University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2006 *Paper* 211

Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting from Violation of Experimental Treatment Assignment

Yue Wang* Maya L. Petersen†

David Bangsberg‡ Mark J. van der Laan**

*Division of Biostatistics, School of Public Health, University of California, Berkeley, ywang@stat.berkeley.edu

†Division of Biostatistics, School of Public Health, University of California, Berkeley, mayaliv@berkeley.edu

‡Epidemiology & Prevention Interventions Center, University of California, San Francisco

**Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting from Violation of Experimental Treatment Assignment

Yue Wang, Maya L. Petersen, David Bangsberg, and Mark J. van der Laan

Abstract

Inverse probability of treatment weighting (IPTW) is frequently used to estimate the causal effects of treatments and interventions. The consistency of the IPTW estimator relies not only on the well-recognized assumption of no unmeasured confounders (Sequential Randomization Assumption or SRA), but also on the assumption of experimentation in the assignment of treatment (Experimental Treatment Assignment or ETA). In finite samples, violations in the ETA assumption can occur due simply to chance; certain treatments become rare or non-existent for certain strata of the population. Such practical violations of the ETA assumption occur frequently in real data, and can result in significant bias in the IPTW estimator of causal effects. This manuscript presents a diagnostic tool for assessing the bias in the IPTW estimator due to violation of the ETA assumption. The Diagnostic of ETA Bias (DEB), implemented in a public R routine, relies on parametric bootstrap sampling from an estimated data-generating distribution. The article presents results of simulations to assess the performance and applications of the diagnostic and to investigate the extent of ETA bias in a range of contexts. In addition, results are presented from two data examples drawn from the treatment of HIV infection, in which DEB is used to assess the ETA bias of the IPTW estimator.

1 Introduction.

Estimation of the causal effects of treatments or interventions using observational data is a major focus of many statistical applications, including much of epidemiology and clinical research. Relying on the counterfactual framework, seminal work by Robins (Robins et al. (2000),Robins (1999),Robins (2000)) introduced marginal structural models (MSM) as a powerful tool for the estimation of causal effects. This work, further developed in van der Laan and Robins (2002), presents three marginal structural model estimators: the G-computation estimator, the Inverse Probability of Treatment Weighted (IPTW) estimator, and the Double-Robust (DR) estimator. Perhaps due in part to its straightforward implementation, the IPTW estimator, in particular, is rapidly becoming a standard tool in the analysis of clinical and epidemiological data.

The consistency of the IPTW estimator relies heavily on the assumption of experimental treatment assignment (ETA). The ETA assumption states that each possible treatment or intervention occurs with some positive probability, regardless of a subject's past. When the ETA assumption is violated, the IPTW estimator is undefined, while the G-computation and DR estimators rely on extrapolation. Importantly, recent work by Neugebauer and van der Laan (2005) has shown that even practical violations of the ETA assumption can lead to substantial bias in the IPTW estimator. Such violations occur commonly in real-life data applications; whenever a given treatment occurs with very low probability for a given stratum of subjects, whether because the subjects' covariates make assignment of the treatment unlikely, or due simply to chance, substantial bias to the IPTW estimator can result.

Despite the demonstrated potential for practical ETA violations to bias the IPTW estimator, real data applications rarely investigate this potential source of bias. One reason for the relative neglect of the ETA assumption in data applications may be the lack of an accessible diagnostic tool for estimating the extent of ETA bias. In this paper we present a straightforward diagnostic tool that provides quantitative estimates of the bias in the IPTW estimator due to violations of the ETA assumption. We refer to this new tool as the "Diagnostic for ETA Bias" (DEB). Under the assumption that the data-generating distribution and treatment mechanism (i.e. the distribution of treatment assignment given a subject's observed past) are correctly specified, DEB provides an estimate of the extent of relative and absolute bias in the IPTW estimator due to ETA violations. The diagnostic presented thus provides additional quantitative output to complement a standard IPTW analysis. If DEB suggests that ETA violations are resulting in major bias in the IPTW estimator, then the investigator should consider implementing alternative estimators.

In addition to the ETA assumption, and the assumption of no unmeasured confounders, or sequential randomization, required by all MSM estimators, the consistency of the IPTW estimator further relies on consistent estimation of the treatment mechanism. However, even assuming the true treatment mechanism is known or

estimated correctly, in practice achieving an IPTW estimator with minimal mean squared error (MSE) can require use of an incorrectly specified treatment mechanism. Consider the following two common examples. 1) The presence in the dataset of individuals with very low probability of receiving their observed treatments given their covariates can result extreme values in estimated weights, increasing the variability of the resulting IPTW estimate. In this setting, truncating extreme values of the weights provides a means of reducing variability, at a cost to bias stemming from the resulting misspecification of the treatment mechanism. 2) Treatment assignment may be heavily influenced by a covariate W that is a very weak confounder of the causal effect of interest. In this case, while the true treatment mechanism clearly includes W, inclusion of W in the treatment mechanism used to estimate the weights can lead to a significant increase in bias due to practical ETA violations, while excluding W will result in only minimal increase in bias due to uncontrolled confounding. In addition to providing an assessment of the extent of ETA bias, the diagnostic tool presented here provides a way of quantifying the effects of choices such as these on the bias-variance tradeoff of the IPTW estimator.

The paper is structured as follows. In Section 2 we review the counterfactual framework for inferring the causal effect of a point treatment. The data structure and model are introduced, including formal definition of the ETA and other underlying assumptions. The three marginal structural model estimators, IPTW, G-computation, and DR, are reviewed. In Section 3 we review how ETA violations can result in bias in the IPTW estimator. We then present our diagnostic tool DEB for estimating the extent of this bias. Briefly, the approach involves drawing parametric bootstrap samples from the estimated data-generating distribution. The G-computation estimator for this known data-generating distribution is consistent, thus deviation between the G-computation estimator and the IPTW estimator applied to the bootstrap samples reflects the extent of bias due to ETA violations (assuming consistent estimation of the treatment mechanism). The R-code for this diagnostic is made publicly available as the routine `bias.ETA()`. In Section 4 we provide the results of several simulations that demonstrate the validity of the diagnostic tool and its performance under different conditions. Simulations are further used to demonstrate how the output of DEB can be used to quantify the effects of choices regarding truncation of weights and the inclusion of covariates in the treatment mechanism. Section 5 describes the application of the diagnostic to two real data examples. Both examples focus on interventions to improve patient adherence to antiretroviral medications. For pedagogical purposes, the manuscript focuses on causal inference for point treatments; however, the approach can be readily applied as a diagnostic of ETA bias in the IPTW estimators of longitudinal treatment effects. Finally, we discuss the implications of our new diagnostic tool, and areas where further research is needed.

2 Review of Causal Inference in the Point Treatment Setting.

In this section, we review the counterfactual framework for causal inference, focusing on the point treatment setting. Three estimators of the marginal structural model parameter of interest are reviewed: IPTW, G-computation, and DR.

