mathbf{0}$ results in treatment effect heterogeneity on the logit scale between active and placebo treatments. We considered scenarios with no confounding, moderate confounding and severe confounding, with the parameters

$$\begin{aligned} &\textit{None}: \quad \boldsymbol{\alpha}_1 = (0.25\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}, \qquad \boldsymbol{\beta}_1 = \mathbf{0}_p \boldsymbol{\beta}_3 = \mathbf{0}_p, \\ &\textit{Moderate}: \quad \boldsymbol{\alpha}_1 = (0.25\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}, \qquad \boldsymbol{\beta}_1 = (0.15\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}, \qquad \boldsymbol{\beta}_3 = (0.1\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}, \\ &\textit{Severe}: \quad \boldsymbol{\alpha}_1 = (0.30\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}, \qquad \boldsymbol{\beta}_1 = (0.25\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}, \qquad \boldsymbol{\beta}_3 = (0.15\mathbf{1}_4^\mathsf{T}, \mathbf{0}_{p-4}^\mathsf{T})^\mathsf{T}. \end{aligned}$$

This setup leads to a standardized mean difference (Cohen (2013)) of approximately -0.38 for the first four covariates and -0.18 for the remaining covariates in the none and moderate settings and -0.45 and -0.22 in the severe setting. To get a sense of the treatment effect heterogeneity, the standard deviation of the true differences in the outcome probability for active vs. placebo treatments in the IPD and AGD trials are 0 and 0 for the none setting, 0.05 and 0.07 for the moderate setting and 0.07 and 0.09 for the severe setting. We initially generated a large sample $\{(Y_i, Z_i, T_i, \mathbf{X}_i) : i = 1,\ldots,N^*\}$ and then randomly subsampled by arm in each trial, as described in Appendix C, to include a fixed number of n patients in all arms in the final sample. We also provide arguments there to justify that the proposed procedures for inference are still valid under this modified sampling scheme.

In each replicate of the data, we calculated the four estimators for Δ, as well as the four estimators for σ^2 discussed in Section 2.4. Besides $\widehat{\sigma}^2_{fo}$, $\widehat{\sigma}^2_{po}$ and $\widehat{\sigma}^2_{cs}$, we also implemented an estimator $\widehat{\sigma}^2_{sw}$ based on a sandwich estimator for regression coefficients when $\widehat{\Delta}$ is implemented through a weighted linear regression, using the `sandwich` package in R with default options (Zeileis (2004)). The approach has been previously considered for $\widehat{\Delta}^{\mathrm{ACB}}$ (Phillippo et al. (2016)) and is essentially a direct calculation of the variance of $\widehat{\Delta}$, treating the weights $\omega(\mathbf{X}_i)$ and treatment assignment T_i as fixed and plugging in estimators for

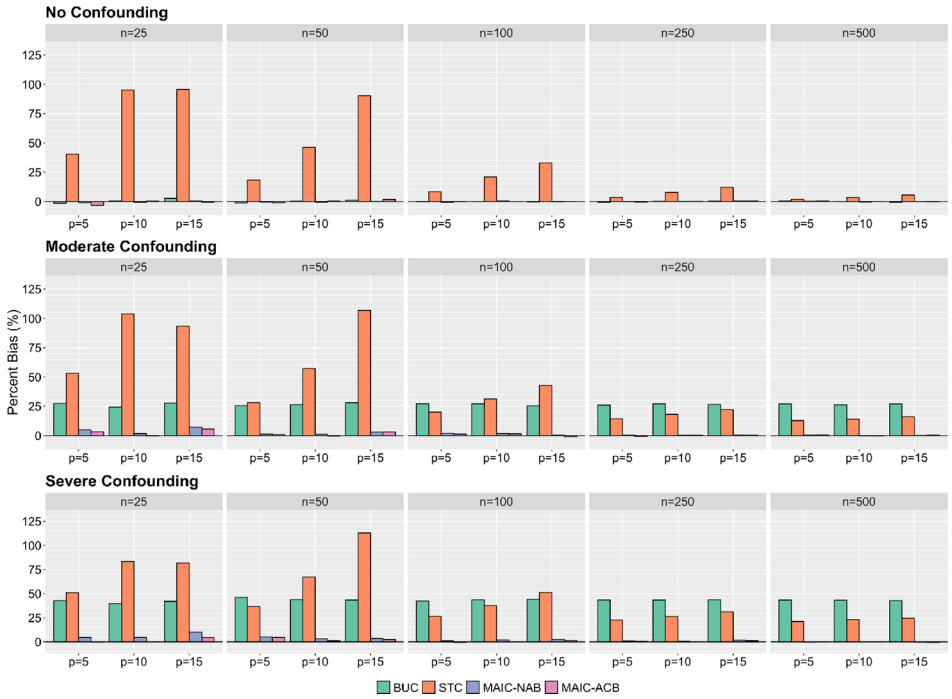
the variance of Y_i that allow for heteroskedasticity. For benchmarking purposes we also calculated an estimator of σ^2 using the full influence function, the estimator that would typically be computed had the IPD from all trials been available. The true Δ was calculated through simulating a large sample and calculating the mean difference of counterfactual outcomes had patients in the AGD population received treatment $Z = 1$ and $Z = 2$, according to $E(Y|\mathbf{X}, Z, T)$ in (16). The percent bias for each estimator was then calculated as $R^{-1}\sum_{r=1}^{R}(\widehat{\Delta}^{(r)} - \Delta)/\Delta$, where $\widehat{\Delta}^{(r)}$ denotes an estimator calculated from data in the rth replicate. The bias simulations were repeated over the different confounding scenarios for sample sizes ranging $n = 25$ to $n = 500$ per arm and $p = 5, 10, 15$. The CI coverage was calculated as $R^{-1}\sum_{r=1}^{R} I[\Delta \in \{\widehat{\Delta}^{(r)} \pm z_{0.975}N^{-1/2}\widehat{\sigma}^{(r)}\}]$, where $z_{0.975}$ denotes the .975 quantile of a standard normal and $\widehat{\sigma}^{(r)}$ denotes an estimator of the asymptotic variance estimated from data observed in the rth replicate. Relative CI length was calculated as $R^{-1}\sum_{r=1}^{R}\widehat{\sigma}^{(r)}/\widehat{\sigma}_{\text{emp}}$, where $\widehat{\sigma}^2_{\text{emp}} = (R-1)^{-1}\sum_{r=1}^{R}\{\widehat{\Delta}^{(r)} - R^{-1}\sum_{r=1}^{R}\widehat{\Delta}^{(r)}\}^2$ is the empirical variance of $\widehat{\Delta}^{(r)}$ over all repetitions. The coverage simulations were conducted under the moderate confounding scenario for $n = 25$ to $n = 500$ per arm with $p = 5, 15$. For each set of simulations, we ran $R = 5000$ replicates to well approximate the tail probabilities when evaluating CI coverage.

3.1. *Simulation results*. The results for the bias simulations are presented in Figure 1. Both MAIC estimators generally exhibited negligible bias across the scenarios considered. The bias of BUC increases with the degree of confounding and results from the nonlinear link function for $E(Y|\mathbf{X}, Z, T)$ and interaction between $\mathbf{X}$ and Z in (16). Similar results hold even when considering contrasts on the log odds-ratio scale in Appendix B. STC also incurred bias that increases with the degree of confounding due to the nonlinear link. STC, however, generally had lower bias than BUC for larger n. Extrapolation based on fitting $m_1(\mathbf{X}; \boldsymbol{\gamma})$ appears to outperform placebo adjustment in BUC in large samples. The bias of STC in small samples also appears to be more sensitive to increasing p than MAIC and BUC.

Results from the coverage simulations are presented in Figure 2. CIs based on $\widehat{\sigma}^2_{fo}$ and $\widehat{\sigma}^2_{sw}$ generally achieve close to nominal coverage for $n \geq 50$ per arm. They are only slightly conservative in terms of both coverage and length, relative to empirical estimates, in large samples. Approximating the behavior of CIs based on the full influence function, CIs based on $\widehat{\sigma}^2_{po}$ also achieve close to nominal coverage for $n \geq 150$ per arm when $p = 5$ but exhibits slight undercoverage for smaller n and when $p = 15$. This could be related to underestimation observed for sandwich estimators in small samples as discussed in Section 2.4 and can be expected with larger number of parameters. Still, in large samples both still show excellent performance. For all the approaches considered, there appears to be a drop in coverage for $n = 25$. CIs for studies with such small samples should be interpreted with caution. The CI based on $\widehat{\sigma}^2_{cs}$ is consistently conservative, achieving coverage of around 97% in most scenarios and exhibiting a length around 5–10% longer than CIs based on empirical estimates of the standard error.

In the Supplementary Material (Cheng, Ayyagari and Signorovitch (2020)), we also provide an extended set of simulation results on both bias and coverage for a range of data settings, including for both binary and continuous Y, to further elucidate when MAIC and its standard error estimators are or are not reliable.

