Rejoinder: Estimands and their Estimators for Clinical Trials Impacted by the COVID-19 Pandemic: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions

Kelly Van Lancker, Sergey Tarima, Jonathan Bartlett, Madeline Bauer, Bharani Bharani-Dharan, Frank Bretz, Nancy Flournoy, Hege Michiels, Camila Olarte Parra, James L. Rosenberger, and Suzie Cro

1. Stijn Vansteelandt’s Discussion Points

We certainly agree with Prof. Vansteelandt that the data from randomized clinical trials are “precious” and that analysis approaches should control the Type I error rate of the null hypothesis tests of no treatment effect where possible (Vansteelandt 2022). He suggests that from the description of intercurrent events 6 and 7, it would be reasonable to assume they would occur with equal rates in both arms of the trial (i.e., both trial arms are equally affected by COVID-19). Although this might sometimes be a reasonable assumption, there are intercurrent events that may well affect the treatment arms differentially. Such an intercurrent event is discussed, for example, in “Randomization Tests to Address Disruptions in Clinical Trials: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions” by Uschner et al. (submitted). Specifically, drug supply issues can affect different treatment arms in different ways. It might, for example, be the case that the standard-of-care is available, while the experimental treatment is not (due to drug supply issues). The study drug might then be replaced with a different drug, or even lead to a disruption in the planned randomization of patients (e.g., through “forced randomization,” Rosenkranz 2011).

1.1. Allowing for Changes in the Trial Population

Prof. Vansteelandt makes an attractive proposal for handling the effects of population shifts in the statistical analysis. He proposes outcome regression based, inverse weighting based, and a doubly robust combination of these, that target the effect in the pre-pandemic population, which exploit outcome information in those patients randomized after the pandemic emerged. These estimators exploit the observed outcomes in the latter, which should lead to improved efficiency relative to estimators which only use data from pre-pandemic patients. These rely on an assumption that patients who enter before the start of the pandemic have similar outcomes as patients with the same baseline covariates who enter the same arm of the trial after the start of the pandemic. We note that often in practice, the (in general conditional) treatment effect is estimated as the coefficient of treatment in a suitable regression model for the outcome given treatment and covariates. In this case, the stated conditional independence assumption would mean that the treatment effect could simply be estimated by fitting this model using data from both pre- and post-pandemic patients (assuming that the covariates adjusted for in this model correspond to those deemed to be required to render the conditional assumption true). Assuming this outcome regression is correctly specified, we believe this estimator would be more efficient than the inverse probability weighted and doubly robust versions (which usually focus on a marginal treatment effect).

Whether such an assumption is deemed reasonable will have to be judged on a case by case basis. In general, we might anticipate that patients with the same treatment and baseline covariates may experience different outcomes after the pandemic emerged, for example, because the pandemic affected the ability of patients to adhere to the intended treatment regime or COVID-19 infection impacted the disease course (Unnikrishnan and Misra 2021; Wang et al. 2022). Nevertheless, we believe the proposed approach is promising, particularly given the property that it controls the Type I error rate under the null hypothesis of no treatment effect even when the aforementioned assumption is violated.
1.2. Removing Effects of Intercurrent Events

Under a hypothetical strategy, removing the effects of the intercurrent events on observed outcomes generally requires adjustment for common causes of the intercurrent event and outcome. Prof. Vansteelandt's proposal to use time of recruitment as an instrumental variable offers the potential to avoid the need for adjustment for such variables. This is particularly attractive given that measurements of such confounders may be incomplete given the unanticipated nature of the pandemic related intercurrent events. We are enthusiastic about further research into this approach, in particular as to how the models should be adapted to account for how the timing or frequency of occurrence of the intercurrent events affects outcomes, and the fact that the impact of these intercurrent events on outcomes may well vary over time.