2.1 The counterfactual framework for causal inference.

Let $X = ((Y(a), a \in \mathcal{A}), W) \sim F_{X0}$ be the full data structure of interest on a randomly sampled subject, where W denotes baseline covariates, $Y(a)$ denotes the outcome on a subject if the subject would have taken treatment a, and $\mathcal{A}$ denotes the set of possible treatments. Such potential outcomes $Y(a)$ are called counterfactuals (e.g. Rubin (1978)). Let A be a random variable with conditional probability distribution $g_0(a \mid X) \equiv P(A = a \mid X)$, $a \in \mathcal{A}$, which denotes the treatment the subject actually took. For each subject, we will only observe the outcome indexed by the treatment the subject took. Thus we observe n i.i.d. observations $O_1, \ldots, O_n$ of the observed data structure

$$O = (A, Y \equiv Y(A), W),$$

where Y denotes the observed outcome corresponding with the treatment taken by the subject. Note that the observed data structure is a missing data structure in which the full data is X, and the missingness variable is A. Consequently, we have that the observed data distribution $P_0 = P_{F_{X0},g_0}$ is indexed by the full data distribution F_{X0} and the conditional density g_0.

In order to make this full data parameter identifiable, we will assume the randomization assumption (RA) and the experimental treatment assignment assumption (ETA) on the treatment mechanism $g_0(A \mid X)$. The randomization assumption states that treatment is randomized within strata of W:

$$g_0(a \mid X) = g_0(a \mid W) \text{ for all } a \in \mathcal{A}.$$

The randomization assumption corresponds with assuming coarsening at random or missing at random in our missing data model (see van der Laan and Robins (2002)).

The experimental treatment assignment assumption states that within each strata of W, each possible treatment has positive probability:

$$\min_{a \in \mathcal{A}} g_0(a \mid W) > 0, \ F_{0W}\text{-a.e.}$$

When stabilized weights are used (discussed below), it is sufficient to assume a V-specific weaker version of the experimental treatment assignment assumption stating that there exists a conditional density $g(\cdot \mid V)$ such that

$$\sup_{a \in \mathcal{A}} \frac{g(a \mid V)}{g(a \mid W)} < \infty, \ F_{0W}\text{-a.e.}$$

Under these two assumptions RA and ETA, the density of O factorizes into a F_X-part and a g-part:

$$p(O) = p(W)p(Y \mid A, W)g(A \mid W) = p(W)\, p(Y(a) \mid W)|_{a=A}\, g(A \mid X).$$

The first two terms correspond with the F_X-part of the density of O, and the last term is the treatment mechanism.

The marginal causal effect of treatment a on the counterfactual outcome Y_a adjusted for V is defined by $E_{F_{X0}}(Y_a|V)$; a marginal structural model can be used to model this effect:

$$E_{F_{X0}}(Y_a|V) = m(a, V|\beta_0),$$

where the parameter of interest is the true causal parameter β_0.

2.2 Estimation of the causal parameter.

There are three well-established estimators of the causal parameter β_0. The traditional approach to estimate β_0 uses the G-computation formula from Robins (1986) and Robins (1987). This estimate relies on the association model used for $E_{F_{X0}}(Y|A, W)$; if the association model is misspecified, the G-computation estimate of β_0 will be inconsistent. Another widely-used methodology is the Inverse Probability of Treatment Weighted (IPTW) estimator, which relies on the treatment model $g(A|W)$; if the treatment model is misspecified, and/or the experimental treatment assignment (ETA) assumption is violated, the IPTW estimate of β_0 will be inconsistent. The Double Robust estimate is the solution of the Double Robust estimating equation. The DR estimate requires assuming a treatment model $g(A|W)$ and a regression model $E(Y|A, W)$, but only one of these models needs to be correctly specified in order to obtain a consistent estimate of β_0. The estimation procedures of the three estimators are briefly reviewed below.

2.2.1 G-computation estimator.

The G-computation estimate of β_0 is based on the likelihood function of the data and the assumption that the association between Y and $\{A, W\}$ is described correctly by a regression model $E(Y|A, W) = Q(A, W)$. The estimate takes two steps:

1. Fit the association model $\hat{Q}(A, W)$ by regressing Y on A and W, and for each subject, calculate $\hat{Y}_{a,W} = \hat{Q}(a, W)$ for every $a \in \mathcal{A}$;

2. Estimate the G-computation estimator by regressing $\hat{Y}_{a,W}$ on a and V for the MSM.

The association model of $E(Y|A, W)$ needs to be correctly specified for the G-computation estimator to be consistent.

2.2.2 IPTW estimator.

The IPTW estimate of β_0 is the solution of the IPTW estimating equation for the following estimating function:

$$D_h(O|\beta, g) = \frac{h(A,V)\varepsilon(\beta)}{g(A|W)},$$

where $\varepsilon(\beta) = Y - m(A,V|\beta)$, $g(A|W)$ is the treatment model, and h is a function of A and V. This estimating function is unbiased at β_0 if the treatment model $g(A|W)$ is correctly specified and ETA assumption holds, i.e., $E_0[D_h(O|\beta, g)] = 0$ if $g = g_0$ and ETA holds.

The IPTW estimator can be easily obtained by regressing Y on A and V with weights using any statistical software routines. Several weighting options are available. The standard, or non-stabilized weights simply use $h(A,V) = \frac{d}{d\beta}m(A,V|\beta)$ and thus the weights $w_i = 1/g(A_i|W_i)$ for each subject. The "stabilized weights" (Robins et al. (2000)) use $h(A,V) = g(A|V)\frac{d}{d\beta}m(A,V|\beta)$, where $g(A|V)$ is a conditional probability of A on V, and $w_i = g(A_i|V_i)/g(A_i|W_i)$. As noted above, use of the stabilized weights allows a weaker form of the ETA assumption.

2.2.3 Double Robust estimator.

The DR estimate of β_0 is the solution of the DR estimating equation for the following estimating function:

$$\begin{aligned} D_h(O|\beta, g, Q) &= \frac{h(A,V)\varepsilon(\beta)}{g(A|W)} - \frac{h(A,V)}{g(A|W)}[Q(A,W) - m(A,V|\beta)] \\ &\quad + \sum_{a\in\mathcal{A}} h(a,V)[Q(a,W) - m(a,V|\beta)]. \end{aligned}$$

This estimating function is unbiased as β_0 if either $Q(A,W)$ is correctly specified or $g(A|W)$ is correctly specified with no violation of ETA, i.e., $E_0[D_h(O|\beta, g, Q)] = 0$ if $Q = Q_0$ or $g = g_0$ and ETA holds.