4. Data application. Duchenne muscular dystrophy (DMD) is a rare neuromuscular disorder in which patients experience progressive muscle degeneration that often leads to loss of ambulation during adolescence, respiratory and cardiac dysfunction in early adulthood and, eventually, premature mortality (Emery, Muntoni and Quinlivan (2015)). Due to the rarity of the disease, recruitment of DMD patients into clinical trials can be challenging. There is strong interest in adopting single-arm designs for future DMD trials and using natural history



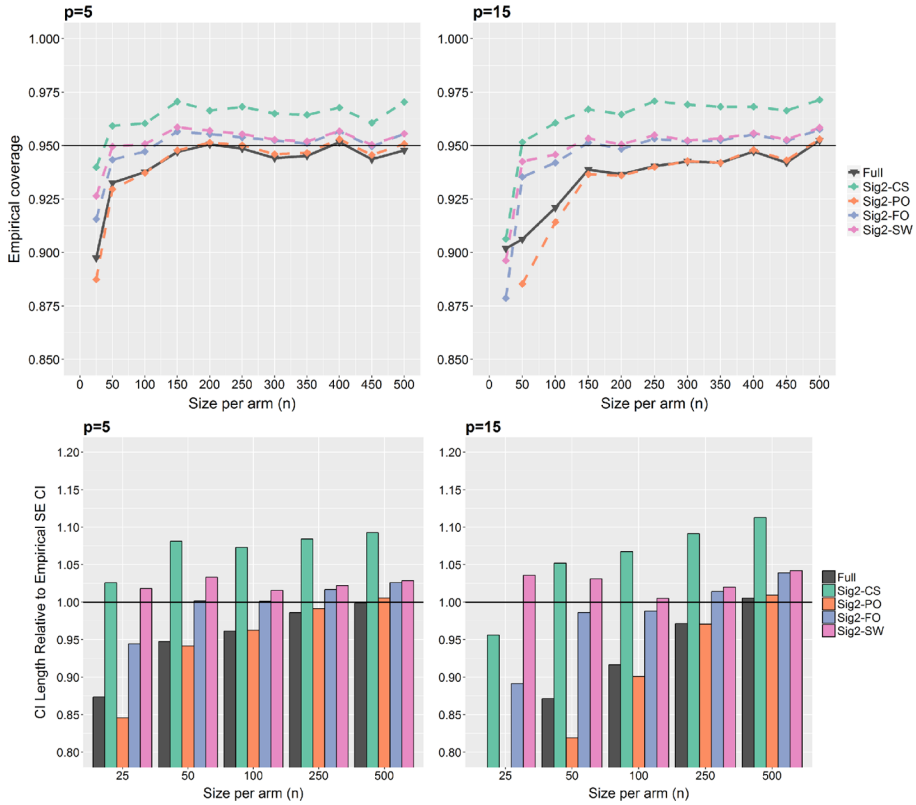
BUC: Method of Bucher et al, STC: Simulated treatment comparison, MAIC-NAB: Nonanchor-based MAIC, MAIC-ACB: Anchor-based MAIC.

FIG. 1. *Percent bias of estimators by degree of confounding*, *sample size per arm* (n) *and number of covariates* (p).

(NH) data for external controls. Such external control groups have previously informed drug approval in other rare diseases (ICH (2000), FDA (2006, 2017)). However, indirect comparisons to external controls is subject to the risk of bias due to potential differences in patient baseline characteristics, along with other factors such as those in Table 1.

To illustrate MAIC, we considered a "negative control" study to assess whether a NH cohort from a prospective noninterventional study (DMD-PRO-01), provided by CureDuchenne, a 501(3)c patient foundation, are sufficiently comparable to the placebo arm of a phase III DMD trial for tadalafil (PBO), provided by Eli Lilly, using IPD from DMD-PRO-01 and AGD from the PBO trial. If the NH setting is sufficiently comparable to the trial, mean outcomes between the studies should be similar after adjustment for differences in observed characteristics. The data were accessed through the Collaborative Trajectory Analysis Project (cTAP), a collaboration aiming to improve clinical trial design and interpretation in DMD.

We focused on a binary outcome for clinically significant worsening of the North Star Ambulatory Assessment (NSAA), defined as decrease of ≥ 3 units from baseline to week 48 after study initiation (Ricotti et al. (2016)). We implemented MAIC-NAB, in addition to STC and a naïve comparison that directly contrasted mean outcomes between the two studies without adjustment for baseline characteristics. Comparisons were conducted on the log odds-ratio scale. For instance, MAIC-NAB estimates are obtained as $\widehat{\Delta}_g = g(\widehat{\mu}_1) - g(\bar{Y}_{22})$ with $g(u) = \log\{u/(1-u)\}$. For MAIC and STC, we include baseline characteristics from Table 2 in $\mathbf{X}$. These characteristics were selected based on studies evaluating prognostic factors for ambulatory outcomes in DMD and were thought to be important confounding factors (Mazzone et al. (2016), Goemans et al. (2016)). Data on other potential confounding factors such as genetic markers, were not available from both studies, and interpretations of the results should bear in mind that other unknown or unobserved factors could still bias the results



Full: Based on full influence function, Sig2-CS: Based on $\widehat{\sigma}^2_{cs}$, Sig2-PO: Based on $\widehat{\sigma}^2_{po}$, Sig2-FO: Based on $\widehat{\sigma}^2_{fo}$, Sig2-SW: Based on $\widehat{\sigma}^2_{sw}$.

FIG. 2. *Coverage and relative length of* 95% *CI's for* $\widehat{\Delta}$ *by size per arm* (n) *and number of covariates* (p) *in the moderate confounding scenario. Some points and bars for* $n = 25$ *are omitted if they take values beyond the axis limits.* $\widehat{\sigma}^2_{fo}$ *and* $\widehat{\sigma}^2_{sw}$ *achieve close to nominal coverage for* $n \geq 50$ *and only slightly conservative in length.*

after adjustment. Estimators of the asymptotic variance on the log odds ratio scale based on strategies discussed in Section 2.4 were obtained (detailed in Appendix A). We report SEs based on $\widehat{\sigma}^2_{fo}$, which we chose since the sample size is not small. For the naïve comparison, we use the usual SE estimator for log odds ratios based on the delta method (Bland and Altman (2000)). Only the point estimate is available for STC. Two-sided p-values were also reported from Wald tests for the null hypothesis that there is no difference in the proportion of patients with NSAA worsening between studies. Comparisons were repeated with and without application of the inclusion/exclusion criteria from the tadalafil trial to DMD-PRO-01 (age 7–14 years, steroid duration $\geq$ six months, baseline six min walk distance 200–400 meters).

Table 2 reports the key characteristics available in both cohorts. Patients in the NH cohort, on average, were younger, had better ambulatory function and rise time and shorter duration of steroid treatment. The indirect comparison results are reported in Table 3. A naïve comparison of outcomes suggests that there were significant differences in NSAA worsening, with odds of worsening about 40% lower in NH vs. PBO (OR = 0.58; 95% CI: 0.34 to 0.99, $p = 0.04$). After MAIC adjustment, the magnitude of differences attenuated and was no longer statistically significant (OR = 1.14, 95% CI: 0.64 to 2.04, $p = 0.66$). The standard error increased slightly on the log odds scale from 0.27 to 0.30 after adjustment with MAIC. Results were similar for STC (OR = 1.20). Findings were also similar after applying the trial

TABLE 2
Baseline characteristics of patients from DMD-PRO-01 (NH) and tadalafil trial (PBO)

Baseline characteristics	NH ($n = 152$)	PBO ($n = 90$)
Age (years)	8.8	9.3
NSAA score	24.4	22.6
Six min walk distance (meters)	374.3	348.5
Rise time $\geq$ five seconds	50.0%	74.4%
Steroid duration $\geq$ 12 months	70%	90%
Height (cm)	122.1	125.5
Weight (kg)	28.2	30.6

NSAA: North Star Ambulatory Assessment.

inclusion/exclusion criteria. These results suggest that NH data is comparable to trial data after appropriate adjustment for baseline characteristics.

5. Discussion. As with any nonrandomized treatment comparison, indirect comparisons through MAIC can be biased by differences in unobserved confounders. Researchers ought to include covariates $\mathbf{X}$ such that Assumptions 2.4 or 2.5 are satisfied as much as possible and recognize the limitations of indirect comparison when confounders are unobserved. In the literature on estimating average treatment effects, it is well known that including covariates associated with only the outcome increases efficiency, while covariates associated with only the exposure decreases efficiency (Lunceford and Davidian (2004), Brookhart et al. (2006), Rotnitzky, Li and Li (2010), de Luna, Waernbaum and Richardson (2011)). The same results are expected to hold when estimating the weights for MAIC, viewing trial selection as the "exposure." It is thus generally advisable to include in $\mathbf{X}$ observed covariates that are known or suspected to be associated with outcomes, even if their imbalance between trials is minor (Rubin abd Thomas (1996), Lunceford and Davidian (2004), Stuart (2010)). As discussed in Section 2.2, with anchor-based MAIC there are cases where it is not necessary to adjust for covariates in $\mathbf{X}$ that are not effect modifiers. However, it is generally difficult to be certain that any covariate is not an effect modifier, especially for novel therapies. A simple objective approach to covariate selection is to include all covariates available for both the IPD and AGD that cannot be ruled out as having no associations with outcomes. Known prognostic covariates that are unobserved should be noted among the limitations. Sensitivity analyses on the impact of adding or removing covariates from $\mathbf{X}$ can also be informative.

Even when all confounders are observed, MAIC still relies on the trial assignment model from (5) being at least approximately correct for a given $\mathbf{X}$ to make appropriate adjustments.

TABLE 3
Indirect comparison of NSAA worsening between NH vs. PBO cohorts

	Method	OR	95% CI	Log OR	SE	p-value
No inclusion/exclusion criteria applied	Naive	0.58	(0.34, 0.99)	−0.54	0.27	0.04
	STC	1.20	–	0.19	–	–
	MAIC-NAB	1.14	(0.64, 2.04)	0.13	0.3	0.66
With inclusion/exclusion criteria applied	Naive	1.78	(0.95, 3.34)	0.58	0.32	0.07
	STC	1.75	–	0.56	–	–
	MAIC-NAB	1.42	(0.69, 2.92)	0.35	0.37	0.34

Naive: Unadjusted comparison, STC: Simulated treatment comparison, MAIC-NAB: Nonanchor-based MAIC.

It is important for researchers to consider its specification, such as whether higher-order polynomial or interaction terms are plausible and the corresponding requisite AGD, such as standard errors or correlations, are available. Other parametric models besides logistic regression could also be used, but the estimating equation that balances trial covariates, as in (4), must admit a unique solution for the model parameters.