1.3. Combining Unbiased and Biased Estimators

When combining an unbiased estimator \( \hat{\theta} \) \((E(\hat{\theta}) = \theta)\) and a biased one \( \hat{\theta} \) \((E(\hat{\theta}) = \theta - \delta)\), the estimator \( \hat{\theta}^0(\hat{\delta}) = \hat{\theta} - \text{var}(\hat{\theta}) \left[ \text{var}(\hat{\theta}) + \text{var}(\hat{\theta}) + \delta^2 \right]^{-1} \hat{\delta} \), with \( \hat{\delta} = \hat{\theta} - \theta \), has the smallest mean squared error but depends on unknown variances and expectation. In practice, we can use

\[
\hat{\theta}^0(\hat{\delta}) = \hat{\theta} - \text{var}(\hat{\theta}) \left[ \text{var}(\hat{\theta}) + \text{var}(\hat{\theta}) + \delta^2 \right]^{-1} \hat{\delta}.
\]

As we replaced \( \delta \) and the unknown variances by respectively \( \hat{\delta} \) and estimated variances, we no longer have the guarantee that the suggested estimator minimizes the mean squared error. We therefore investigate this in Monte Carlo simulations in Section 1.3.1.

Prof. Vansteelandt asked about the impact of \( \delta^2 \) on the large sample properties of \( \hat{\theta}^0(\hat{\delta}) \) and expressed concern about its \( O(n^{-1/2}) \) bias, which raises questions about its practical utility. In addition, he stated that it was not clear how to make inference in such settings.

To address these questions we assume, for simplicity, that \( \hat{\theta} \) is a sample mean based on an independent sample \( X_1, \ldots, X_n \) with \( E(X_i) = \theta \) and \( \text{var}(X_i) = \sigma^2 \). The estimator \( \hat{\theta} \) is also a sample mean but from another independent sample \( Y_1, \ldots, Y_m \) with \( E(Y_i) = \theta - \delta \) and \( \text{var}(Y_i) = \sigma^2 \). Assuming consistent variance estimation (say, \( \hat{\sigma} \to \sigma \), in probability), it is sufficient to assume that \( \sigma^2 \) is known and only on the impact of estimation of \( \delta \). Then, \( \hat{\theta}^0(\hat{\delta}) = \hat{\theta} - \sigma^2 \left[ \sigma^2 + 1 + n/m \right] \hat{\delta} \).

Tarima et al. (2020) used local asymptotics to evaluate statistical properties of a more general version of \( \hat{\theta}^0(\hat{\delta}) \). Here, we use a similar reasoning to address the effect of estimating \( \delta^2 \) on large sample properties. Namely, since clinical trials are typically designed to detect an effect size proportional to \( n^{-1/2} \), asymptotic properties should be considered against alternatives that shrink to zero at rate \( n^{-1/2} \). This leads to interest in an effect size \( \delta = h \cdot n^{-1/2} = O(n^{-1/2}) \), where \( h \) is a local alternative. The case of \( \delta = O(n^{-1/2+\epsilon}) \), \( \epsilon > 0 \), is not interesting because the impact of \( Y_1, \ldots, Y_m \) on estimation of \( \theta \) is asymptotically suppressed and the asymptotic distributions of \( \sqrt{n} (\hat{\theta} - \theta) \) and \( \sqrt{n} (\hat{\theta}^0(\hat{\delta}) - \theta) \) become equivalent.

We derive the asymptotic distribution of \( \hat{\theta}^0(\hat{\delta}) \) from which statistical inference can be performed, under the assumption that \( (n/m) \to p \), where \( 0 < |p| < \infty \). When \( \delta = h/\sqrt{n} \),

\[
\sqrt{n} (\hat{\theta} - \theta) \to \mathcal{N} (0, \sigma^2) \quad \text{and} \quad \sqrt{m} (\hat{\theta} - \theta - h/\sqrt{n}) \to \mathcal{N} (0, \sigma^2).
\]

From

\[
\sqrt{n} \delta = \sqrt{n} (\hat{\delta} - \delta) = \sqrt{n} (\hat{\theta} - \theta) - \frac{n}{m} \sqrt{m} (\hat{\theta} - \theta - h/\sqrt{n}) = \xi_n - \xi_n.
\]

we find that \( \sqrt{n} \delta \to \xi_n - \xi_n \to \xi - \xi \), where \( \xi \to \mathcal{N} (h, p\sigma^2) \).

