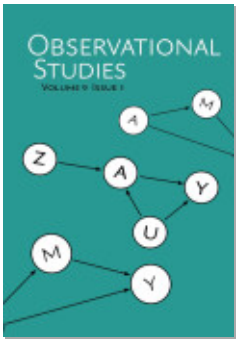
The Pursuit of Efficiency versus Robustness: A Learning Experience from Analyzing a Semiparametric Nonignorable Propensity Score Model



Samidha Shetty, Yanyuan Ma, Jiwei Zhao

Observational Studies, Volume 9, Issue 1, 2023, pp. 97-104 (Article)

Published by University of Pennsylvania Press
DOI: *https://doi.org/10.1353/obs.2023.0009*

Observational Studies 9.1 (2023): 97-104 Published Jan 2023

The Pursuit of Efficiency versus Robustness: A Learning Experience from Analyzing a Semiparametric Nonignorable Propensity Score Model

Samidha Shetty **sss269@psu.edu**
Department of Statistics
Pennsylvania State University
University Park, PA 16802

Yanyuan Ma **yzm63@psu.edu**
Department of Statistics
Pennsylvania State University
University Park, PA 16802

Jiwei Zhao **jiwei.zhao@wisc.edu**
Department of Biostatistics & Medical Informatics
University of Wisconsin-Madison
Madison, WI 53726

Abstract

Rosenbaum and Rubin's pioneering work on "The Central Role of the Propensity Score in Observational Studies for Causal Effects" has shaped the landscape of the literature in causal inference and missing data analysis. In the past decades, the concept of propensity score has been used not only under ignorability assumption, but also under nonignorability assumption. The nice properties of double robustness and semiparametric efficiency are well known under ignorability; however, the situation is a lot more sophisticated under nonignorability. In this paper, we summarize what we have learnt from analyzing a semiparametric nonignorable propensity score model. It turns out that, under nonignorability, the efficient estimator for the quantity of interest might be too complicated to be practically implemented. On the other hand, by sacrificing the efficiency to some extent, one type of robust estimators is much easier to derive and implement; hence is recommended. This is a general tradeoff between efficiency and robustness in a typical semiparametric model.

Keywords: Efficiency, Ignorability, Influence function, Nonignorability, Propensity score, Robustness

1. Propensity Score

The propensity score was introduced in Rosenbaum and Rubin (1983). It is a seminal work. In the past four decades, it has flourished numerous novel ideas and fascinating methods for estimating causal effects and for analyzing data with missing values. More importantly, it has shaped the landscape of the literature on causal inference and missing data analysis. The idea of using propensity score model has been not only well studied in the discipline of statistics, but also fruitfully applied in social sciences, health sciences, and biomedical studies.

In this work, our discussion mainly extends the propensity score from under ignorable assumption to nonignorable. Though the elegant double robustness and semiparametric efficiency are crystal clear under ignorability, they are not automatic under nonignorability. Without sacrificing any efficiency, the optimal estimator under nonignorability might be too complicated to be practically implementable. We use a semiparametric nonignorable propensity score model as an exemplar to elucidate a tradeoff between efficiency and robustness.

Our discussion can be presented in the context of either causal inference or missing data analysis, and we take the latter simply for brevity. We first introduce some notations. Throughout, we denote scalar Y as the outcome, and binary variable R as the indicator of whether Y is observed or not; i.e., $R = 1$ if Y is observed and $R = 0$ if otherwise. We use $\mathbf{X}$ to collect all the covariates and we assume $\mathbf{X}$ is fully observed. Suppose the interest is to estimate some summary of the outcome Y, say, $E\{\boldsymbol{\zeta}(Y)\}$, where $\boldsymbol{\zeta}(\cdot)$ is a known function. In applications, we have N independent and identically distributed copies of $(R, RY, \mathbf{X})$, and we denote n the sample size with completely observed data.