Beyond unobserved confounding and model specification, Assumption 2.3 is also an important assumption. When $P(T=1|\mathbf{X}=\mathbf{x})$ is close to 0 for some $\mathbf{x}\in\mathcal{X}$, there exist some subpopulation in the AGD population that does not have strong overlap with the IPD population. This leads some observations in the IPD to receive extreme weights and dominate the reweighted sample, resulting in poor performance in point and standard error estimation. A useful diagnostic tool for lack of overlap is the effective sample size, defined by $\widetilde{N}_{1z}=\{\sum_{T_i=1,Z_i=z}\omega(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}_1)\}^2/\sum_{T_i=1,Z_i=z}\omega(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}_1)^2$ for weighted samples from arm $z=0,1$ of the IPD. $\widetilde{N}_{1z}$ is an approximation from importance sampling that downweights the sample size by the approximate relative efficiency between a sample average using data from a target distribution and a weighted average from a proposal distribution (Kong (1992)). Violations or near-violations of Assumption 2.3 would tend to inflate some $\omega(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}_1)$ and deflate $\widetilde{N}_{1z}$, signaling poor overlap. In simulations, coverage in cases when $\widetilde{N}_{11}\leq p$ ranged 68–88% for CIs based on $\widehat{\sigma}^2_{fo}$, in the small sample scenario with $n=25$, $p=5,15$, and moderate confounding. This suggests that inferences based on $\widehat{\Delta}$ are suspect in real datasets when $\widetilde{N}_{11}$ is less than or close to p. While $\widetilde{N}_{1z}$ is termed the "effective sample size" for arm $z=0,1$, it has no direct bearing on statistical inferences, including standard error and CI estimation and should be viewed as a rough indicator only. Poor performance in estimation can also result when p is large or covariates in $\mathbf{X}$ are heavily correlated which yields large standard errors for $\widehat{\boldsymbol{\alpha}}_1$. One possible remedy would be to add a regularization term to (8).

As illustrated in Section 4, it is possible to consider a nonefinitive test for the adequacy of Assumptions 2.1–2.4 and specification of trial selection model using data from a common comparator arm, if available, provided the studies are deemed to be sufficiently similar in other aspects. Let

$$\widehat{\Delta}_0=\sum_{i=1}^{N}I(Z_i=0,T_i=1)\omega(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}_1)Y_i/\sum_{i=1}^{N}I(Z_i=0,T_i=1)\omega(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}_1)-\bar{Y}_{20}$$

denote the difference in outcomes under the common comparator after weighting. Under the null that these assumptions hold, $N^{1/2}\widehat{\Delta}_0\xrightarrow{d}N(0,\sigma_0^2)$, where σ_0^2 is analogous to σ^2 for outcomes under the common comparator. Consequently, a test that rejects when $N^{1/2}|\widehat{\Delta}_0|/\widehat{\sigma}_0>z_{1-\alpha/2}$ is a level α test, where $\widehat{\sigma}_0^2$ is an estimator of σ_0^2 that can be constructed using similar strategies as in Section 2.4. The type I error may be conservative if a conservative strategy for estimating σ_0^2 is used. Rejection in such a test suggests violation of some assumption, but failure to reject does not verify the assumptions. For instance, there may be unobserved confounders that impact outcomes in the active treatment arms but not for the placebo arm.

MAIC enables estimation of causal contrasts between studies when IPD is available for one study and only AGD is available for other studies. This is a common occurrence for researchers who have access to data from their own study but only published AGD from other studies. However, we are keen to emphasize that appropriate sharing of IPD is the preferred approach, when possible, as it offers important advantages over settings where IPD is available only from some studies. With the full IPD, the pooled data can be regarded as observational data for which the wide array of methods developed for causal inference

can be applied. In particular, while MAIC only enables estimation of causal contrasts in the AGD population, having the full IPD offers additional flexibility in allowing for estimation of treatment contrasts in other target populations. This not only provides insight into treatment effect heterogeneity between populations but can also help circumvent issues with violations or near-violations of study assumptions discussed above. For example, if one study enrolled a more inclusive patient population than the other, that study can be considered as the $T = 1$ study so that it would be more plausible for Assumption 2.3 to be satisfied. Similarly, if one study collected a richer set of baseline prognostic covariates in $\mathbf{X}$, that study can be set as the $T = 1$ study to more convincingly satisfy Assumption 2.4. Access to full IPD also enables diagnostic assessments of the goodness of fit and calibration of the propensity score model which is a standard for propensity score analyses.

Another direction of future research will be to consider efficient estimation of Δ. In particular, it would be of interest to consider whether any estimator with only AGD available in one trial can still achieve the known semiparametric efficiency bound for Δ (Hahn (1998)). Recently, entropy balancing, a method that estimates causal effects via weights that minimize the relative entropy with the distribution of covariates in the control population, has been shown to be locally semiparametric efficient if the logit of the propensity scores and mean outcomes in the treated population are both linear in the covariates $\mathbf{X}$ (Zhao and Percival (2017)). Moreover, it was be also shown to be doubly robust in that it is consistent if either the logit propensity score or outcome model is linear in $\mathbf{X}$. Since entropy-balancing coincides with $\widehat{\Delta}$ when a logistic regression model is used for trial assignment, this immediately indicates that MAIC is also locally semiparametric efficient and doubly robust in the same sense. It would be of interest to consider whether doubly robust and efficient estimators are available for more general models when full IPD has been withheld for one treatment group. Developments of extensions for MAIC along these lines are underway.

APPENDIX A: INFERENCES ON NONLINEAR SCALE

When the treatment effect on a different scale of contrast is of interest, Δ_g can be estimated by first estimating $E\{Y(1)|T = 2\}$ and $E\{Y(2)|T = 2\}$ and then applying a specified $g(\cdot)$ transformation. The same considerations for identification and estimation of $E\{Y(1)|T = 2\}$ and $E\{Y(2)|T = 2\}$ applies as in Sections 2.2 and 2.3. The estimator in this case would then be

$$\widehat{\Delta}_g = g(\widehat{\mu}_1) - g(\bar{Y}_{22}), \tag{17}$$

where $\widehat{\mu}_1$ is the weighted average for treatment 1 using the IPD.

Applying the delta method, when $g(u)$ is differentiable and nonzero valued at $u = \mu_1$ and $u = \mu_2$, the influence function for $\widehat{\Delta}_g$ is given by

$$\begin{aligned} N^{1/2}(\widehat{\Delta}_g - \Delta_g) = N^{-1/2} \sum_{i=1}^{N} \psi_i^{\mu_2}(\Delta^*, \mu_1^*) + \psi_i^{\mu_1}(\mu_1^*, \boldsymbol{\alpha}_1^*) \\ + \widetilde{\psi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \mu_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + \widetilde{\psi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1^*, \mu_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1), \end{aligned}$$

where

$$\begin{aligned} &\psi_i^{\mu_2}(\Delta, \mu_1) = \varphi_i^{\mu_2}(\Delta, \mu_1) g'(\mu_1 - \Delta), \\ &\psi_i^{\mu_1}(\mu_1, \boldsymbol{\alpha}_1) = \varphi_i^{\mu_1}(\mu_1, \boldsymbol{\alpha}_1) g'(\mu_1), \\ &\widetilde{\psi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \mu_1, \boldsymbol{\mu}_{\mathbf{X}_2}) \\ &\quad = J^{\mu_1}(\boldsymbol{\alpha}_1)^{-1} \widetilde{\boldsymbol{C}}_1(\mu_1, \boldsymbol{\alpha}_1)^{\mathsf{T}} \mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})^{-1} \mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) g'(\mu_1), \end{aligned}$$

$$\widetilde{\psi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1, \mu_1, \boldsymbol{\mu}_{\mathbf{X}_2}) = J^{\mu_1}(\boldsymbol{\alpha}_1)^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1, \boldsymbol{\alpha}_1)^{\mathsf{T}}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})g'(\mu_1),$$

are modifications of the original influence function with $g'(u) = \frac{\partial}{\partial u}g(u)$. In parallel to (12), the asymptotic variance of $\widehat{\Delta}_g$ can be expressed as

$$\begin{aligned}\sigma_g^2 = &\{\mathrm{Var}(\psi_i^{\mu_1}) + \mathrm{Var}(\psi_i^{\mu_2})\} + \{\mathrm{Var}(\widetilde{\psi}_i^{\boldsymbol{\alpha}_1}) + 2\,\mathrm{Cov}(\psi_i^{\mu_1}, \widetilde{\psi}_i^{\boldsymbol{\alpha}_1})\} \\ &+ \{\mathrm{Var}(\widetilde{\psi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}) + 2\,\mathrm{Cov}(\psi_i^{\mu_2}, \widetilde{\psi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}})\},\end{aligned} \tag{18}$$

where the arguments of the components of the influence function are suppressed but implicitly evaluated at their respective truth. Based on similar considerations, the three proposed estimators for σ_g^2 are

$$\begin{aligned}\widehat{\sigma}_{g,fo}^2 &= \widehat{\mathrm{Var}}\{\psi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1)\} + \widehat{\mathrm{Var}}\{\psi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}, \\ \widehat{\sigma}_{g,po}^2 &= \widehat{\mathrm{Var}}\{\psi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1)\} + \widehat{\mathrm{Var}}\{\widetilde{\psi}_i^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\mu}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2})\} + \widehat{\mathrm{Var}}\{\psi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}, \\ \widehat{\sigma}_{g,cs}^2 &= \widehat{\mathrm{Var}}\{\psi_i^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1) + \widetilde{\psi}_i^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\mu}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2})\} + \widehat{\mathrm{Var}}\{\psi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}, \\ &\quad + \widehat{V}^{\boldsymbol{\mu}_{\mathbf{X}_2}}g'(\widehat{\mu}_1)^2 + 2\big[\widehat{\mathrm{Var}}\{\psi_i^{\mu_2}(\widehat{\Delta}, \widehat{\mu}_1)\}\widehat{V}^{\boldsymbol{\mu}_{\mathbf{X}_2}}g'(\widehat{\mu}_1)^2\big]^{1/2}.\end{aligned} \tag{19}$$

As in Section 2.4, $\widehat{\sigma}_{g,fo}^2$ tends to outperform $\widehat{\sigma}_{g,po}^2$ in small samples, whereas $\widehat{\sigma}_{g,cs}^2$ provides a conservative estimate.