Then

\[
\sqrt{n} \left( \hat{\theta}^0(\hat{\delta}) - \theta \right) \to \mathcal{N} (\hat{\theta} - \theta - \frac{\sigma^2}{\sigma^2 + 1 + n/m} \hat{\delta}^2, \hat{\sigma}^2(\xi - \xi)^2),
\]

The asymptotic distribution of \( \sqrt{n} (\hat{\theta} - \theta) \) is normal and only depends on \( \sigma \), whereas the asymptotic distribution of \( \sqrt{n} (\hat{\theta}^0(\hat{\delta}) - \theta) \) depends on three parameters: \( \sigma \), \( p \), and \( h \).

Above, we assumed that \( \sigma \) is known and only \( p \) and \( h \) need to be estimated. The \( \hat{\theta} = \sqrt{n} (\hat{\theta} - \theta) \) and \( \hat{\sigma} = n/m \) are reasonable estimator choices, which can be used by parametric bootstrap procedures to approximate the asymptotic distribution shown in Equation (1). By analogy with the use of the sample standard deviation instead of \( \sigma \) in Gaussian asymptotics, these estimators may impact the finite sample distribution of \( \hat{\theta}^0(\hat{\delta}) \). Figure 1 shows results of Monte Carlo simulations for large sample characteristics (quantiles, bias, square root of the mean squared error (RMSE), and Type I error rate) of the estimators under given values of \( \sigma \), \( h \), and \( p \).

1.3.1. Illustrative Example

Let \( n = 100 \), \( m = 1000 \) and \( \sigma = 1 \). Then, the asymptotic 95% confidence interval (CI) for \( \sqrt{n} (\hat{\theta} - \theta) \) is defined by the 0.025- and 0.975-level quantiles of a standard normal distribution: \(-1.96 \) to \( 1.96 \). The width of this confidence interval is \( 3.92 \).

An asymptotic 95% confidence interval for \( \sqrt{n} (\hat{\theta}^0(\hat{\delta}) - \theta) \) depends on \( p \) and \( h \). Figure 1(a) shows how 0.025- and 0.975-level quantiles change with \( h \). The width of the confidence interval for \( \sqrt{n} (\hat{\theta}^0(\hat{\delta}) - \theta) \) is shown in green at \( m = 10 \cdot n \) and changes from \( \approx 3.11 \) to \( \approx 4.20 \). The best reduction of the CI when compared to the \( \hat{\theta} \) case is 20.8% and, in the worse case scenario, it becomes 7.1% wider. In addition, we also show the extreme case of \( m = \infty \) (blue): the smallest reduction is 21.3% and in a worst case the CI can become 8.2% wider.

Figure 1(b) shows how bias depends on \( h \). At \( h = 0 \) the asymptotic distribution is symmetric and there is no bias, but as \( h \) departs from 0, the bias starts increasing, reaching its maximum \( \approx 0.37 \) (\( \approx 0.35 \)) at around \( h = \pm 1.8 \) (\( h = \pm 1.8 \)) and \( m = \infty \) (\( m = 1000 \)).
Figure 1. Monte-Carlo simulations with $10^6$ repetitions for $\zeta_n$, $\eta_m$, and $\eta_\infty$. The subscript $m$ was added to highlight how many $Y_i$ are used by $\hat{\theta}_0(\delta)$. $n = 100$, $m = 1000$.