2. Propensity Score under Ignorability: Double Robustness and Semiparametric Efficiency

We first briefly review the results under ignorability assumption. Denote $\pi(y, \mathbf{x})$ as $\mathrm{pr}(R = 1 \mid y, \mathbf{x})$. Ignorability simply means $\pi(y, \mathbf{x}) = \pi(\mathbf{x})$; that is,

$$\mathrm{pr}(R = 1 \mid Y, \mathbf{X}) = \mathrm{pr}(R = 1 \mid \mathbf{X}). \tag{1}$$

Equivalently, R and Y are conditionally independent given the value of $\mathbf{X}$. The ignorability assumption has been termed missing-at-random (MAR) in traditional missing data analysis (Little and Rubin, 2019). In causal inference, it has been variously described as unconfoundedness, selection on observables, or, exogeneity; see, e.g., Imbens (2004); Imbens and Rubin (2015).

Under assumption (1), the joint distribution from one single observation of $(R, RY, \mathbf{X})$ is

$$f_{\mathbf{X}}(\mathbf{x})\{\pi(\mathbf{x}) f_{Y|\mathbf{X}}(y, \mathbf{x})\}^r \{1 - \pi(\mathbf{x})\}^{1-r}, \tag{2}$$

where $f_{Y|\mathbf{X}}(y, \mathbf{x})$ encodes the conditional distribution of Y given $\mathbf{X}$, and $f_{\mathbf{X}}(\mathbf{x})$ the marginal distribution of $\mathbf{X}$. Then, following the routine of characterizing the geometric structure of the semiparametric model (Bickel et al., 1993; Tsiatis, 2006), one can derive, the nuisance tangent space

$$\mathcal{T} = \mathcal{T}_1 \oplus \mathcal{T}_2 \oplus \mathcal{T}_3,$$

where $\mathcal{T}_1 = [\{r - \pi(\mathbf{x})\}\mathbf{a}(\mathbf{x}) : \forall \mathbf{a}(\mathbf{x})]$ is the nuisance tangent space of $\pi(\mathbf{x})$, $\mathcal{T}_2 = \{r\mathbf{b}(y, \mathbf{x}) : E(\mathbf{b} \mid \mathbf{x}) = \mathbf{0}\}$ is the nuisance tangent space of $f_{Y|\mathbf{X}}(y, \mathbf{x})$, and $\mathcal{T}_3 = \{\mathbf{c}(\mathbf{x}) : E(\mathbf{c}) = \mathbf{0}\}$ is the nuisance tangent space of $f_{\mathbf{X}}(\mathbf{x})$.

To estimate $E\{\boldsymbol{\zeta}(Y)\}$, its efficient influence function is

$$\begin{aligned}\boldsymbol{\phi}_{\text{eff}} &= \underbrace{\frac{r}{\pi(\mathbf{x})}[\boldsymbol{\zeta}(y) - E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}]}_{\in \mathcal{T}_2} + \underbrace{E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\} - E\{\boldsymbol{\zeta}(Y)\}}_{\in \mathcal{T}_3}, \\ &= \frac{r}{\pi(\mathbf{x})}[y - E\{\boldsymbol{\zeta}(Y)\}] - \frac{r - \pi(\mathbf{x})}{\pi(\mathbf{x})}[E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\} - E\{\boldsymbol{\zeta}(Y)\}].\end{aligned} \tag{3}$$

The well-known property of double robustness and semiparametric efficiency of estimating $E\{\boldsymbol{\zeta}(Y)\}$ can be easily seen from analyzing its efficient influence function $\boldsymbol{\phi}_{\text{eff}}$. We concisely write below.

1. Robustness to the misspecification of the propensity score $\pi(\mathbf{x})$: We have $E(\boldsymbol{\phi}_{\text{eff}}) = \mathbf{0}$ as long as $E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}$ is correctly specified. This is also indicated by the fact that $\boldsymbol{\phi}_{\text{eff}}$ is orthogonal to $\mathcal{T}_1$, the nuisance tangent space corresponding to the propensity score.