APPENDIX B: SIMULATION RESULTS ON LOGIT SCALE

We repeated the simulations to assess the bias of the corresponding estimators modified to estimate the treatment effect on the logit scale, that is, Δ_g with $g(u) = \log\{u/(1-u)\}$. This has previously been the recommended scale for binary outcomes HTA applications (Phillippo et al. (2016)). The performance of CIs for MAIC-NAB based on the modified estimators of the asymptotic variance were also assessed. The same simulation settings as those in Section 3 were used throughout. MAIC-ACB was calculated based on $\widehat{\Delta}_g^{\mathrm{ACB}} = \widehat{\Delta}_g - g\{\sum_{i=1}^N I(Z_i = 0, T_i = 1)\omega(\mathbf{X}_i; \widehat{\boldsymbol{\alpha}}_1)Y_i / \sum_{i=1}^N I(Z_i = 0, T_i = 1)\omega(\mathbf{X}_i; \widehat{\boldsymbol{\alpha}}_1)\} + g(\bar{Y}_{20})$.

The bias results are reported in Table 4. Similar patterns of bias occur as when estimating Δ. BUC and STC incur substantial bias in scenarios with confounding, whereas MAIC has negligible bias except with extremely low sample size case with $n = 25$ per arm where no method is reliable. The CI performance results are presented in Figure 3. CIs based on $\widehat{\sigma}_{g,fo}^2$ and $\widehat{\sigma}_{g,sw}^2$ achieve near nominal coverage when $n \geq 50$ per arm and their lengths do not surpass those based on the empirical estimate by more than 5% for most n. CIs based on $\widehat{\sigma}_{g,po}^2$ exhibit good coverage when sample size is large relative to p but tend to undercover when n is relatively small. CIs based on $\widehat{\sigma}_{g,cs}^2$ are the most conservative.

APPENDIX C: ALTERNATIVE SAMPLING SCHEMES

When estimating the standard error of $\widehat{\Delta}$, a way to avoid the issue of the lack of full IPD is to consider $\boldsymbol{\mu}_{\mathbf{X}_2}^*$ and μ_2^* as fixed parameters. That is, one may consider $\boldsymbol{\mu}_{\mathbf{X}_2}^* = \bar{\mathbf{X}}_2$ or $\mu_2^* = \bar{Y}_{22}$. Under either of these assumptions, the asymptotic variance σ^2 would exclude terms involving either $\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}$ or $\varphi_i^{\mu_2}$, respectively. Inference based on these assumptions clearly then would not acknowledge sampling variability from estimating these parameters which may or may not be justified depending on the context of the problem. In this article we primarily focus on the more difficult case where estimates from the AGD are considered to be random.

TABLE 4

Percent bias of estimators of Δ_g by degree of confounding, total size (N) and number of covariates (p)

		No confounding			Moderate confounding			Severe confounding		
Size	Estimator	$p=5$	$p=10$	$p=15$	$p=5$	$p=10$	$p=15$	$p=5$	$p=10$	$p=15$
$n=25$	BUC	5%	3%	2%	36%	32%	41%	50%	45%	53%
	STC	1315%	11,064%	3752%	1816%	9903%	3392%	2498%	9066%	2951%
	MAIC-NAB	14%	27%	365%	20%	24%	322%	23%	41%	362%
	MAIC-ACB	0%	−7%	−122%	4%	−11%	−44%	−3%	2%	−173%
$n=50$	BUC	3%	2%	0%	28%	32%	33%	44%	42%	47%
	STC	31%	146%	10,724%	42%	346%	11,908%	53%	702%	11,294%
	MAIC-NAB	4%	8%	6%	7%	9%	15%	7%	10%	10%
	MAIC-ACB	3%	0%	−4%	0%	4%	7%	−1%	−2%	0%
$n=100$	BUC	3%	2%	−1%	30%	29%	28%	44%	46%	46%
	STC	14%	28%	66%	23%	38%	101%	33%	45%	209%
	MAIC-NAB	3%	3%	2%	3%	4%	4%	6%	3%	6%
	MAIC-ACB	3%	1%	−1%	2%	2%	1%	1%	2%	2%
$n=250$	BUC	1%	2%	1%	29%	29%	27%	42%	44%	46%
	STC	6%	10%	21%	17%	21%	29%	23%	29%	39%
	MAIC-NAB	2%	2%	2%	2%	2%	1%	0%	2%	2%
	MAIC-ACB	1%	2%	0%	1%	2%	0%	−2%	1%	2%
$n=500$	BUC	0%	0%	1%	28%	27%	28%	41%	43%	44%
	STC	3%	5%	9%	14%	16%	21%	21%	25%	29%
	MAIC-NAB	1%	1%	1%	1%	0%	1%	−1%	1%	2%
	MAIC-ACB	0%	0%	1%	0%	−1%	1%	−2%	0%	1%

BUC: Method of Bucher et al. STC: Simulated treatment comparison, MAIC-NAB: Nonanchor-based MAIC, MAIC-ACB: Anchor-based MAIC.

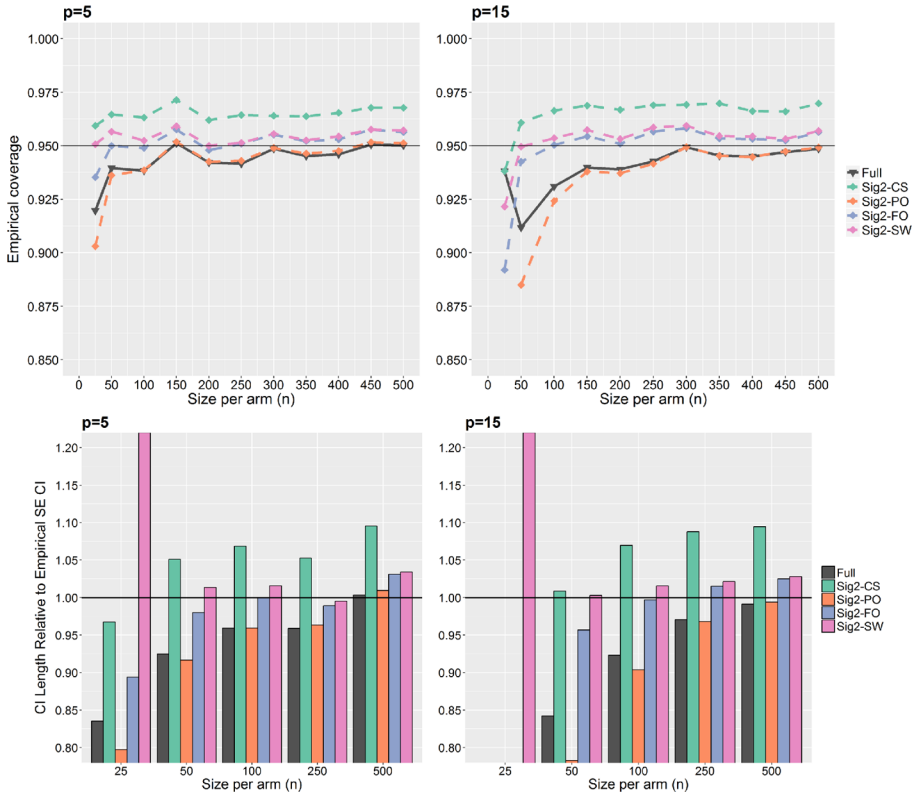
A related issue regarding sampling is that the allocation of patients among trials, in practice, may be constrained such that each trial and arm enrolls a fixed, or nearly fixed, number of patients. While formally this sampling scheme differs from the problem setup, which does not assume any sample size constraints, it can be accommodated as a simple extension that does not impact the asymptotic analysis. In particular, one could consider patients to be subsampled by trial and arm to meet size constraints after initially being sampled without constraint from the super-population. Under the assumptions on the initial sample required when sampling without constraint, the proposed procedures for inference about Δ based on $\widehat{\Delta}$ would still be valid using the subsampled data as long as the subsampling was random by trial and arm.

More concretely, suppose N^* total patients were initially sampled without constraint. Let $R_i \in \{0, 1\}$ for $i = 1, \ldots, N^*$ be an indicator of whether each patient is subsampled into the final sample. If a fixed n_{tz} number of patients were enrolled into arm z of trial t, the constraint on the subsampling is that $\sum_{i=1}^{N^*} R_i I(T_i = t, Z_i = z) = n_{tz}$, for $t = 1, 2$ and $z = 0, 1, 2$. If subsampling was random by trial and arm, then

$$R \perp\!\!\!\perp \{\mathbf{X}, Y(0), Y(1), Y(2)\} | T, Z. \tag{20}$$

When Assumption 2.1 holds in the initial sample, then we also have that $R \perp\!\!\!\perp \{\mathbf{X}, Y(0), Y(1), Y(2)\}|T$. But under this independence,

$$E\{Y(2)|T=2\} = E\{Y(2)|T=2, R=1\} \quad \text{and}$$
$$E\{Y(1)|T=2\} = E\{Y(1)|T=2, R=1\}.$$



Full: Based on full influence function, Sig2-CS: Based on $\widehat{\sigma}^2_{g,cs}$, Sig2-PO: Based on $\widehat{\sigma}^2_{g,po}$, Sig2-FO: Based on $\widehat{\sigma}^2_{g,fo}$, Sig2-SW: Based on $\widehat{\sigma}^2_{g,sw}$.