The general solution of the DR estimating equation can be searched with Newton-Raphson algorithm. That is,

$$\hat{\beta}^{k+1} = \hat{\beta}^k + C^{-1}(\hat{\beta}^k)\frac{1}{n}\sum_{i=1}^{n} D_h(O_i|g_n, Q_n, \hat{\beta}^k),$$

where $C(\hat{\beta}^k)$ is a square matrix,

$$\begin{aligned} C(\hat{\beta}^k) &= -\frac{d}{d\beta}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{h(A_i,V_i)}{g(A_i|W_i)}\varepsilon(\hat{\beta}^k)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{h(A_i,V_i)}{g(A_i|W_i)}\frac{d}{d\beta}m(A_i,V_i|\hat{\beta}^k). \end{aligned}$$

For continuous outcome Y, the DR estimate has a closed form solution for linear functions of $m(A, V|\beta)$. Let the p-dimension linear function $m(A, V|\beta) = Z\beta$, where Z is the design matrix of A and V with dimension $n \times p$. The DR estimating equation can be written as

$$Z^T[(Y - Q(A, W))\frac{g(A|V)}{g(A|W)}] + \sum_{a \in \mathcal{A}} Z_a^T[Q(a, W)g(a|V)] - \sum_{a \in \mathcal{A}} Z_a^T[g(a|V)Z_a]\beta = 0.$$

The DR estimate is $\hat{\beta} = C^{-1}C_0$, where

$$C_0 = Z^T[(Y - Q(A, W))\frac{g(A|V)}{g(A|W)}] + \sum_{a \in \mathcal{A}} Z_a^T[Q(a, W)g(a|V)],$$

and

$$C = \sum_{a \in \mathcal{A}} Z_a^T[g(a|V)Z_a].$$

3 Bias of the IPTW Estimator Due to Violation of ETA Assumption.

Bias in the IPTW estimator can arise due to several causes. The first is misspecification of the marginal structural model itself. However, following the lead of Neugebauer and van der Laan (2006), we simply consider our causal parameter of interest β to be the projection of the true causal parameter onto our model $m(a, V|\beta)$. Second, misspecification of the treatment mechanism ($g(A|W)$) can lead to bias in the IPTW estimator. However, throughout this article, we assume that the model of the treatment mechanism is correctly specified. While this may seem counterintuitive, recall that the goal of our diagnostic tool is not to assess all bias in the IPTW estimator, but rather to assess the extent to which, given a correctly specified treatment mechanism, the IPTW estimator remains biased due to ETA violations.

The bias in the IPTW estimator is defined as the difference between the true causal parameter of interest and the expectation (under the true data generating distribution) of the IPTW estimator applied to a finite sample. Under the assumption that the treatment mechanism is correctly specified, this bias is predominantly composed of bias due to ETA violations (with some additional bias due to the finite sample estimate $g_{P_n}(A|W)$ of the true treatment mechanism ($g_{P_0}(A|W)$)). In this paper then, we call the bias in the IPTW estimator, under the assumption that the $g_{P_n}(A|W)$ is correctly specified, $ETA.Bias$:

$$ETA.Bias(\hat{\Psi}_{IPTW}, P_0, n) = E_{P_0}\hat{\Psi}_{IPTW}(P_n) - \Psi(P_0),$$

where $\Psi(P_0)$ is the causal parameter of interest, $P_n \rightarrow \hat{\Psi}_{IPTW}(P_n)$ is the IPTW estimator, and P_n denotes the empirical distribution of a sample of n i.i.d observations

from the true observed data distribution P_0. That is, $ETA.Bias$ is the bias of the IPTW estimator $\hat{\Psi}_{IPTW}$ applied to a sample of n i.i.d. observations (i.e, P_n) from P_0.

This bias can be decomposed as

$$\begin{aligned} ETA.Bias(\hat{\Psi}_{IPTW}, P_0, n) &= (\hat{\Psi}_{IPTW}(P_0) - \Psi(P_0)) \\ &\quad +(E_{P_0}\hat{\Psi}_{IPTW}(P_n) - \hat{\Psi}_{IPTW}(P_0)). \end{aligned}$$

The first part of the bias, $\hat{\Psi}_{IPTW}(P_0) - \Psi(P_0)$, is due to theoretical violation of the ETA. In other words, this quantity will contribute to the overall bias whenever certain subjects have zero probability of receiving a given treatment. For example, if W includes a covariate that is an absolute contraindication for receiving a treatment of interest (for example, A is a teratogenic drug and thus is never given to pregnant women), then for some W in the population P_0, the treatment probability $g(a|W)$ will always equal 0. Such a situation arises frequently in the context of clinical treatments.

The second part of the bias $E_{P_0}\hat{\Psi}_{IPTW}(P_n) - \hat{\Psi}_{IPTW}(P_0)$ comes from the finite sample size of the empirical population P_n. Much of the finite sample bias is due to practical violations of the ETA assumption. Under some circumstances, the treatment probability $g(a|W)$ is theoretically positive for any $a \in \mathcal{A}$. However, as the sample size decreases, the probability of receiving a given treatment may approach zero for certain subjects (strata of W) and a practical violation of the ETA assumption arise. This is due to the fact that, while with large sample size each possible treatment may be represented among every group of patients, as sample size shrinks, by chance certain treatments may no longer occur in some groups of patients. For example, if only five Hispanic women are present in the sample, it can easily occur by chance that none of the five receives a given treatment.

3.1 Estimating bias due to ETA violations using the bootstrap.

Since the true observed data distribution P_0 is unknown in the real world, our approach to estimating $ETA.Bias$ relies on an estimate $\hat{P}_0$ of P_0. To calculate this estimate, we employ a parametric bootstrap. That is, our diagnostic DEB is defined as:

$$E_{\hat{P}_0}\hat{\Psi}_{IPTW}(P_n^{\#}) - \Psi(\hat{P}_0),$$

where $P_n^{\#}$ is the empirical distribution of the bootstrap sample obtained from sampling $\hat{P}_0$. Implementation of DEB involves a parametric bootstrap in two steps.

Step 1. Estimating $\hat{P}_0$ of P_0.

Given the empirical distribution P_n, to estimate $\hat{P}_0$ of P_0 is to estimate from the observed dataset an association model $Q_{\hat{P}_0}(A, W)$ for $E_0(Y|A, W)$, a treatment model

$g_{\hat{P}_0}(A|W)$ for $g_0(A|W)$, and $P(W)$. If we define $Q_{\hat{P}_0}(A,W) = Q_{P_n}(A,W)$ and $g_{\hat{P}_0}(A|W) = g_{P_n}(A|W)$, then the true causal parameter of interest for the known distribution $\hat{P}_0$ is the same as the G-computation estimator of the observed data:

$$\Psi(\hat{P}_0) = \hat{\Psi}_{Gcomp}(P_n).$$

For the IPTW estimator, the above statement can only be correct when the ETA assumption holds for $g_{P_n}(A|W)$.

We treat the model of the treatment mechanism as known (in fact it was likely estimated in the process of implementing the IPTW estimator). $P(W)$ can be estimated empirically from the observed data. In order for DEB to perform well as a diagnostic of ETA bias, it remains to do a good job estimating $Q_{P_n}(A,W)$. Use of machine-learning algorithms and cross validation can help to achieve this goal.

Step 2. Generating $P_n^{\#}$ by sampling from $\hat{P}_0$.

In the second step, we assume $\hat{P}_0$ is the true data generating distribution. Bootstrap samples $P_n^{\#}$, each with n i.i.d observations, are generated by sampling from $\hat{P}_0$.