2. Robustness to the misspecification of $E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}$: We have $E(\boldsymbol{\phi}_{\text{eff}}) = \mathbf{0}$ as long as the propensity score $\pi(\mathbf{x})$ is correctly specified. If $E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}$ is misspecified as zero, $\boldsymbol{\phi}_{\text{eff}}$ becomes the influence function of the so-called inverse probability weighting estimator.

3. Semiparametric efficiency: if both $\pi(\mathbf{x})$ and $E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}$ are correctly specified, $\boldsymbol{\phi}_{\text{eff}}$ leads to the semiparametrically efficient estimator. The efficiency lower bound is $E(\boldsymbol{\phi}_{\text{eff}}\boldsymbol{\phi}_{\text{eff}}^{\mathrm{T}})$.

3. Propensity Score under Nonignorability: Efficiency versus Robustness

Ignorability may not hold in various applications such as patient-reported outcomes, electronic health records, and mobile health. See, e.g., Frankel et al. (2012); Gomes et al. (2016); Mercieca-Bebber et al. (2016); Ayilara et al. (2019); Groenwold (2020); Carreras et al. (2021); Goldberg et al. (2021); Lim et al. (2021). If the ignorability assumption (1) is not satisfied, the propensity score is called nonignorable, or missing-not-at-random (MNAR) in the missing data literature. Nonignorability does exist in causal inference as well, where researchers could assume the treatment assignment to depend on potential outcomes, or, in general, the unmeasured confounder.

To make the propensity score assumption as flexible as possible while still achieving model identifiability (Rotnitzky and Robins, 1997), a semiparametric nonignorable propensity score has been proposed in Shao and Wang (2016) and further studied in Shetty et al. (2022). The propensity score is defined as

$$\pi(y, \mathbf{x}) = \pi(y, \mathbf{u}, \boldsymbol{\beta}, g) = \text{expit}\{h(y, \boldsymbol{\beta}) + g(\mathbf{u})\}, \tag{4}$$

where $\text{expit}(\cdot) = \exp(\cdot)/\{1 + \exp(\cdot)\}$, $\boldsymbol{\beta}$ is an unknown d-dimensional parameter (parametric component), $h(\cdot)$ is a known function, $h(0, \boldsymbol{\beta}) = 0$, and $g(\cdot)$ is an arbitrary unspecified function (nonparametric component). The covariate $\mathbf{X}$ can be split as $\mathbf{X} = (\mathbf{U}^{\mathrm{T}}, \mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}$, and $\mathbf{Z}$ is termed the shadow variable in the literature. The shadow variable concept enables

the model identifiability by assuming that the propensity score does not depend on the shadow variable. It is also possible that the nonignorable propensity score depends on the entire variable $\mathbf{X}$, but then in such case other assumptions have to be imposed to attain the model identifiability. Putting this restriction aside, the assumption in (4) does generalize the ignorability assumption (1), and becomes (1) if the true value of $\boldsymbol{\beta}$ is zero.

Under assumption (4), the joint likelihood of $(\mathbf{X}, R, RY)$ can be written as

$$f_{\mathbf{X},R,RY}(\mathbf{x}, r, ry) = f_{\mathbf{X}}(\mathbf{x})\{f_{Y|\mathbf{X}}(y, \mathbf{x})\pi(y, \mathbf{u}, \boldsymbol{\beta}, g)\}^r\{1 - \int f_{Y|\mathbf{X}}(y, \mathbf{x})\pi(y, \mathbf{u}, \boldsymbol{\beta}, g)dy\}^{1-r}.$$