FIG. 3. *Coverage and relative length of* 95% *CI's for* $\widehat{\Delta}$ *by size per arm* (n) *and number of covariates* (p) *in the moderate confounding scenario. Some points and bars for* $n = 25$ *are omitted if they take values beyond the axis limits.*

Moreover, Assumptions 2.1–2.4 can be shown to still hold conditional on $R = 1$, provided they hold in the initial sample. The same arguments to identify Δ from Section 2.2 thus hold under distributions that are conditional on $R = 1$. $\widehat{\Delta}$ will then be consistent for Δ when the subsampled data is used, if $\omega(\mathbf{X})$ among the subsampled data can be identified and estimated. But if trial assignment among the initial sample follows a logistic regression model as in (5), then

$$\text{logit}\, P(T = 2|\mathbf{X}, R = 1) = \boldsymbol{\alpha}_T^{\mathsf{T}} \overrightarrow{\mathbf{X}},$$

where $\boldsymbol{\alpha}_T = (\alpha_{0,T}, \boldsymbol{\alpha}_1)$. That is, the trial assignment among the subsample *still* follows a logistic regression model with a possibly different intercept $\alpha_{0,T}$. This is analogous to the result that the probability of an outcome given covariates among those sampled in a case-control study still follows the same logistic regression with a different intercept when the prospective model is logistic regression (Prentice and Pyke (1979)). The arguments for identification of $\omega(\mathbf{X})$ in Section 2.2 are thus also still valid, and estimation can proceed from solving (8) among the subsampled data.

APPENDIX D: PROOFS OF THEOREMS

D.1. Proof of Theorem 2.1. Let the estimating equation from (8) be denoted

$$\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) = N^{-1} \sum_{i=1}^{N} \exp\{\boldsymbol{\alpha}_1^{\mathsf{T}}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2}) I(T_i = 1). \tag{21}$$

We will first outline the consistency of $\widehat{\boldsymbol{\alpha}}_1$ for $\boldsymbol{\alpha}_1^*$. It can be shown that there exists a $C_N = O_p(1)$ such that

$$\begin{aligned}\sup_{\boldsymbol{\alpha}_1}\|\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}) - \mathbf{U}_N^{\widetilde{\boldsymbol{\alpha}}_1,1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\| &\le C_N\|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*\| \\ &= o_p(1),\end{aligned} \tag{22}$$

by using that $\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})$ is continuously differentiable in $\boldsymbol{\mu}_{\mathbf{X}_2}$, the parameter space for $(\boldsymbol{\alpha}_1^\mathsf{T}, \boldsymbol{\mu}_{\mathbf{X}_2}^\mathsf{T})^\mathsf{T}$, denoted by Θ, is compact and $\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}$ is a consistent estimator for $\boldsymbol{\mu}_{\mathbf{X}_2}^*$. Let $\Theta_{\boldsymbol{\alpha}_1}$ be the parameter space for $\boldsymbol{\alpha}_1$. Moreover, it can be shown that there exists a $D_N = O_p(1)$ such that, for any $\widetilde{\widetilde{\boldsymbol{\alpha}}}_1, \widetilde{\boldsymbol{\alpha}}_1 \in \Theta_{\boldsymbol{\alpha}_1}$,

$$\|\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\widetilde{\widetilde{\boldsymbol{\alpha}}}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) - \mathbf{U}_N^{\boldsymbol{\alpha}_1}(\widetilde{\boldsymbol{\alpha}}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\| \le D_N\|\widetilde{\widetilde{\boldsymbol{\alpha}}}_1 - \widetilde{\boldsymbol{\alpha}}_1\|, \tag{23}$$

using that $\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)$ is continuously differentiable in $\boldsymbol{\alpha}_1$ and $\Theta_{\boldsymbol{\alpha}_1}$ is compact. Let $\mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}) = E[\exp\{\boldsymbol{\alpha}_1^\mathsf{T}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})I(T_i = 1)]$. Since $\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) \xrightarrow{p} \mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)$ pointwise for each $\boldsymbol{\alpha}_1 \in \Theta_{\boldsymbol{\alpha}_1}$, we have

$$\sup_{\boldsymbol{\alpha}_1}\|\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) - \mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\| = o_p(1) \tag{24}$$

using Lemma 2.9 of Newey and McFadden (1994). This verifies that

$$\begin{aligned}&\sup_{\boldsymbol{\alpha}_1}\|\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}) - \mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\| \\ &\quad\le \sup_{\boldsymbol{\alpha}_1}\|\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}) - \mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\| + \sup_{\boldsymbol{\alpha}_1}\|\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) - \mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\| \\ &\quad= o_p(1).\end{aligned} \tag{25}$$

Now, since $E[\exp\{\boldsymbol{\alpha}_1^\mathsf{T}(\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2})\}I(T_i = 1)]$ is strictly convex in $\boldsymbol{\alpha}_1$, $\mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})$ has a unique solution in $\widetilde{\boldsymbol{\alpha}}_1$ for a given $\boldsymbol{\mu}_{\mathbf{X}_2}$. Hence, $\widehat{\boldsymbol{\alpha}}_1 \xrightarrow{p} \boldsymbol{\alpha}_1^*$ by Theorem 5.9 of van der Vaart (1998). We now turn to obtaining the influence function for $\widehat{\boldsymbol{\alpha}}_1$. Let $U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})$ denote the jth component of $\mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})$, for $j = 1, \ldots, p$. An expansion of $U_{N,j}^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2})$ around $(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)$ yields

$$\begin{aligned}&U_{N,j}^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}) \\ &\quad= U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + \frac{\partial}{\partial\boldsymbol{\alpha}_1^\mathsf{T}}U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*) \\ &\qquad+ \frac{\partial}{\partial\boldsymbol{\mu}_{\mathbf{X}_2}^\mathsf{T}}U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)(\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*) + \frac{\partial^2}{\partial\boldsymbol{\alpha}_1^{\otimes 2}}U_{N,j}^{\boldsymbol{\alpha}_1}(\widetilde{\boldsymbol{\alpha}}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*)^{\otimes 2} \\ &\qquad+ \frac{\partial^2}{\partial\boldsymbol{\mu}_{\mathbf{X}_2}^{\otimes 2}}U_{N,j}^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widetilde{\boldsymbol{\mu}}_{\mathbf{X}_2})(\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*)^{\otimes 2} \\ &\qquad+ \frac{\partial^2}{\partial\boldsymbol{\alpha}_1\partial\boldsymbol{\mu}_{\mathbf{X}_2}}U_{N,j}^{\boldsymbol{\alpha}_1}(\widetilde{\widetilde{\boldsymbol{\alpha}}}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*)(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*)(\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*),\end{aligned} \tag{26}$$

where $\widetilde{\boldsymbol{\alpha}}_1$ and $\widetilde{\widetilde{\boldsymbol{\alpha}}}_1$ are intermediates on the line segment between $\widehat{\boldsymbol{\alpha}}_1$ and $\boldsymbol{\alpha}_1^*$, and $\widetilde{\boldsymbol{\mu}}_{\mathbf{X}_2}$ is an intermediate between $\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}$ and $\boldsymbol{\mu}_{\mathbf{X}_2}^*$. Using that $U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2})$ is continuously differentiable

in $(\boldsymbol{\alpha}_1^{\mathsf{T}}, \boldsymbol{\mu}_{\mathbf{X}_2}^{\mathsf{T}})^{\mathsf{T}}$ and Θ is compact,

$$\begin{aligned}&\frac{\partial^2}{\partial \boldsymbol{\alpha}_1^{\otimes 2}} U_{N,j}^{\boldsymbol{\alpha}_1}(\widetilde{\boldsymbol{\alpha}}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) = O_p(1), \frac{\partial^2}{\partial \boldsymbol{\mu}_{\mathbf{X}_2}^{\otimes 2}} U_{N,j}^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widetilde{\boldsymbol{\mu}}_{\mathbf{X}_2}) = O_p(1),\\ &\frac{\partial^2}{\partial \boldsymbol{\alpha}_1 \partial \boldsymbol{\mu}_{\mathbf{X}_2}} U_{N,j}^{\boldsymbol{\alpha}_1}(\widetilde{\widetilde{\boldsymbol{\alpha}}}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) = O_p(1).\end{aligned} \tag{27}$$

Moreover, we have

$$\begin{aligned}&\frac{\partial}{\partial \boldsymbol{\alpha}_1^{\mathsf{T}}} U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\\ &\quad = E\big[\exp\{\boldsymbol{\alpha}_1^{*\mathsf{T}}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}_2}^*)\}(X_j - \mu_{\mathbf{X}_2,j}^*)(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}_2}^*)^{\mathsf{T}} I(T=1)\big] + o_p(1)\\ &\quad = -J_j^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1),\\ &\frac{\partial}{\partial \boldsymbol{\mu}_{\mathbf{X}_2}^{\mathsf{T}}} U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) = -E\big[\exp\{\boldsymbol{\alpha}_1^{*\mathsf{T}}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}_2}^*)\} I(T=1)\mathbf{1}_j^{\mathsf{T}}\big] + o_p(1),\end{aligned} \tag{28}$$

where X_j and $\mu_{\mathbf{X}_2,j}$ denote the jth element of $\mathbf{X}$ and $\boldsymbol{\mu}_{\mathbf{X}_2}$ and $\mathbf{1}_j$ denotes a $p \times 1$ vector that is 1 in the jth position and 0 in the other positions. We also used that $\boldsymbol{\alpha}_1^*$ is the solution to $\mathbf{U}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1, \boldsymbol{\mu}_{\mathbf{X}_2}^*) = \mathbf{0}$.