Throughout the bootstrap simulation, we use $Q_{\hat{P}_0}(A,W)$ and $g_{\hat{P}_0}(A|W)$ as the data generation models, and calculate the IPTW estimator of each sample based on an estimate $\hat{g}_{\hat{P}_0}(A|W)$. The whole sampling distribution of the bootstrap IPTW estimator $\hat{\Psi}_{IPTW}(P_n^{\#})$ reflects the behavior of $\hat{\Psi}_{IPTW}(P_n^{\#})$ under $\hat{P}_0$. The estimate $ETA.Bias$ is then obtained by comparing the bootstrap IPTW estimator with the true causal parameter under the same $\hat{P}_0$.

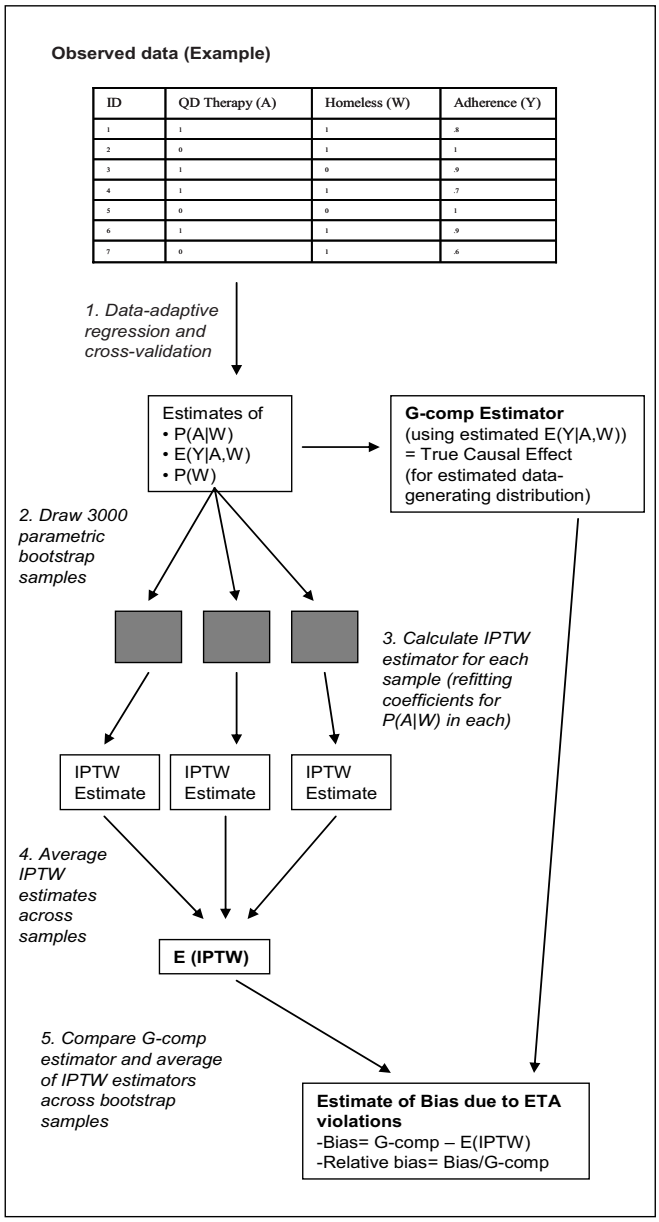
In Figure (1), we illustrate the steps of $ETA.Bias$ estimation with a hypothetical example based on the data analysis in Section (5).

Of course, in reality the estimate $\hat{P}_0$ of P_0 employed in Step 1 may be biased. However, the bootstrap simulation in the second step does not consider the true observed data distribution P_0. Thus the estimate of $ETA.Bias$ does not reflect the true bias $E_{P_0}\hat{\Psi}_{IPTW}(P_n) - \Psi(P_0)$ when $\hat{P}_0$ is misspecified from P_0, but it does measure the bias due to violation of the ETA in the world where $\hat{P}_0$ is the true data generating distribution.

DEB is a powerful practical diagnostic tool to use in real data analyses. Given an observed dataset and user-specified models $Q_{P_n}(A,W)$ and $g_{P_n}(A|W)$ (estimated in the course of implementing the G-computation, IPTW and DR estimators), this two step simulation approach immediately provides an estimate of bias due to both practical and theoretical ETA violations, under the assumption that the data-generating distribution is correctly specified. As is demonstrated in the following sections, such an estimate can readily identify situations where ETA violations occur.

DEB is publicly available as a R routine `bias.ETA()` (http://www.stat.berkeley.edu/ laan/Software/index.html). The routine takes the original data as input and performs bootstrap simulations with the user-specified

Figure 1: Step-by-step estimation of $ETA.Bias$



information such as the functional forms of the MSM and the nuisance parameter models. An example of the R routine is provided in the appendix.

4 Simulation Study.

Four simulations were carried out to investigate the behavior of DEB under a range of conditions. Specifically, the behavior of the estimator under varying intensities of practical ETA violations, sample sizes, and using both stabilized and non-stabilized weights, was compared. Simulation was further used to illustrate how the diagnostic output can be used not only to assess the extent of ETA bias present, but also to assess the consequences for bias and variance of choices regarding specification of the treatment mechanism and truncation of the resulting weights. As discussed in the Introduction, these choices can simultaneously affect bias due to incomplete control of confounding, bias due to ETA violations, and variance resulting from the use of extreme weights. Simulations demonstrate how DEB can be used to quantify these tradeoffs.

4.1 Data generation.

Let $W = (W_1, W_2, W_3, W_4)$ be a vector of baseline covariates, A be a binary treatment, Y be a post-treatment outcome, and $V = W_1$ be the adjustment variable used in the treatment-specific mean. Given the true treatment mechanism $g_0(A \mid W)$, and the function $E(Y|A, W) = Q_0(A, W)$, we generate $O = (W, A, Y)$ in the following manner:

1. Set the baseline covariates $W = (W_1, W_2, W_3, W_4)$ and $V = W_1$, where $W_i \sim U(0, 1)$ for $i = 1$ to 4;
2. Generate the observed treatment variable A from $A \mid W \sim g_0$;
3. Generate the observed outcomes $Y(A)$ as $Y(A) = Q_0(A, W) + \varepsilon$, where $\varepsilon \sim N(0, 1)$;
4. Set $O = (W, Y(A), A)$.

The following data generation models are used, unless otherwise noted:

$$\begin{aligned} Q_0(A, W) &= -1 + A + W_1 + A \times W_1 + W_2 + W_1 \times W_3, \\ g_{0,\gamma}(A|W) &= logit^{-1}(-1 + W_1 - \gamma W_2 + W_1 \times W_3), \end{aligned}$$

where $\gamma = \{1, 2, 3, 5\}$. In this model, γ reflects the intensity of practical ETA violations.