Figure 1(c) shows the results of Monte Carlo experiments for $h$ taking values between $-5$ and $5$ with 0.1 increments, and $p$ taking values 0.1, 1 and 10. Each Monte Carlo experiment is based on $10^6$ repetitions with $n = 100$ and $\sigma = 1$. All in all, the figure supports the theory that the MSE is always improved (black) when $\theta_0(\delta)$ is used; the highest improvement happens at $h = 0$ and $p = 0.1$ (70% reduction in RMSE). The estimator $\theta_0(\delta)$ (green) is not as good as $\theta_0(\delta)$ (black) but we still observe 18% reduction in RMSE at $h = 0$ and $p = 0.1$; in the worst case scenario, RMSE can increase by 11%.

Figure 1(d) shows that hypothesis testing based on $\hat{\theta}_0(\delta)$ provides good control of Type I error when the nuisance parameters are known. Table 1 shows Type I error in a setting when the nuisance parameters are estimated and a parametric bootstrap is used to calculate $p$-values and 95% confidence intervals.

The use of parametric bootstrap procedures makes Monte Carlo simulations computationally intensive. We performed a small Monte-Carlo simulation study with 2000 repetitions at $n = 1000$ and $\sigma = 1$ (see Table 1). The unknown parameters $h$, $p$ and $\sigma$ in the distribution of $\hat{\theta}_0(\delta)$ were estimated with $\hat{h}$, $\hat{p}$, and
and a parametric bootstrap was applied to calculate p-values and 95% confidence intervals (10,000 bootstrap resamples were used within the procedure). Table 1 shows that the use of a parametric bootstrap for \( \hat{\delta} \) provides good control of Type I error and coverage under the considered values of \( \delta = h / \sqrt{1000} \). The MSE is reduced for smaller values of \( \delta \) and close to 0.001 (the variance of \( \hat{\delta} \) when \( \delta = 0.1 \)).

### 1.3.2. Summary on Combining Biased and Unbiased Estimators

Combining biased and unbiased estimators is not a trivial task and researchers need to weigh all costs and pros associated with the combined estimator. Since the closed-form mathematical expression of the distribution of the combined estimators is not readily available, performing a Monte Carlo study for a specific setting would be a good starting point. Plus, expert opinion would be a valuable source of information on possible values of true \( \delta \). As shown in Figure 1(a) some benefits are expected if the researcher combines biased and unbiased estimators and believes that the magnitude of the bias \( \delta \) is rather small, but at the same time, the researcher does not want to jeopardize the validity of the study if this belief is not true. In our illustrative example the width of the 95% confidence interval was narrower for \( \hat{\delta}^b (\delta) \) than \( \hat{\delta} \) when \( |h| < 2.2 \) (\( |h| < 2.4 \)) at \( m = 1000 \) \((m = \infty)\).

In our original manuscript, we mainly focused on the estimation problem, but the combined estimator can also be used for hypothesis testing (Van Lancker et al. 2022). When the asymptotic distribution of \( \hat{\delta}^b (\delta) \) is used for hypotheses testing, asymptotic control of the Type I error (say 5%) is ensured because the asymptotic distribution of Equation (1) is known under the null. Figure 1(d) shows good control of the Type I error for testing based on \( \hat{\delta}^b (\delta) \) (green) and \( \hat{\delta} \) (red). For the considered settings in Table 1, the Type I error seems to be controlled when \( p \)-values are obtained via the parametric bootstrap.

Thus, the asymptotically justified estimator combining unbiased and biased estimators seems to perform reasonably well in large samples, although further extensive simulations are strongly recommended. Its potential moderate benefits and minor losses depend on the parameters of its asymptotic distribution: the local shift \( h \) and the ratio of sample sizes \( p \). Before making design, analysis, and hypothesis testing decisions on the use of the combined estimator, we therefore strongly advise researchers to evaluate the performance of the combined estimator (RMSE, coverage probabilities, Type I error rate, power, etc.) in a series of Monte Carlo simulation studies under plausible settings. Additional benefits of using the combined estimator include (a) its model independence since only two moments need to exist to build the estimator, (b) robustness to bias because the combined estimator detects and suppresses the effect of bias at large samples, and (c) its known asymptotic distribution allows performing various types of statistical inference.