Using the fact that

$$f_{Y|\mathbf{X}}(y, \mathbf{x}) = \frac{f_{Y|\mathbf{X},R=1}(y, \mathbf{x})/\pi(y, \mathbf{u}, \boldsymbol{\beta}, g)}{\int f_{Y|\mathbf{X},R=1}(t, \mathbf{x})/\pi(t, \mathbf{u}, \boldsymbol{\beta}, g)dt},$$

where $f_{Y|\mathbf{X},R=1}(y, \mathbf{x})$ is the conditional distribution of Y given $\mathbf{X}$ and $R = 1$, it can be rewritten as

$$\begin{aligned} & f_{\mathbf{X},R,RY}(\mathbf{x}, r, ry) \qquad (5) \\ = \; & f_{\mathbf{X}}(\mathbf{x}) \left\{ \frac{f_{Y|\mathbf{X},R=1}(y, \mathbf{x})}{\int f_{Y|\mathbf{X},R=1}(t, \mathbf{x})/\pi(t, \mathbf{u}, \boldsymbol{\beta}, g)dt} \right\}^r \left\{ 1 - \frac{1}{\int f_{Y|\mathbf{X},R=1}(t, \mathbf{x})/\pi(t, \mathbf{u}, \boldsymbol{\beta}, g)dt} \right\}^{1-r}. \end{aligned}$$

Clearly, model (5) is much more complicated than the model (2). To estimate $E\{\boldsymbol{\zeta}(Y)\}$ under model (5), one has to first estimate both the parameter $\boldsymbol{\beta}$ and the nonparametric components $f_{\mathbf{X}}(\cdot)$, $f_{Y|\mathbf{X},R=1}(\cdot)$ and $g(\mathbf{u})$. While the estimation of $f_{\mathbf{X}}(\cdot)$ and $f_{Y|\mathbf{X},R=1}(\cdot)$ does not involve missing data and can be done by using various off-the-shelf methods, the estimation of $g(\mathbf{u})$ might not be straightforward. How to estimate $g(\mathbf{u})$ turns out to be pivotal as we learn from this nonignorable propensity score model.

3.1 Efficient Estimation of $\boldsymbol{\beta}$ and $E\{\boldsymbol{\zeta}(Y)\}$: Feasible?

Under assumption (4), Shetty et al. (2022) showed that the efficient score for estimating $\boldsymbol{\beta}$ is

$$\mathbf{S}_{\text{eff}}(\mathbf{x}, r, ry) = \mathbf{g}(\mathbf{x})[1 - r\{1 + e^{-g(\mathbf{u})-h(y,\boldsymbol{\beta})}\}], \qquad (6)$$

where

$$\begin{aligned} \mathbf{g}(\mathbf{x}) &= \frac{\mathbf{a}(\mathbf{u})E\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\} - E\{e^{-h(Y,\boldsymbol{\beta})}\mathbf{h}'_{\boldsymbol{\beta}}(Y, \boldsymbol{\beta}) \mid \mathbf{x}, 1\}}{E\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\} + e^{-g(\mathbf{u})}E\{e^{-2h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}}, \\ \mathbf{a}(\mathbf{u}) &= \frac{E\left[E\{e^{-h(Y,\boldsymbol{\beta})}\mathbf{h}'_{\boldsymbol{\beta}}(Y, \boldsymbol{\beta}) \mid \mathbf{x}, 1\}E\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}/d(\mathbf{x}) \mid \mathbf{u}, 1\right]}{E\left([E\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}]^2/d(\mathbf{x}) \mid \mathbf{u}, 1\right)}, \text{ and} \\ d(\mathbf{x}) &= E\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\} + e^{-g(\mathbf{u})}E\{e^{-2h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}. \end{aligned}$$

Although it is specific, the efficient score $\mathbf{S}_{\text{eff}}(\mathbf{x}, r, ry)$ has a very complicated form and its implementation is not straightforward. Essentially, an estimate of $g(\mathbf{u})$ is needed. To pursue

the efficient estimation of $\boldsymbol{\beta}$, one might have to first use some approximation technique to locate an appropriate estimator for $g(\mathbf{u})$.