The true $\Psi(P_0)$ can be derived from the function $E(Y|A,W)$. That is,

$$\begin{aligned}\Psi(P_0|\beta) = E[Y(a)|V] &= E[E(Y|a,W)|V] \\ &= E[-1 + a + W_1 + AW_1 + W_2 + W_1W_3|W_1] \\ &= -0.5 + a + 1.5W_1 + AW_1.\end{aligned}$$

Four simulation studies are reported. In Simulation 1, $ETA.Bias$ as estimated by DEB is compared with the true bias due to ETA violations. Simulation 2 evaluates the effects on $ETA.Bias$ and MSE of using stabilized vs. non-stabilized weights, both in the absence of truncation and at a range of truncation values. Simulation 3 compares estimates of relative $ETA.Bias$ for two sample sizes and for a varying intensity of practical ETA violations. Finally, Simulation 4 investigates the effect on $ETA.Bias$ and MSE of employing a treatment mechanism that excludes a covariate that is highly predictive of treatment, and is thus the source of substantial ETA violation.

4.2 Simulation 1. True $ETA.Bias$ vs. estimated $ETA.Bias$.

Data P_n were generated using $Q_0(A,W)$ and $g_0(A|W)$, with the sample size $n = 2000$ and $\gamma = 3$. All IPTW estimators used non-stabilized non-truncated weights.

The true finite sample bias of the IPTW estimator

$$ETA.Bias = E_{P_0}\hat{\Psi}_{IPTW}(P_n) - \Psi(P_0)$$

was calculated as follows: First, $E_{P_0}\hat{\Psi}_{IPTW}(P_n)$ was calculated by drawing 3000 parametric bootstrap samples using the true data-generating distribution $Q_0(A,W)$ and $g_0(A|W)$ (i.e. sampling from P_0). The IPTW estimator was applied across bootstrap samples, in each using a correctly specified model for the treatment mechanism, but with coefficients refit in that sample. This provided an estimate of the expectation (and of the entire distribution) under P_0 of $\hat{\Psi}_{IPTW}(P_n)$. The true causal parameter $\Psi(P_0) = (-0.5, 1, 1.5, 1)$ was known. The true finite sample bias was calculated as the difference between these two quantities.

The true finite sample bias was compared to the DEB-based estimate of $ETA.Bias = E_{\hat{P}_0}\hat{\Psi}_{IPTW}(P_n^{\#}) - \Psi(\hat{P}_0)$ under correctly specified models of $Q_{P_n}(A,W)$ and $g_{P_n}(A|W)$ (equivalent to the true data-generating models $Q_0(A,W)$ and $g_0(A|W)$, but with coefficients refit using P_n). $Q_{P_n}(A,W)$ and $g_{P_n}(A|W)$ were used to draw 3000 parametric bootstrap samples (sampling from $\hat{P}_0$). Again, the IPTW estimator was applied across samples, using the model $g_{P_n}(A|W)$ with coefficients refit for each sample. This provided an estimate of $E_{\hat{P}_0}\hat{\Psi}_{IPTW}(P_n^{\#})$. The difference between this estimate and the true causal parameter for the data-generating distribution $\hat{P}_0$, $\Psi_{Gcomp}(\hat{P}_0)$ formed the estimate of $ETA.Bias$, under consistent estimation of P_0 using $\hat{P}_0$.

Result.

The true and estimated $ETA.Bias$ from Simulation 1 are provided in Table (1). The simulation demonstrates that under correct specification of the data generating distribution, the estimated $ETA.Bias$ is close to the true finite sample bias. The results of this simulation suggest that DEB provides a reasonable estimate of the bias of the IPTW estimator due to the practical violations of the ETA assumption.

Table 1: Simulation 1. Performance of the $ETA.Bias$ estimate.

Estimate	$Intercept$	a	V	aV
True finite sample bias (sampling from P_0)	0.0095	-0.1268	-0.0100	0.1346
Estimated $ETA.Bias$ (sampling from $\hat{P}_0$)	0.0086	-0.1086	-0.0094	0.1217

4.3 Simulation 2. Truncation of stabilized and non-stabilized weights.

This simulation explored how DEB can be used to quantify the effects of truncation of weights on bias and MSE of the IPTW estimator. We estimated the $ETA.Bias$ and calculated the MSE of the IPTW estimator for non-truncated weights and weights truncated at fixed points. The effects of truncation were estimated both in the context of non-stabilized weights $1/g(A|W)$ and stabilized weights $g(A|V)/g(A|W)$.

The treatment model was $g(A|W) = logit^{-1}(4 + W_1 - 10W_2 + W_1W3)$. Sample size for all scenarios was $n = 2000$. In truncating the weights, we used truncation values 5, 10 and 20 for the stabilized weights, while for non-stabilized weights, we applied truncation levels 0.05, 0.1 and 0.2 to the predicted probability of treatment given covariates.

Result.

Table (2) presents the estimated $ETA.Bias$ and MSE of the IPTW estimator using stabilized vs. non-stabilized weights, each truncated at several different values. The true and estimated $ETA.Bias$ for various truncation levels are plotted for stabilized and unstabilized weights in Figure (2). We focus our interest on estimates of the coefficients for a and aV, which reflect the causal effect of interest.

As anticipated given the weaker version of the ETA assumption required by stabilized weights, in the absence of truncation, stabilized weights were associated with both lower $ETA.Bias$, and a smaller MSE. For both stabilized and non-stabilized weights, increased truncation resulted in a larger estimated bias, reflecting increased

Figure 2: Simulation 2. Estimated and true $ETA.Bias$ for a range of truncation values, using stabilized and non-stabilized weights