The focus on asymptotics ensures us that the combination of unbiased and biased estimators can be extended to asymptotically unbiased estimators where the asymptotic bias is \( o (1 / \sqrt{n}) \), which is typical for many MLEs under certain regularity conditions.

## 2. Akacha and Lyu’s Discussion Points

We are pleased to see support from Drs. Akacha and Lyu for using hypothetical estimands, which remove the effects of administrative and operational related intercurrent events related to the pandemic (Akacha and Lyu 2022). It is interesting to hear that received regulatory feedback has argued that intercurrent events relating to operational complications should alternatively be handled with a treatment policy strategy. We also hope that these discussions help harmonize the choice of suitable estimands and highlight here that the hypothetical estimands explored could be used for either primary or supplementary analysis. The use of supplementary estimands to handle COVID-19 related intercurrent events in a different manner to the primary estimand may help to provide a fuller picture of treatment performance and any impact of COVID-19 on trial results.

We moreover agree with Drs. Akacha and Lyu that comparing estimators which exploit post-intercurrent event data to estimate hypothetical estimands is an important area for future study. It should be assessed on a case by case basis whether borrowing information from the outcomes observed after the intercurrent event is sensible. For instance, this could be the case where due to supply chain issues, the study drug was temporarily discontinued. In this context, we may consider that patients who have a positive outcome even without the drug, would have had a positive outcome if the study drug was not discontinued. If exploiting post-intercurrent information is deemed reasonable, there are different available estimators that can handle it. The paper by Lasch and Guizzaro (2022) compared various methods which exploit post-intercurrent event data when estimating hypothetical estimands, and concluded that Loh’s g-estimation approach (Loh et al. 2020) appeared to be best. The investigation by Lasch and Guizzaro (2022) did not include G-formula (G-computation) estimators which used post-intercurrent event data. However, it is known that, at least in special cases, G-estimators and G-formula estimators of controlled direct effects can be identical (Vansteelandt 2009). Future research should investigate to what extent this equivalence holds.

### 2.1. Doubly Robust Estimators

Drs. Akacha and Lyu raise important questions about the viability of doubly robust augmented inverse probability estimators. They note that under misspecification of both models, augmented inverse probability weighting can sometimes do worse than likelihood (imputation)-based methods (Kang and Schafer 2007). These can occur in particular when some individuals have very small propensity score values, and consequently, are assigned very large weights. In this case, estimators based on inverse weighting (including the “standard” doubly robust estimator) can exhibit poor performance. However, there are analogous settings where likelihood based estimators can perform worse than weighting and doubly robust methods (Tsatis and Davidian 2007). Moreover, alternative doubly robust estimators have been developed which achieve comparable or improved performance relative to existing methods, even with some estimated propensity scores close to zero (Cao, Tsatis, and Davidian 2009). More recent developments in this direction include the biased-reduced doubly robust.
estimator that improves performance under model misspecification (Vermeulen and Vansteelandt 2015). In addition, the targeted minimum loss-based estimation (TMLE) framework (see e.g., van der Laan and Gruber 2012) allows the use of machine learning in the construction of doubly robust estimators, placing minimal assumptions on the distribution of the data. It would be a useful contribution to the clinical trials literature to investigate these approaches in Monte Carlo simulation studies.

We admit that obtaining stable inverse weights can be problematic. A first approach—which was mentioned by Drs. Akacha and Lyu—is to truncate the extreme weights. Although one could specify the level at which one truncates in advance, these type of approaches are generally ad hoc and can introduce bias. A more principled approach is to estimate the weights in a way that promotes stability, that is, prevents extreme weights (e.g., Imai and Ratkovic 2014; Zubizarreta 2015; Avayan and Vansteelandt 2021). It’s an interesting question which of these recently introduced methods are best suited for prespecification in the analysis plan of confirmatory clinical trials.