Now, for each $j = 1, \ldots, p$, since $U_{N,j}^{\boldsymbol{\alpha}_1}(\widehat{\boldsymbol{\alpha}}_1, \widehat{\boldsymbol{\mu}}_{\mathbf{X}_2}) = 0$, rearranging from above yields

$$\begin{aligned}&\big\{-J_j^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1)\big\} N^{1/2}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*)\\ &\quad = -N^{1/2} U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\\ &\qquad - \frac{\partial}{\partial \boldsymbol{\mu}_{\mathbf{X}_2}^{\mathsf{T}}} U_{N,j}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) N^{1/2}(\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1),\end{aligned} \tag{29}$$

where we use that $\widehat{\mu}_{\mathbf{X}_2}$ is $N^{1/2}$-consistent. Considering the p components simultaneously yields

$$\begin{aligned}&\big\{-\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1)\big\} N^{1/2}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*)\\ &\quad = -N^{1/2} \mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\\ &\qquad + E\big[\exp\{\boldsymbol{\alpha}_1^{*\mathsf{T}}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}_2}^*)\} I(T=1)\mathbf{1}_{p\times p}^{\mathsf{T}}\big] N^{1/2}(\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1).\end{aligned} \tag{30}$$

Since $\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} = \{\sum_{i=1}^N \mathbf{X}_i I(T_i = 2)\}/\{\sum_{i=1}^N I(T_i = 2)\}$, its influence function is given by

$$N^{1/2}(\widehat{\boldsymbol{\mu}}_{\mathbf{X}_2} - \boldsymbol{\mu}_{\mathbf{X}_2}^*) = N^{-1/2} \sum_{i=1}^N (\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2}^*) \frac{I(T_i = 2)}{P(T_i = 2)} + o_p(1). \tag{31}$$

Finally, since $\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)$ is nonsingular,

$$\begin{aligned}&N^{1/2}(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*)\\ &\quad = \mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)^{-1} \Bigg(N^{1/2} \mathbf{U}_N^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\\ &\qquad - E\big[\exp\{\boldsymbol{\alpha}_1^{*\mathsf{T}}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}_2}^*)\} I(T=1)\big] N^{-1/2} \sum_{i=1}^N (\mathbf{X}_i - \boldsymbol{\mu}_{\mathbf{X}_2}^*) \frac{I(T_i = 2)}{P(T_i = 2)} \Bigg) + o_p(1)\end{aligned} \tag{32}$$

$$= N^{-1/2} \sum_{i=1}^{N} \mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)^{-1} \{\mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + \mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*)\} + o_p(1)$$

$$= N^{-1/2} \sum_{i=1}^{N} \boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*, \boldsymbol{\mu}_{\mathbf{X}_2}^*) + o_p(1).$$

D.2. Proof of Theorem 2.2. Let the estimating equation associated with $\widehat{\Delta}$ be

$$U_N^{\Delta}(\Delta, \mu_1) = N^{-1} \sum_{i=1}^{N} (\mu_1 - Y_i - \Delta) I(T_i = 2, Z_i = 2). \tag{33}$$

We will directly identify the influence function expansion for $\widehat{\Delta}$, which implies that $\widehat{\Delta} \xrightarrow{p} \Delta^*$, when the identification assumptions hold and the trial assignment is correctly specified. First, we decompose $U_N^{\Delta}(\widehat{\Delta}, \widehat{\mu}_1)$:

$$\begin{aligned} U_N^{\Delta}(\widehat{\Delta}, \widehat{\mu}_1) &= U_N^{\Delta}(\Delta^*, \mu_1^*) + U_N^{\Delta}(\widehat{\Delta}, \mu_1^*) - U_N^{\Delta}(\Delta^*, \mu_1^*) + U_N^{\Delta}(\widehat{\Delta}, \widehat{\mu}_1) - U_N^{\Delta}(\widehat{\Delta}, \mu_1^*) \\ &= U_N^{\Delta}(\Delta^*, \mu_1^*) + N^{-1} \sum_{i=1}^{N} -I(T_i = 2, Z_i = 2)(\widehat{\Delta} - \Delta^*) \\ &\quad + N^{-1} \sum_{i=1}^{N} I(T_i = 2, Z_i = 2)(\widehat{\mu}_1 - \mu_1^*). \end{aligned} \tag{34}$$

Since $U_N^{\Delta}(\widehat{\Delta}, \widehat{\mu}_1) = 0$, rearranging from above,

$$\begin{aligned} N^{1/2}(\widehat{\Delta} - \Delta^*) &= N^{1/2} \left\{ U_N^{\Delta}(\Delta^*, \mu_1^*) + N^{-1} \sum_{i=1}^{N} I(T_i = 2, Z_i = 2)(\widehat{\mu}_1 - \mu_1^*) \right\} \\ &\quad \Big/ \left\{ N^{-1} \sum_{i=1}^{N} I(T_i = 2, Z_i = 2) \right\} \\ &= N^{-1/2} \sum_{i=1}^{N} (\mu_1^* - Y_i - \Delta^*) \frac{I(Z_i = 2, T_i = 2)}{P(Z_i = 2, T_i = 2)} \\ &\quad + N^{1/2}(\widehat{\mu}_1 - \mu_1) + o_p(1). \end{aligned} \tag{35}$$

Now, let the estimating equation for $\widehat{\mu}_1$ be denoted

$$U_N^{\mu_1}(\mu_1, \boldsymbol{\alpha}_1) = N^{-1} \sum_{i=1}^{N} (Y_i - \mu_1) \exp(\boldsymbol{\alpha}_1^{\mathsf{T}} \mathbf{X}_i) I(T_i = 1, Z_i = 1). \tag{36}$$

We can decompose $U_N^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1)$ as

$$\begin{aligned} &U_N^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1) \\ &\quad = U_N^{\mu_1}(\mu_1^*, \boldsymbol{\alpha}_1^*) + U_N^{\mu_1}(\widehat{\mu}_1, \boldsymbol{\alpha}_1^*) - U_N^{\mu_1}(\mu_1^*, \boldsymbol{\alpha}_1^*) + U_N^{\mu_1}(\widehat{\mu}_1, \widehat{\boldsymbol{\alpha}}_1) - U_N^{\mu_1}(\widehat{\mu}_1, \boldsymbol{\alpha}_1^*) \\ &\quad = U_N^{\mu_1}(\mu_1^*, \boldsymbol{\alpha}_1^*) - N^{-1} \sum_{i=1}^{N} \exp(\boldsymbol{\alpha}_1^{*\mathsf{T}} \mathbf{X}_i) I(T_i = 1, Z_i = 1)(\widehat{\mu}_1 - \mu_1^*) \\ &\qquad + \frac{\partial}{\partial \boldsymbol{\alpha}_1^{\mathsf{T}}} U_N^{\mu_1}(\mu_1^*, \boldsymbol{\alpha}_1^*)(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*) + \frac{\partial^2}{\partial \boldsymbol{\alpha}_1^{\otimes 2}} U_N^{\mu_1}(\widehat{\mu}_1, \widetilde{\boldsymbol{\alpha}}_1)(\widehat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_1^*)^{\otimes 2} \end{aligned} \tag{37}$$

$$+\frac{\partial}{\partial\mu_1}\frac{\partial}{\partial\boldsymbol{\alpha}_1^{\mathsf{T}}}U_N^{\mu_1}(\widetilde{\mu}_1,\boldsymbol{\alpha}_1^*)(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1^*)(\widehat{\mu}_1-\mu_1),$$

where $\widetilde{\boldsymbol{\alpha}}_1$ and $\widetilde{\mu}_1$ are such that $\|\widetilde{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1^*\|\le\|\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1^*\|$ and $\|\widetilde{\mu}_1-\mu_1^*\|\le\|\widehat{\mu}_1-\mu_1^*\|$. Now, using that $U_N^{\mu_1}(\mu_1,\boldsymbol{\alpha}_1)$ is twice continuously differentaible and Θ is compact, it can be shown that

$$\frac{\partial^2}{\partial\boldsymbol{\alpha}_1^{\otimes 2}}U_N^{\mu_1}(\widehat{\mu}_1,\widetilde{\boldsymbol{\alpha}}_1)=O_p(1)\quad\text{and}\quad\frac{\partial}{\partial\mu_1}\frac{\partial}{\partial\boldsymbol{\alpha}_1^{\mathsf{T}}}U_N^{\mu_1}(\widetilde{\mu}_1,\boldsymbol{\alpha}_1^*)=O_p(1). \tag{38}$$

Since $U_N^{\mu_1}(\widehat{\mu}_1,\widehat{\boldsymbol{\alpha}}_1)=0$, rearranging yields

$$\begin{aligned}&\Bigg\{N^{-1}\sum_{i=1}^N\exp(\boldsymbol{\alpha}_1^{*\mathsf{T}}\mathbf{X}_i)I(T_i=1,Z_i=1)+o_p(1)\Bigg\}N^{1/2}(\widehat{\mu}_1-\mu_1^*)\\&\quad=N^{1/2}\Bigg\{U_N^{\mu_1}(\mu_1^*,\boldsymbol{\alpha}_1^*)+\frac{\partial}{\partial\boldsymbol{\alpha}_1^{\mathsf{T}}}U_N^{\mu_1}(\mu_1^*,\boldsymbol{\alpha}_1^*)(\widehat{\boldsymbol{\alpha}}_1-\boldsymbol{\alpha}_1^*)\Bigg\}+o_p(1).\end{aligned} \tag{39}$$

Now, substituting the influence function expansion from Theorem 2.1 yields that

$$\begin{aligned}N^{1/2}(\widehat{\mu}_1-\mu_1^*)&=N^{-1/2}\sum_{i=1}^N J^{\mu_1}(\boldsymbol{\alpha}_1^*)^{-1}(Y_i-\mu_1^*)\exp(\boldsymbol{\alpha}_1^{*\mathsf{T}}\mathbf{X}_i)I(T_i=1,Z_i=1)\\&\quad+J^{\mu_1}(\boldsymbol{\alpha}_1^*)^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1^*,\boldsymbol{\alpha}_1^*)^{\mathsf{T}}\boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*,\boldsymbol{\mu}_{\mathbf{X}_2}^*)+o_p(1).\end{aligned} \tag{40}$$