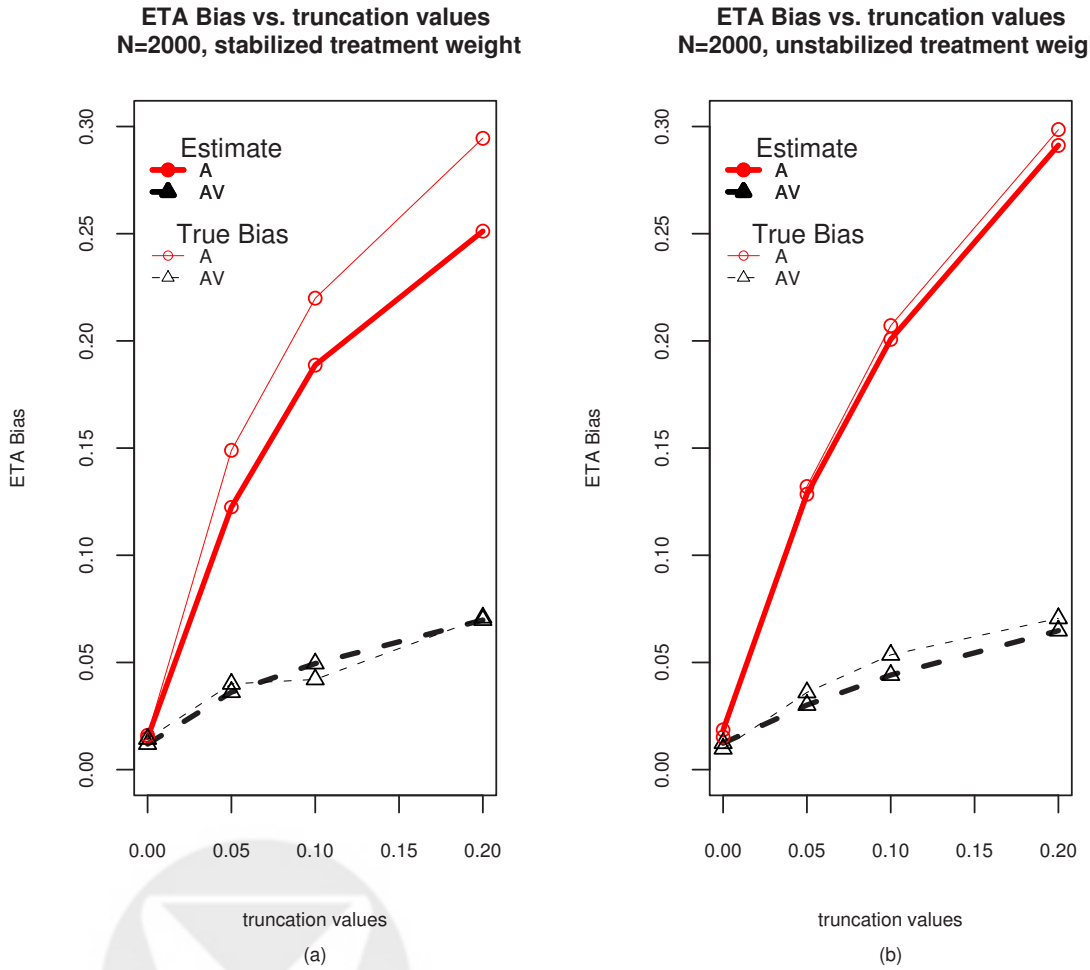


Table 2: Simulation 2. Estimated $ETA.Bias$ and MSE of IPTW estimator of unstabilized/stabilized treatment weights with no truncation and with different truncation levels.

Non-stabilized treatment weights								
	$ETA.Bias$				MSE			
Truncation	Intercept	a	V	aV	Intercept	a	V	aV
0.2	0.0891	-0.2913	0.0530	0.0648	0.0083	0.0857	0.0041	0.0071
0.1	0.0451	-0.2007	0.0614	0.0441	0.0026	0.0417	0.0060	0.0066
0.05	0.0161	-0.1285	0.0575	0.0301	0.0013	0.0190	0.0072	0.0096
0	0.0011	-0.0185	0.0066	0.0123	0.0023	0.0105	0.0123	0.0317
Stabilized treatment weights								
	$ETA.Bias$				MSE			
Truncation	Intercept	a	V	aV	Intercept	a	V	aV
5	0.0811	-0.2512	0.0258	0.0698	0.0069	0.0639	0.0020	0.0076
10	0.0504	-0.1887	0.0323	0.0494	0.0031	0.0371	0.0034	0.0071
20	0.0157	-0.1224	0.0409	0.0361	0.0011	0.0175	0.0053	0.0099
Inf	0.0016	-0.0149	0.0024	0.0119	0.0016	0.0073	0.0077	0.0209

misspecification of the treatment mechanism. In addition, results from both stabilized and non-stabilized weights show that truncation can provide an improved bias-variance tradeoff, by decreasing the variability of the weights and thus of the IPTW estimator. In the case of the stabilized weights, for example, the results suggest that MSE for the causal coefficients on a is minimized by no truncation, while the MSE for the coefficient on aV is minimized by truncating weights at 10.

4.4 Simulation 3. Varying sample size and intensity of practical ETA violations.

This simulation investigated the extent of $ETA.Bias$ under treatment models $g(A|W)$ that violated the practical ETA assumption at different levels ($\gamma = \{1, 2, 3, 5\}$); as γ increased, practical ETA violations became more intense. The effect of increasing practical ETA violations was investigated for two sample sizes $n_1 = 2000$ and $n_2 = 200$, and in the presence and absence of truncation. Non-stabilized weights were used throughout, and truncation, when employed, was at the level of 0.1. In addition, true finite sample biases were calculated as described in Simulation 1.

Result.

The results from Simulation 3 are provided in Table (3) and plotted in Figure (3). Again, we focus on the estimated coefficients on a and aV, as these reflect the causal effect of interest. Several points are worth noting. First, as expected, the estimated $ETA.Bias$ increases as the intensity of practical ETA violation in the treatment model (γ) increases. Second, in the absence of truncation, a decrease in sample size results in an increase in estimated $ETA.Bias$, as expected given that, given the same treatment model, more practical ETA violations should occur at smaller sample size. Interestingly, estimated bias does not necessarily decrease with increasing sample size when truncation is employed. This reflects that fact that truncation of weights corresponds to misspecification of the treatment mechanism, resulting in an estimator that is not asymptotically consistent. In this setting, $ETA.Bias$ reflects both bias due to practical ETA violations, and bias due to misspecification of the treatment mechanism, which does not necessarily decrease with increasing sample size.

4.5 Simulation 4. Selection of covariates for inclusion in the treatment mechanism used for estimation of weights.

This simulation explored the use of DEB to investigate the bias-variance tradeoffs implied by excluding certain covariates from the treatment mechanism used for estimation of the weights. Covariates W were generated $W_i \sim U(-1, 1)$ for $i = 1, 2, 3$, the treatment model used was $g_0(A|W) = logit^{(-1)}(1 + W_1 + 10W_3)$ and Y was generated as $Q_0(A, W) = -1 + A + W_1 + W_2 + W_3$. Sample size was 2000, and stabilized weights with no truncation were used. As a strong predictor of A, W_3 was thus a major source of practical ETA violation, as well as having a moderate effect on Y. $ETA.Bias$ and MSE were estimated for three IPTW estimators relying on three alternative treatment mechanisms to generate weights: 1) a correctly specified model; 2) an incorrectly specified model, including $W3$; and 3) an incorrectly specified model, excluding W_3.

Result.

The results of Simulation 4 are shown in Table (4). $ETA.Bias$ under the the correctly specified model, including W_3, consisted of bias due to ETA violations, while $ETA.Bias$ under the incorrectly specified treatment models also included bias due to misspecification of the treatment mechanism. Interestingly, inclusion of W_3 in either a correct or misspecified treatment model resulted in a larger estimates $ETA.Bias$ than use of a misspecified treatment model excluding W_3, suggesting that in this example, bias due to uncontrolled confounding by W_3 is less important that bias resulting from ETA violations due to W_3. Similarly, the MSE in this simulation was minimized by using a misspecified treatment mechanism excluding W_3.

Figure 3: Simulation 3. Estimated and true $ETA.Bias$ given intensity of practical ETA violation (gamma).