As for multiple imputation, inverse weighting methods (including doubly robust augmented inverse probability weighting) have the attractive feature that the missingness can be conditional on different variables, hence, potentially including (different) auxiliary information for different intercurrent events and missing data patterns. This can in principle be accommodated by modeling the intercurrent events separately and multiplying the weights. This, however, makes the rather strong assumption that the different intercurrent events are independent of each other. Further research and Monte Carlo simulations would be useful to evaluate the effects of weakening this assumption.

Finally, Drs. Akacha and Lyu raised a question about the role of methods which move beyond double robustness by allowing multiple models for both the propensity score and the outcome regression models. While the approaches in Han and Wang (2013) might be relevant, it is important to realize that these estimators are still in a sense doubly robust rather than truly multiply robust (Wang and Tchetgen Tchetgen 2018). The approach of Han and Wang (2013) combines different models for two components of the likelihood, the conditional outcome mean and the propensity score. Multiply robust estimators instead require specification of three or more components of the likelihood. Whether this is possible or not depends on the estimand of interest.

**Supplementary Materials**

The R code to obtain the plots in Figure 1 is available in the supplementary materials.

**Acknowledgments**

We would like to thank the discussants Mouna Akacha, Tianmeng Lyu, and Stijn Vansteelandt for their interesting and thoughtful comments on our paper “Estimators for Clinical Trials Impacted by the COVID-19 Pandemic: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions.” (Van Lancker et al. 2022). We also would like to extend our gratitude to Toshimitsu Hamasaki for making this a discussion article.

**Disclosure statement**

J.B.'s institution has received consultancy fees for the author's advice on statistical methodology from AstraZeneca, Bayer, Novartis, Roche. J.B. has received consultancy fees from Bayer and Roche, and fees for provision of online courses from Roche.

**Funding**

K.V.L. is supported by the Ghent University Special Research Fund under award number BOF1P07421. S.T. is partially supported by the Department of Health and Human Services of the National Institutes of Health under award number R40MC41748. J.B. and C.O.P. are supported by UK Medical Research Council under grant agreement MR/T023953/1. H.M. is supported by VLAIO under Baekeland grant agreement HBC.2019.2155. S.C. is supported by an NIHR advanced research fellowship (NIHR 300593). The views expressed are those of the authors and not necessarily those of Ghent University Special Research Fund, VLAIO, the National Institutes of Health, the UK Medical Research Council, the NIHR or the Department of Health and Social Care.

**ORCID**

Kelly Van Lancker http://orcid.org/0000-0003-3458-3788

Sergey Tarima http://orcid.org/0000-0001-7542-1094

Jonathan Bartlett https://orcid.org/0000-0001-7117-0195

Frank Bretz http://orcid.org/0000-0002-2008-8340

Nancy Flournoy http://orcid.org/0000-0003-1299-9299

Hege Michiels http://orcid.org/0000-0002-3165-2330

Camila Olarte Parra http://orcid.org/0000-0003-0263-4392

James L. Rosenberger http://orcid.org/0000-0003-2939-5475

Suzie Cro http://orcid.org/0000-0002-6113-1173

**References**

Akacha, M., and Lyu, T. (2022), "Comment on ‘Estimands and Their Estimators for Clinical Trials’ by Chapman and Ehmke: A Report from the NIHSS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions,” *Statistics in Biopharmaceutical Research*. [122]

Avayan, V., Vansteelandt, S. (2021), “Stable Inverse Probability Weighting Estimation for Longitudinal Studies,” *Scandinavian Journal of Statistics*, 48, 1046–1067. [123]