Further, to achieve the efficient estimation of $E\{\boldsymbol{\zeta}(Y)\}$, in the supplementary material of this paper, we derive the efficient influence function for estimating $E\{\boldsymbol{\zeta}(Y)\}$ as

$$\begin{aligned}\boldsymbol{\phi}_{\text{eff}}(\mathbf{x}, r, ry) &= \frac{r}{\pi(y,\mathbf{u})}\left[\boldsymbol{\zeta}(y) - \mathbf{b}_3(\mathbf{x}) + \mathbf{c}_3(\mathbf{u})\frac{1-w(\mathbf{x})}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1}\right] \\ &\quad -E\{\boldsymbol{\zeta}(Y)\} + \mathbf{b}_3(\mathbf{x}) - \mathbf{c}_3(\mathbf{u})\frac{1-w(\mathbf{x})}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1} \\ &\quad +\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{S}_{\text{eff}}(\mathbf{x}, r, ry), \end{aligned} \tag{7}$$

where

$$\begin{aligned}\mathbf{b}_3(\mathbf{x}) &= \frac{E\{\boldsymbol{\zeta}(Y)\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1}, \\ \mathbf{c}_3(\mathbf{u}) &= E\left(\left[\frac{1-w(\mathbf{x})}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1} - \pi(Y,\mathbf{u})\right]\{\pi^{-1}(Y,\mathbf{u}) - 1\}\boldsymbol{\zeta}(Y) \mid \mathbf{u}\right) d_3(\mathbf{u})^{-1}, \\ d_3(\mathbf{u}) &= E\left(\frac{\{1-w(\mathbf{x})\}^2}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1} \mid \mathbf{u}\right), \\ \mathbf{M}_1 &= E\left\{E[\mathbf{g}(\mathbf{x})\{1-w(\mathbf{x})\} \mid \mathbf{u}]\frac{\{1-w(\mathbf{x})\}\{1-\pi(Y,\mathbf{u})\}\mathbf{h}'_{\boldsymbol{\beta}}{}^{\mathrm{T}}(Y)}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1} d_3(\mathbf{u})^{-1}\right. \\ &\quad \left. -\mathbf{g}(\mathbf{x})\{1-\pi(Y,\mathbf{u})\}\mathbf{h}'_{\boldsymbol{\beta}}{}^{\mathrm{T}}(Y)\right\}, \\ \mathbf{M}_2 &= E\left([\mathbf{b}_3(\mathbf{X}) - \boldsymbol{\zeta}(Y)]\mathbf{h}'_{\boldsymbol{\beta}}(Y,\boldsymbol{\beta})^{\mathrm{T}}\{1-\pi(Y,\mathbf{u})\}\right) \\ &\quad -E\left[E\left(\{1-w(\mathbf{x})\}\mathbf{b}_3(\mathbf{x}) - \{1-w(\mathbf{x})\}E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}\right.\right. \\ &\quad \left. +\frac{1-w(\mathbf{x})}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1}\mathbf{a}_3(\mathbf{u})\left[w(\mathbf{x})E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1\right] \mid \mathbf{u}\right) \\ &\quad \left.\frac{\{1-w(\mathbf{x})\}\{1-\pi(Y,\mathbf{u})\}\mathbf{h}'_{\boldsymbol{\beta}}(Y)^{\mathrm{T}}}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1} d_3(\mathbf{u})^{-1}\right] \\ &\quad -E\left\{\frac{\{1-w(\mathbf{x})\}\{1-\pi(Y,\mathbf{u})\}}{E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}\} - 1}\mathbf{a}_3(\mathbf{U})\mathbf{h}'_{\boldsymbol{\beta}}(Y)^{\mathrm{T}}\right\}, \\ \mathbf{a}_3(\mathbf{u}) &= \frac{E[\{\pi(Y,\mathbf{u}) - w(\mathbf{x})\}\boldsymbol{\zeta}(Y) \mid \mathbf{u}]}{E\{w(\mathbf{x}) - w^2(\mathbf{x}) \mid \mathbf{u}\}}. \end{aligned}$$