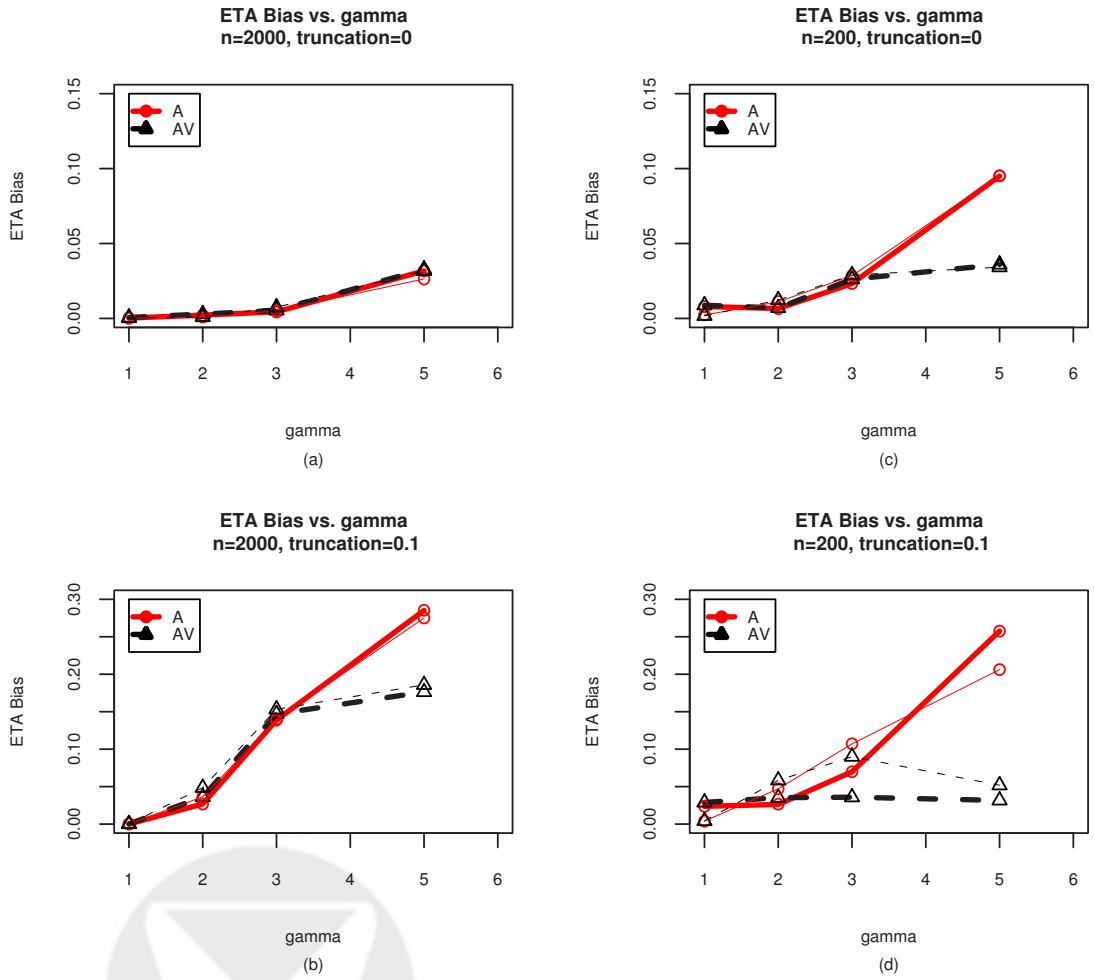


Table 3: Simulation 3. Estimated $ETA.Bias$ for varying intensity of practical ETA violations (γ). True finite sample biases are inside the parentheses.

Truncation level=0, N=2000				
γ	Intercept	a	V	aV
1.00	-0.0005(0.0194)	0.0003(-0.0002)	0.0005(-0.0161)	-0.0006(0.0003)
2.00	0.0005(0.0205)	-0.0022(-0.0008)	-0.0010(-0.0171)	0.0029(0.0009)
3.00	0.0000(0.0205)	-0.0044(-0.0051)	-0.0001(-0.0168)	0.0053(0.0075)
5.00	0.0002(0.0198)	-0.0317(-0.0264)	-0.0004(-0.0166)	0.0328(0.0317)
Truncation level=0.1, N=2000				
γ	Intercept	a	V	aV
1.00	0.0005(0.0198)	0.0000(-0.0001)	-0.0003(-0.0155)	-0.0002(0.0000)
2.00	0.0024(0.0232)	-0.0269(-0.0365)	-0.0036(-0.0204)	0.0362(0.0485)
3.00	0.0103(0.0299)	-0.1388(-0.1398)	-0.0104(-0.0267)	0.1469(0.1532)
5.00	0.0127(0.0323)	-0.2852(-0.2750)	-0.0038(-0.0214)	0.1763(0.1859)
Truncation level=0, N=200				
γ	Intercept	a	V	aV
1.00	-0.0001(-0.0632)	-0.0050(-0.0023)	0.0013(0.0384)	0.0088(0.0018)
2.00	-0.0011(-0.0631)	-0.0066(-0.0111)	0.0044(0.0381)	0.0090(0.0122)
3.00	-0.0011(-0.0646)	-0.0232(-0.0285)	0.0020(0.0380)	0.0262(0.0290)
5.00	-0.0013(-0.0631)	-0.0950(-0.0953)	0.0020(0.0352)	0.0360(0.0342)
Truncation level=0.1, N=200				
γ	Intercept	a	V	aV
1.00	0.0005(-0.0617)	-0.0237(-0.0041)	0.0008(0.0356)	0.0289(0.0041)
2.00	0.0028(-0.0567)	-0.0262(-0.0471)	-0.0040(0.0270)	0.0353(0.0584)
3.00	0.0038(-0.0543)	-0.0699(-0.1070)	0.0003(0.0282)	0.0359(0.0900)
5.00	0.0122(-0.0520)	-0.2576(-0.2062)	0.0028(0.0356)	0.0316(0.0519)

5 Data analysis.

We applied our diagnostic tool to two data examples drawn from the treatment of patients infected with HIV. Antiretroviral drugs greatly reduce morbidity and mortality due to HIV infection. However, this efficacy requires that complex regimens be taken as prescribed over a patient's lifetime. Less than perfect adherence risks a loss of viral suppression, which can lead both to increased morbidity and to the emergence of resistant virus. Thus interventions capable of improving adherence to antiretroviral drugs are needed. We investigated two interventions hypothesized to affect patient adherence: 1) the use of a pill box organizer, and; 2) the use of a once daily antiretroviral therapy regimen (as compared to a regimen with more frequent dosing). Marginal structural models were used to estimate the effect of these two in-

Table 4: Simulation 4. Estimated $ETA.Bias$ and MSE of IPTW estimator using different treatment models.

	$ETA.Bias$			MSE		
$g_1(A\|W)$	Intercept	a	V	Intercept	a	V
$W_1 + W_3$	0.1904	-0.2760	0.0427	0.0971	0.1680	0.0510
$W_1 + W_2 + W_3$	0.1801	-0.2675	0.0621	0.0900	0.1573	0.0517
$W_1 + W_2$	0.0003	0.0001	-0.0008	0.0006	0.0007	0.0010

terventions based on the IPTW, G-computation and Double Robust estimators. DEB was used as a diagnostic of the IPTW estimator for both causal effects of interest.