Cao, W., Tsiatis, A. A., and Davidian, M. (2009), "Improving Efficiency and Robustness of the Doubly Robust Estimator for a Population Mean with Incomplete Data," *Biometrika*, 96, 723–734. [122]

Han, P., and Wang, L. (2013), "Estimation with Missing Data: Beyond Double Robustness," *Biometrika*, 100, 417–430. [123]

Imai, K., and Ratkovic, M. (2014), "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society, Series B*, 76, 243–263. [123]

Kang, J. D., and Schafer, J. L. (2007), "Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data," *Statistical Science*, 22, 523–539. [122]

Lasch, F., and Guizzaro, L. (2022), “Estimators for Handling COVID-19-Related Intercurrent Events with a Hypothetical Strategy,” *Pharmaceuti- cal Statistics*, 21, 1258–1280. [122]

Loh, W. W., Moerkerke, B., Loey, T., Poppe, L., Crombez, G., and Vansteelandt, S. (2020), “Estimation of Controlled Direct Effects in Longitudinal Medication Analyses with Latent Variables in Randomized Studies,” *Multivariate Behavioral Research*, 55, 763–785. [122]

Rosenkranz, G. K. (2011), "The Impact of Randomization on the Analysis of Clinical Trials," *Statistics in Medicine*, 30, 3475–3487. [119]

Tarima, S., Tuyishimire, B., Sparapani, R., Rein, L., and Meurer, J. (2020), "Estimation Combining Unbiased and Possibly Biased Estimators," *Journal of Statistical Theory and Practice, 14*, 1–20. [120]

Tsiatis, A. A., and Davidian, M. (2007), "Comment: Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a
Population Mean from Incomplete Data,” *Statistical Science*, 22, 569–573. [122]

Unnikrishnan, R., and Misra, A. (2021), “Diabetes and COVID-19: A Bidirectional Relationship,” *Nutrition & Diabetes*, 11, 1–5. [119]

van der Laan, M. J., and Gruber, S. (2012), “Targeted Minimum Loss Based Estimation of Causal Effects of Multiple Time Point Interventions,” *The International Journal of Biostatistics*, 8. [123]

Van Lancker, K., Tarima, S., Bartlett, J., Bauer, M., Bharani-Dharan, B., Bretz, F., Flournoy, N., Michiels, H., Olarte Parra, C., Rosenberger, J. L., and Cro, S. (2022), “Estimands and their Estimators for Clinical Trials Impacted by the COVID-19 Pandemic: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions,” *Statistics in Biopharmaceutical Research*, 1–12. https://doi.org/10.1080/19466315.2022.2094459. [122,123]

Vansteelandt, S. (2009), “Estimating Direct Effects in Cohort and Case–Control Studies,” *Epidemiology*, 20, 851–860. [122]

——— (2022), “Comment on ‘Estimands and Their Estimators for Clinical trials Impacted by the COVID-19 Pandemic: A Report from the NISS Ingram Olkin Forum Series on Unplanned Clinical Trial Disruptions,’” *Statistics in Biopharmaceutical Research*. [119]

Vermeulen, K., and Vansteelandt, S. (2015), “Bias-Reduced Doubly Robust Estimation,” *Journal of the American Statistical Association*, 110, 1024–1036. [123]

Wang, L., Davis, P. B., Volkow, N. D., Berger, N. A., Kaelber, D. C., and Xu, R. (2022), “Association of COVID-19 with New-Onset Alzheimer’s Disease,” *Journal of Alzheimer’s Disease*, 1–4 (Preprint). [119]

Wang, L., and Tchetgen Tchetgen, E. (2018), “Bounded, Efficient and Multiply Robust Estimation of Average Treatment Effects using Instrumental Variables,” *Journal of the Royal Statistical Society*, Series B, 80, 531–550. [123]

Zubizarreta, J. R. (2015), “Stable Weights that Balance Covariates for Estimation with Incomplete Outcome Data,” *Journal of the American Statistical Association*, 110, 910–922. [123]