Note that, in the efficient influence function $\boldsymbol{\phi}_{\text{eff}}(\mathbf{x}, r, ry)$ in (7), $\mathbf{S}_{\text{eff}}(\mathbf{x}, r, ry)$ is the efficient score for $\boldsymbol{\beta}$ given in (6). Also, $w(\mathbf{x})$ is defined as $w(\mathbf{x}) = \left[E\{\pi^{-1}(Y,\mathbf{u}) \mid \mathbf{x}, 1\}\right]^{-1}$. If the true value of $\boldsymbol{\beta}$ is zero hence the model (4) becomes ignorable, it can be easily checked that $\mathbf{b}_3(\mathbf{x}) = E\{\boldsymbol{\zeta}(Y) \mid \mathbf{x}\}$, $\mathbf{c}_3(\mathbf{u}) = \mathbf{M}_2 = \mathbf{0}$ and the efficient influence function (7) becomes the one in (3) in Section 2.

The expression of the efficient influence function (7), while explicit, is very complex. This complexity is the result of the structure of the problem setting itself. The terms involve the unknown non-parametric function $g(\mathbf{u})$. In order to achieve efficient estimation of $E\{\boldsymbol{\zeta}(Y)\}$ one would need to further plug in the estimators for $\boldsymbol{\beta}$, $g(\cdot)$, which will make

the above equation more convoluted. In addition, the expectations in various expressions are conditional on $\mathbf{x}$, $E(\cdot \mid \mathbf{x})$, and estimating them involve first estimating conditional expectations given $\mathbf{x}, r = 1$, $E(\cdot \mid \mathbf{x}, r = 1)$, and converting them to $E(\cdot \mid \mathbf{x})$ via $\widehat{E}(\cdot \mid \mathbf{x}) = \widehat{w}(\mathbf{x})\widehat{E}\{\cdot \times \pi^{-1}(Y, \mathbf{u}) \mid \mathbf{x}, r = 1\}$. In obtaining all these expectations, to avoid possible model misspecification so that efficiency can be guaranteed, non-parametric techniques need to be used, and the nested structure of the conditional expectations will further complicate the implementation of the estimator. These obstacles, in addition to the large number of terms, will make the implementation of the efficient estimator very computationally expensive.

3.2 Robust Estimation of $\boldsymbol{\beta}$ and $E\{\boldsymbol{\zeta}(Y)\}$: Feasible

Shetty et al. (2022) discovered that the efficient score function in (6) has mean zero property even when the unknown function $g(\mathbf{u})$ is replaced by a working model $g^*(\mathbf{u})$, due to the special form of the propensity score. This leads to the estimating equation $\sum_{i=1}^N \widehat{\mathbf{S}}^*_{\text{eff}}(\mathbf{x}_i, r_i, r_i y_i) = \mathbf{0}$, i.e.,

$$\sum_{i=1}^{N} \frac{\widehat{\mathbf{a}}^*(\mathbf{u}_i)\widehat{E}\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}_i, 1\} - \widehat{E}\{e^{-h(Y,\boldsymbol{\beta})}\mathbf{h}'_{\boldsymbol{\beta}}(Y, \boldsymbol{\beta}) \mid \mathbf{x}_i, 1\}}{\widehat{E}\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}_i, 1\} + e^{-g^*(\mathbf{u}_i)}\widehat{E}\{e^{-2h(Y,\boldsymbol{\beta})} \mid \mathbf{x}_i, 1\}}[1 - r_i\{1 + e^{-g^*(\mathbf{u}_i)-h(y_i,\boldsymbol{\beta})}\}] = \mathbf{0},$$