Returning to (39), we conclude that

$$\begin{aligned}N^{1/2}(\widehat{\Delta}-\Delta^*)&=N^{-1/2}\sum_{i=1}^N\varphi_i^{\mu_2}(\Delta^*,\mu_1^*)+\varphi_i^{\mu_1}(\mu_1^*,\boldsymbol{\alpha}_1^*)\\&\quad+J^{\mu_1}(\boldsymbol{\alpha}_1^*)^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1^*,\boldsymbol{\alpha}_1^*)^{\mathsf{T}}\boldsymbol{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1^*,\boldsymbol{\mu}_{\mathbf{X}_2}^*)+o_p(1)\\&=N^{-1/2}\sum_{i=1}^N\varphi_i(\Delta^*,\mu_1^*,\boldsymbol{\alpha}_1^*,\boldsymbol{\mu}_{\mathbf{X}_2}^*)+o_p(1).\end{aligned} \tag{41}$$

D.3. Proof of Lemma 2.1. From direct calculation,

$$\begin{aligned}&\operatorname{Var}\{\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\mu_1,\boldsymbol{\mu}_{\mathbf{X}_2})\}\\&\quad=J^{\mu_1}(\boldsymbol{\alpha}_1)^{-2}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\\&\qquad\times\operatorname{Var}\{\mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})\}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)\\&\quad=J^{\mu_1}(\boldsymbol{\alpha}_1)^{-2}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)^{\mathsf{T}}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\\&\qquad\times E\{(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})^{\mathsf{T}}e^{2\boldsymbol{\alpha}_1^{\mathsf{T}}(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})}I(T_i=1)\}\\&\qquad\times\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1),\\&\operatorname{Cov}\{\varphi_i^{\mu_1}(\mu_1,\boldsymbol{\alpha}_1),\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\mu_1,\boldsymbol{\mu}_{\mathbf{X}_2})\}\\&\quad=J^{\mu_1}(\boldsymbol{\alpha}_1)^{-2}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)^{\mathsf{T}}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\\&\qquad\times E\{\mathbf{U}_i^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})(Y_i-\mu_1)e^{\boldsymbol{\alpha}_1^{\mathsf{T}}\mathbf{X}_i}I(Z_i=1,T_i=1)\}\\&\quad=J^{\mu_1}(\boldsymbol{\alpha}_1)^{-2}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)^{\mathsf{T}}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}e^{\boldsymbol{\alpha}_1^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{X}_2}}\end{aligned} \tag{42}$$

$$
\begin{aligned}
&\times E\{(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})(Y_i-\mu_1)e^{2\boldsymbol{\alpha}_1^\top(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})}I(Z_i=1,T_i=1)\},\\
&\operatorname{Var}\{\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\mu}_{\mathbf{X}_2},\mu_1,\boldsymbol{\alpha}_1)\}\\
&\quad= J^{\mu_1}(\boldsymbol{\alpha}_1)^{-2}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\\
&\qquad\times\operatorname{Var}\{\mathbf{U}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})\}\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)\\
&\quad=\frac{E[e^{\boldsymbol{\alpha}_1^\top(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})}I(T_i=1)]^2}{J^{\mu_1}(\boldsymbol{\alpha}_1)^2P(T_i=2)}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)^\top\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\\
&\qquad\times\operatorname{Var}(\mathbf{X}_i|T_i=2)\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1),\\
&\operatorname{Cov}\{\varphi^{\mu_2}(\Delta,\mu_1),\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}}(\boldsymbol{\alpha}_1,\mu_1,\boldsymbol{\mu}_{\mathbf{X}_2})\}\\
&\quad=\frac{E[e^{\boldsymbol{\alpha}_1^\top(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})}I(T_i=1)]}{J^{\mu_1}(\boldsymbol{\alpha}_1)P(T_i=2)}\\
&\qquad\times\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)^\top\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})^{-1}\widetilde{\boldsymbol{C}}_2.
\end{aligned}
$$

In the case where (2.1), (2.2), (2.4) and (2.3) hold and (5) is correctly specified, note that

$$
\begin{aligned}
\widetilde{\boldsymbol{C}}_1(\mu_1,\boldsymbol{\alpha}_1)&=E\{\mathbf{X}_i(Y_i-\mu_1)\omega(\mathbf{X}_i)e^{-\alpha_0}I(Z_i=1,T_i=1)\}\\
&=E\{\mathbf{X}_i\big(Y_i(1)-\mu_1\big)P(T_i=2|\mathbf{X}_i)\}e^{-\alpha_0}P(Z_i=1|T_i=1)\\
&=\boldsymbol{C}_1e^{-\alpha_0}P(Z_i=1|T_i=1)P(T_i=2),\\
\mathbf{J}^{\boldsymbol{\alpha}_1}(\boldsymbol{\alpha}_1,\boldsymbol{\mu}_{\mathbf{X}_2})&=-E\big[(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})^\top\omega(\mathbf{X}_i)e^{-\alpha_0-\boldsymbol{\alpha}_1^\top\boldsymbol{\mu}_{\mathbf{X}_2}}I(T_i=1)\big]\\
&=-E\big[(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})^\top P(T_i=2|\mathbf{X}_i)\big]e^{-\alpha_0-\boldsymbol{\alpha}_1^\top\boldsymbol{\mu}_{\mathbf{X}_2}}\\
&=-\operatorname{Var}(\mathbf{X}_i|T_i=2)P(T_i=2)e^{-\alpha_0-\boldsymbol{\alpha}_1^\top\boldsymbol{\mu}_{\mathbf{X}_2}},
\end{aligned}
\tag{43}
$$

and

$$
\begin{aligned}
E\big[e^{\boldsymbol{\alpha}_1^\top(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2})}I(T_i=1)\big]&=E\big[\omega(\mathbf{X}_i)e^{-\alpha_0-\boldsymbol{\alpha}_1^\top\boldsymbol{\mu}_{\mathbf{X}_2}}I(T_i=1)\big]\\
&=P(T_i=2)e^{-\alpha_0-\boldsymbol{\alpha}_1^\top\boldsymbol{\mu}_{\mathbf{X}_2}},\\
J^{\mu_1}(\boldsymbol{\alpha}_1)&=E\{\omega(\mathbf{X}_i)I(Z_i=1,T_i=1)\}e^{-\alpha_0}\\
&=P(T_i=2)P(Z_i=1|T_i=1)e^{-\alpha_0}.
\end{aligned}
\tag{44}
$$

Consequently, under these assumptions after evaluating parameters at the truth and simplification,

$$
\begin{aligned}
\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1})&=P(T_i=2)^{-1}\boldsymbol{C}_1^\top\operatorname{Var}(\mathbf{X}_i|T_i=2)^{-1}\\
&\quad\times E\{(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2}^*)(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2}^*)^\top\omega(\mathbf{X}_i)|T_i=2\}\operatorname{Var}(\mathbf{X}_i|T_i=2)^{-1}\boldsymbol{C}_1,\\
\operatorname{Cov}(\varphi_i^{\mu_1},\widetilde{\varphi}_i^{\boldsymbol{\alpha}_1})&=-P(T_i=2)^{-1}\boldsymbol{C}_1\operatorname{Var}(\mathbf{X}_i|T_i=2)^{-1}\\
&\quad\times E\{(\mathbf{X}_i-\boldsymbol{\mu}_{\mathbf{X}_2}^*)(Y_i(1)-\mu_1^*)\omega(\mathbf{X}_i)|T_i=2\},\\
\operatorname{Var}(\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}})&=P(T_i=2)^{-1}\boldsymbol{C}_1^\top\operatorname{Var}(\mathbf{X}_i|T_i=2)^{-1}\boldsymbol{C}_1,\\
\operatorname{Cov}(\varphi_i^{\mu_2},\widetilde{\varphi}_i^{\boldsymbol{\mu}_{\mathbf{X}_2}})&=-P(T_i=2)^{-1}\boldsymbol{C}_1^\top\operatorname{Var}(\mathbf{X}_i|T_i=2)^{-1}\boldsymbol{C}_2.
\end{aligned}
\tag{45}
$$

Acknowledgments. The authors would like to thank the patients and families participating in the DMD-PRO-01 and tadalafil DMD trials as well as our collaborators at the Collaborative Trajectory Analysis Project (cTAP), CureDuchenne and Eli Lilly, for making the data available for research. Much of this work was done when the first author was at the Harvard T.H. Chan School of Public Health supported by NIH Grant T32CA009337. We would also like to thank the anonymous referees, an Associate Editor and the Editor for their helpful comments and suggestions.

SUPPLEMENTARY MATERIAL

The statistical performance of matching-adjusted indirect comparisons (DOI: 10.1214/20-AOAS1359SUPP; .zip). The Supplementary Materials report an extended set of simulation results on both bias of various indirect comparison estimators and coverage of proposed standard error estimators for MAIC. The results include settings for both binary and continuous Y as well as different combinations of the level of confounding bias and magnitude and direction of effect modification. The simulations in each setting were repeated for sample sizes ranging from $n = 25$ to $n = 500$ per arm and $p = 5, 10, 15, 20$.

REFERENCES

ADJEI, A. A., CHRISTIAN, M. and IVY, P. (2009). Novel designs and end points for phase II clinical trials. *Clin. Cancer Res.* **15** 1866–1872.

ALTMAN, D., SONG, F., SAKAROVITCH, C., DEEKS, J., D'AMICO, R., BRADBURN, M. and EASTWOOD, A. G. A. (2005). Indirect comparisons of competing interventions. *Health Technology Assessment*.