where at any $\mathbf{u}, \mathbf{x}$,

$$\begin{aligned}
\widehat{\mathbf{a}}^*(\mathbf{u}) &= \frac{\widehat{E}\left[\widehat{E}\{e^{-h(Y,\boldsymbol{\beta})}\mathbf{h}'_{\boldsymbol{\beta}}(Y, \boldsymbol{\beta}) \mid \mathbf{x}, 1\}\widehat{E}\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}/\widehat{d}^*(\mathbf{x}) \mid \mathbf{u}, 1\right]}{\widehat{E}\left([\widehat{E}\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}]^2/\widehat{d}^*(\mathbf{x}) \mid \mathbf{u}, 1\right)}, \\
\widehat{d}^*(\mathbf{x}) &= \widehat{E}\{e^{-h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\} + e^{-g^*(\mathbf{u})}\widehat{E}\{e^{-2h(Y,\boldsymbol{\beta})} \mid \mathbf{x}, 1\}.
\end{aligned}$$

Under the misspecified $g^*(\mathbf{u})$, solving the above estimating equation still leads to a consistent estimator for $\boldsymbol{\beta}$, which reveals a robustness property of the procedure. The conditional expectations in the above equations, which are functions of fully observed data, were estimated using parametric or nonparametric methods.

Next, they proposed a robust estimator for the quantity of interest, $E\{\zeta(Y)\}$, using the fact that $E\{\boldsymbol{\zeta}(Y)\} = E\{\boldsymbol{\zeta}(Y) \mid R = 1\}\text{pr}(R = 1) + E\{\boldsymbol{\zeta}(Y) \mid R = 0\}\text{pr}(R = 0)$. The first term $E\{\boldsymbol{\zeta}(Y) \mid R = 1\}\text{pr}(R = 1)$ only depends on the observed data, hence is proposed to be estimated by $N^{-1}\sum_{i=1}^N r_i\boldsymbol{\zeta}(y_i)$. Further, $E\{\boldsymbol{\zeta}(Y) \mid R = 0\} = \int E\{\boldsymbol{\zeta}(Y) \mid R = 0, \mathbf{x}\}f_{\mathbf{X}|R=0}(\mathbf{x})d\mathbf{x}$ and

$$E\{\boldsymbol{\zeta}(Y) \mid R = 0, \mathbf{x}\} = \frac{E\{\boldsymbol{\zeta}(Y)(\pi^{-1} - 1) \mid \mathbf{x}, 1\}}{E\{(\pi^{-1} - 1) \mid \mathbf{x}, 1\}} = \frac{E[\boldsymbol{\zeta}(Y)\exp\{-h(Y, \boldsymbol{\beta})\} \mid \mathbf{x}, 1]}{E[\exp\{-h(Y, \boldsymbol{\beta})\} \mid \mathbf{x}, 1]}.$$

Hence $E\{\boldsymbol{\zeta}(Y) \mid R = 0)\text{pr}(R = 0)$ can be estimated by $\frac{1}{N}\sum_{i=1}^N (1-r_i)\frac{E[\boldsymbol{\zeta}(Y)\exp\{-h(Y,\boldsymbol{\beta})\}|\mathbf{x}_i,1]}{E[\exp\{-h(Y,\boldsymbol{\beta})\}|\mathbf{x}_i,1]}$. Therefore they estimated $E\{\boldsymbol{\zeta}(Y)\}$ by

$$\frac{1}{N}\sum_{i=1}^{N}\left(r_i\boldsymbol{\zeta}(y_i) + (1 - r_i)\frac{\widehat{E}[\boldsymbol{\zeta}(Y)\exp\{-h(Y, \boldsymbol{\beta})\} \mid \mathbf{x}_i, 1]}{\widehat{E}[\exp\{-h(Y, \boldsymbol{\beta})\} \mid \mathbf{x}_i, 1]}\right).$$

In summary, the estimation procedures proposed in Shetty et al. (2022) relied on the efficient score for $\boldsymbol{\beta}$; however, they do not require the estimation or modeling of $g(\mathbf{u})$. They used a working model $g^*(\mathbf{u})$ and showed that the resulting estimator is still consistent even if the working model is misspecified.

4. Discussion

Our research mainly generalizes the assumption on the propensity score from ignorable to nonignorable. This is practically important, especially in analyzing data with missing values in various biomedical studies. Our main finding is that, the nice and clean property of double robustness and semiparametric efficiency cannot be straightforwardly extended from ignorability to nonignorability. In general, it is a lot more sophisticated and mathematically involved. In the nonignorable propensity score we analyze in this work, we advocate, instead of pursuing the efficient estimation of $\boldsymbol{\beta}$ and $E\{\boldsymbol{\zeta}(Y)\}$, one should consider the robust estimation which, albeit less efficient, is empirically easier and more feasible.

Acknowledgments

This research was partially supported by the National Science Foundation (2122074) of the United States.

References

Olawale F Ayilara, Lisa Zhang, Tolulope T Sajobi, Richard Sawatzky, Eric Bohm, and Lisa M Lix. Impact of missing data on bias and precision when estimating change in patient-reported outcomes from a clinical registry. *Health and Quality of Life Outcomes*, 17(1):1–9, 2019.

Peter J Bickel, J Klaassen, YA'Acov Ritov, and Jon A Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press Baltimore, 1993.

Giulia Carreras, Guido Miccinesi, Andrew Wilcock, Nancy Preston, Daan Nieboer, Luc Deliens, Mogensm Groenvold, Urska Lunder, Agnes van der Heide, and Michela Baccini. Missing not at random in end of life care studies: multiple imputation and sensitivity analysis on data from the action study. *BMC Medical Research Methodology*, 21(1):1–12, 2021.

Martin R Frankel, Michael P Battaglia, Lina Balluz, and Tara Strine. When data are not missing at random: implications for measuring health conditions in the behavioral risk factor surveillance system. *BMJ Open*, 2(4):e000696, 2012.

Simon B Goldberg, Daniel M Bolt, and Richard J Davidson. Data missing not at random in mobile health research: assessment of the problem and a case for sensitivity analyses. *Journal of Medical Internet Research*, 23(6):e26749, 2021.

Manuel Gomes, Nils Gutacker, Chris Bojke, and Andrew Street. Addressing missing data in patient-reported outcome measures (proms): Implications for the use of proms for comparing provider performance. *Health Economics*, 25(5):515–528, 2016.

Rolf HH Groenwold. Informative missingness in electronic health record systems: the curse of knowing. *Diagnostic and Prognostic Research*, 4(1):1–6, 2020.

Guido W Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1):4–29, 2004.

Guido W Imbens and Donald B Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

David K Lim, Naim U Rashid, Junier B Oliva, and Joseph G Ibrahim. Handling non-ignorably missing features in electronic health records data using importance-weighted autoencoders. *arXiv preprint arXiv:2101.07357*, 2021.

Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data*, volume 793. John Wiley & Sons, 2019.

Rebecca Mercieca-Bebber, Michael J Palmer, Michael Brundage, Melanie Calvert, Martin R Stockler, and Madeleine T King. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (pro) data: a systematic review. *BMJ Open*, 6(6):e010938, 2016.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

Andrea Rotnitzky and James Robins. Analysis of semi-parametric regression models with non-ignorable non-response. *Statistics in Medicine*, 16:81–102, 1997.

Jun Shao and Lei Wang. Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika*, 103(1):175–187, 2016.

Samidha Shetty, Yanyuan Ma, and Jiwei Zhao. Avoid estimating the unknown function in a semiparametric nonignorable propensity model. *arXiv preprint arXiv:2108.04966*, 2022.

Anastasios A Tsiatis. *Semiparametric Theory and Missing Data*. New York: Springer, 2006.