BLAND, J. M. and ALTMAN, D. G. (2000). Statistics notes. The odds ratio. *BMJ* **320** 1468. https://doi.org/10.1136/bmj.320.7247.1468

BROOKHART, M. A., SCHNEEWEISS, S., ROTHMAN, K. J., GLYNN, R. J., AVORN, J. and STÜRMER, T. (2006). Variable selection for propensity score models. *Am. J. Epidemiol.* **163** 1149–1156.

BUCHER, H. C., GUYATT, G. H., GRIFFITH, L. E. and WALTER, S. D. (1997). The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J. Clin. Epidemiol.* **50** 683–691.

CARO, J. J. and ISHAK, K. J. (2010). No head-to-head trial? Simulate the missing arms. *Pharmacoeconomics* **28** 957–967.

CHENG, D., AYYAGARI, R. and SIGNOROVITCH, J. (2020). Supplement to "The statistical performance of matching-adjusted indirect comparisons: Estimating treatment effects with aggregate external control data." https://doi.org/10.1214/20-AOAS1359SUPP

COHEN, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, San Diego.

DE LUNA, X., WAERNBAUM, I. and RICHARDSON, T. S. (2011). Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98** 861–875. MR2860329 https://doi.org/10.1093/biomet/asr041

EMA CHMP (2006). Guideline on clinical trials in small populations. London: EMEA.

EMA CHMP (2018). Assessment Report: Kymriah (International non-proprietary name: Tisagenlecleucel). Available at https://www.ema.europa.eu/en/documents/assessment-report/kymriah-epar-public-assessment-report_en.pdf. [Online; accessed 1-April-2020].

EMERY, A. E., MUNTONI, F. and QUINLIVAN, R. C. (2015). *Duchenne Muscular Dystrophy*. OUP, Oxford.

FAY, M. P. and GRAUBARD, B. I. (2001). Small-sample adjustments for Wald-type tests using sandwich estimators. *Biometrics* **57** 1198–1206. MR1950428 https://doi.org/10.1111/j.0006-341X.2001.01198.x

FDA (2001). Guidance for Industry. E 10 choice of control group and related issues in clinical trials. US Department of Health and Human Services, Federal Drug Administration.

FDA (2006). Drug Approval Package: Myozyme (Alglucosidase Alfa). Available at https://www.accessdata.fda.gov/drugsatfda_docs/nda/2006/125141s000_MyozymeTOC.cfm. [Online; accessed 11-June-2019].

FDA (2017). FDA Approves First Treatment for a Form of Batten Disease. Available at https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm555613.htm. [Online; accessed 11-June-2019].

GOEMANS, N., VANDEN HAUWE, M., SIGNOROVITCH, J., SWALLOW, E., SONG, J. and PROJECT, C. T. A. (2016). Individualized prediction of changes in 6-minute walk distance for patients with Duchenne muscular dystrophy. *PLoS ONE* **11**.

HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66** 315–331. MR1612242 https://doi.org/10.2307/2998560

HAINMUELLER, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Polit. Anal.* **20** 25–46.

HARTMAN, E., GRIEVE, R., RAMSAHAI, R. and SEKHON, J. S. (2015). From sample average treatment effect to population average treatment effect on the treated: Combining experimental with observational studies to estimate population treatment effects. *J. Roy. Statist. Soc. Ser. A* **178** 757–778. MR3348358 https://doi.org/10.1111/rssa.12094

ICH (2000). Choice of control group and related issues in clinical trials E10.

IMAI, K. and RATKOVIC, M. (2014). Covariate balancing propensity score. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 243–263. MR3153941 https://doi.org/10.1111/rssb.12027

ISHAK, K. J., PROSKOROVSKY, I. and BENEDICT, A. (2015). Simulation and matching-based approaches for indirect comparison of treatments. *Pharmacoeconomics* **33** 537–549.

JANSEN, J. P., FLEURENCE, R., DEVINE, B., ITZLER, R., BARRETT, A., HAWKINS, N., LEE, K., BOERSMA, C., ANNEMANS, L. et al. (2011). Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: Report of the ISPOR task force on indirect treatment comparisons good research practices: Part 1. *Value Health* **14** 417–428.

KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22** 523–539. MR2420458 https://doi.org/10.1214/07-STS227

KAUERMANN, G. and CARROLL, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *J. Amer. Statist. Assoc.* **96** 1387–1396. MR1946584 https://doi.org/10.1198/016214501753382309

KONG, A. (1992). A note on importance sampling using standardized weights. Univ. Chicago, Dept. Statistics, Tech. Rep. 348.

LI, F., MORGAN, K. L. and ZASLAVSKY, A. M. (2018). Balancing covariates via propensity score weighting. *J. Amer. Statist. Assoc.* **113** 390–400. MR3803473 https://doi.org/10.1080/01621459.2016.1260466

LU, G. and ADES, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Stat. Med.* **23** 3105–3124.

LUNCEFORD, J. K. and DAVIDIAN, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Stat. Med.* **23** 2937–2960. https://doi.org/10.1002/sim.1903

MAZZONE, E. S., CORATTI, G., SORMANI, M. P., MESSINA, S., PANE, M., D'AMICO, A., COLIA, G., FANELLI, L., BERARDINELLI, A. et al. (2016). Timed rise from floor as a predictor of disease progression in Duchenne muscular dystrophy: An observational study. *PLoS ONE* **11**.

NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, *Vol. IV. Handbooks in Econom.* **2** 2111–2245. North-Holland, Amsterdam. MR1315971

NIE, L., ZHANG, Z., RUBIN, D. and CHU, J. (2013). Likelihood reweighting methods to reduce potential bias in noninferiority trials which rely on historical data to make inference. *Ann. Appl. Stat.* **7** 1796–1813. MR3127969 https://doi.org/10.1214/13-AOAS655

NIXON, R. M., BANSBACK, N. and BRENNAN, A. (2007). Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat. Med.* **26** 1237–1254. MR2345718 https://doi.org/10.1002/sim.2624

PHILLIPPO, D., ADES, T., DIAS, S., PALMER, S., ABRAMS, K. and WELTON, N. (2016). NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. (Technical Support Documents). Decision Support Unit, ScHARR, University of Sheffield: NICE Decision Support Unit.

POCOCK, S. J. (1976). The combination of randomized and historical controls in clinical trials. *J. Chronic Dis.* **29** 175–188.

PRENTICE, R. L. and PYKE, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66** 403–411. MR0556730 https://doi.org/10.1093/biomet/66.3.403

RICOTTI, V., RIDOUT, D. A., PANE, M., MAIN, M., MAYHEW, A., MERCURI, E., MANZUR, A. Y. and MUNTONI, F. (2016). The NorthStar ambulatory assessment in Duchenne muscular dystrophy: Considerations for the design of clinical trials. *J Neurol Neurosurg Psychiatry* **87** 149–155.

ROTNITZKY, A., LI, L. and LI, X. (2010). A note on overadjustment in inverse probability weighted estimation. *Biometrika* **97** 997–1001. MR2746169 https://doi.org/10.1093/biomet/asq049

RUBIN, D. B. and THOMAS, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics* 249–264.

SIGNOROVITCH, J. E., WU, E. Q., ANDREW, P. Y., GERRITS, C. M., KANTOR, E., BAO, Y., GUPTA, S. R. and MULANI, P. M. (2010). Comparative effectiveness without head-to-head trials. *Pharmacoeconomics* **28** 935–945.

SIGNOROVITCH, J. E., SIKIRICA, V., ERDER, M. H., XIE, J., LU, M., HODGKINS, P. S., BETTS, K. A. and WU, E. Q. (2012). Matching-adjusted indirect comparisons: A new tool for timely comparative effectiveness research. *Value Health* **15** 940–947.

SIGNOROVITCH, J., SWALLOW, E., KANTOR, E., WANG, X., KLIMOVSKY, J., HAAS, T., DEVINE, B. and METRAKOS, P. (2013). Everolimus and sunitinib for advanced pancreatic neuroendocrine tumors: A matching-adjusted indirect comparison. *Experimental Hematology & Oncology* **2** 32.

SNAPINN, S. and JIANG, Q. (2011). Indirect comparisons in the comparative efficacy and non-inferiority settings. *Pharm. Stat.* **10** 420–426.

STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812 https://doi.org/10.1214/09-STS313

STUART, E. A., COLE, S. R., BRADSHAW, C. P. and LEAF, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. *J. Roy. Statist. Soc. Ser. A* **174** 369–386. MR2898850 https://doi.org/10.1111/j.1467-985X.2010.00673.x

SUTTON, A., ADES, A., COOPER, N. and ABRAMS, K. (2008). Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* **26** 753–767.

SWALLOW, E., SONG, J., YUAN, Y., KALSEKAR, A., KELLEY, C., PEEPLES, M., MU, F., ACKERMAN, P. and SIGNOROVITCH, J. (2016). Daclatasvir and sofosbuvir versus sofosbuvir and ribavirin in patients with chronic hepatitis C coinfected with HIV: A matching-adjusted indirect comparison. *Clinical Therapeutics* **38** 404–412.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

WELLS, G., SULTAN, S., CHEN, L., KHAN, M. and COYLE, D. (2009). Indirect evidence: Indirect treatment comparisons in meta-analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health 1–94.

ZEILEIS, A. (2004). Econometric computing with HC and HAC covariance matrix estimators.

ZHANG, Z., NIE, L., SOON, G. and HU, Z. (2016). New methods for treatment effect calibration, with applications to non-inferiority trials. *Biometrics* **72** 20–29. MR3500570 https://doi.org/10.1111/biom.12388

ZHAO, Q. and PERCIVAL, D. (2017). Entropy balancing is doubly robust. *J. Causal Inference* **5**.

ZUBIZARRETA, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *J. Amer. Statist. Assoc.* **110** 910–922. MR3420672 https://doi.org/10.1080/01621459.2015.1023805