5.1 Data Structure and Methods.

Data were drawn from the Research on Access to Care in the Homeless (REACH) cohort, an observational cohort of HIV-infected patients in San Francisco, California. Patients in REACH are followed longitudinally, and data collected on current and past treatment with antiretroviral drugs, lab values including CD4 T cell count and plasma HIV RNA level (viral load), homelessness, and recreational drug use. Adherence to antiretroviral therapy is assessed monthly using unannounced pill counts (Bangsberg et al. (2000)). In the current analyses, we estimated the effect of pill box organizer use/once daily therapy for a given month on adherence the same month; thus a single subject could contribute multiple observations.

The data for a single observation consisted of binary treatment status for a given month (A=I(pill box use) in analysis 1; A=I(once daily therapy) in analysis 2), a continuous outcome Y defined as adherence for the same month; and a list of 29 covariates W from the previous month (including baseline covariates such as age and sex, and time-varying covariates such as recent homelessness and recreational drug use). The chronological data for a given observation thus consisted of $O = (W, A, Y)$. In total, 237 individuals contributed a total of 2504 observations to analysis 1. Because once daily therapy was not available when follow-up was initiated (a theoretical ETA violation), data for analysis 2 were restricted to dates after once daily therapy first became available, resulting in a slightly smaller sample of 1445 observations among 196 individuals.

For both analyses, the following MSM was assumed: $m(a|\beta) = \beta_0 + \beta_1 a$. The causal parameter of interest was $E(Y_1 - Y_0) = \beta_1$. G-computation, IPTW and DR estimators were calculated, with standard errors estimated based on 200 non-parametric bootstrap samples. The IPTW estimator used stabilized weights $g(A)/g(A|W)$; the weights were truncated by 10 from above.

Models of both nuisance parameters, $g(A|W)$ and $Q(A, W)$, were fit using the Deletion/Substitution/Addition algorithm, with 5 fold cross validation (Sinisi and

van der Laan (2004)). The algorithm was used to search aggressively among all models with a maximum of 10 terms and two-way interactions. We present estimates of the causal effect of each intervention based on the IPTW, DR, and G-computation estimators, and discuss the results of DEB-based $ETA.Bias$ estimation.

5.2 Results: The effect of pill box use and once daily therapy on adherence.

All three marginal structural model estimators suggested that pill box use improves adherence by about four percent; the effect was significant for the G-computation and DR estimators, and of borderline significance for the IPTW estimator (Table (5)). Estimation of $ETA.Bias$ using DEB suggested that the bias due to ETA.violations in estimates of both β_0 and β_1 was minimal, about one percent relative bias in the estimate of the causal parameter of interest, β_1 (Table (6)).

Table 5: Estimates of the effect of pill box use on adherence.

Estimator	$\hat{\beta}_0$(standard error)	$\hat{\beta}_1$(standard error)
G-computation	0.701 (0.019)	0.045 (0.013)
Double Robust	0.702 (0.019)	0.041 (0.015)
IPTW (stabilized)	0.698 (0.021)	0.041 (0.021)

Table 6: Bias and relative bias of the IPTW estimator of the effect of pill box use due to violation of the ETA assumption. Relative $ETA.Bias = ETA.Bias/\beta_{Gcomp}$. $\beta^{\#}_{IPTW}$ denotes the IPTW estimator of β applied to the bootstrap samples.

Estimate	$\hat{\beta}_0$	$\hat{\beta}_1$
Mean of $\beta^{\#}_{IPTW}$	0.693	0.041
$ETA.Bias$	-0.000173	0.000452
Relative $ETA.Bias$	-0.000246	0.0101

The estimated causal effect of once daily therapy on adherence is presented in Table (7). In contrast to the positive effect estimated for pill box use, none of the MSM estimators supported the hypothesis that use of once daily therapy affects adherence. The standard errors for all three estimators were large, and the point estimates of β_1 based on the G-computation and DR estimators were close to zero (a point estimate of 0.1 percent improvement in adherence for both estimators). Interestingly, the IPTW estimate in this analysis diverged from the other two estimators (a point estimate of a 2.9 percent improvement in adherence due to once daily therapy). Violation of the ETA assumption provides a plausible explanation for this divergence; DEB suggested

a relative bias of nearly 850 percent in the IPTW estimator due to ETA violations (Table (8)).

Table 7: Estimates of the effect of once daily therapy on adherence.

Estimator	$\hat{\beta}_0$(standard error)	$\hat{\beta}_1$(standard error)
G-computation	0.735 (0.021)	0.001 (0.025)
Double Robust	0.736 (0.022)	0.001 (0.030)
IPTW (stabilized)	0.729 (0.029)	0.029 (0.065)

Table 8: Bias and relative bias of the IPTW estimator of the effect of once daily therapy due to violation of the ETA assumption. Relative $ETA.Bias = ETA.Bias/\beta_{Gcomp}$. $\beta^{\#}_{IPTW}$ denotes the IPTW estimator of β applied to the bootstrap samples.

Estimate	$\hat{\beta}_0$	$\hat{\beta}_1$
Mean of $\beta^{\#}_{IPTW}$	0.735	0.0094
$ETA.Bias$	0.000117	-0.00965
Relative $ETA.Bias$	0.000159	-8.493

6 Discussion.

The Diagnostic for ETA Bias (DEB) introduced in this article represents a straightforward method for assessing the extent of bias in the IPTW estimator arising from violations of the ETA assumption, under the assumption that the data-generating distribution is correctly specified. The consistency of the IPTW estimator relies on correct specification of the treatment mechanism; when the treatment mechanism is misspecified, the IPTW estimator will be biased regardless of the extent of ETA violation. Thus when presenting DEB, we have treated the treatment mechanism as a given. In reality of course, it can rarely be assumed that the treatment mechanism is correctly specified; in this case sensitivity analysis using alternative models of $g(A|W)$ is warranted. While DEB can serve as a complementary diagnostic to such analyses, alone it does not help to identifying bias due to misspecification of the treatment mechanism; this must be identified using background knowledge and alternative tools. Misspecification of the F_X part of the likelihood, on which the G-computation estimator depends, can also affect performance of DEB. Thus, when estimating $Q_n(A, W)$ one should ideally employ a range of model-fitting algorithms and cross-validation. Such effort will benefit the analyst dually by improving both

the performance of the ETA diagnostic and the consistency of the G-computation estimator.

We have further illustrated how DEB can be used to quantify the bias-variance effects of practical choices in the implementation of the IPTW estimator. Such choices include whether and at what level to truncate the weights, and whether or not to deliberately exclude from the treatment mechanism a covariate known to be a major ETA violator. We feel the resulting output is helpful as an additional diagnostic tool in implementing an IPTW data analysis. However, it should be stressed that the estimates of bias and MSE provided by the diagnostic again depend on correct specification of the treatment mechanism and data-generating distribution. In settings where there are a small number of very unlikely treatment assignments, standard methods of estimating the treatment mechanism are unlikely to capture these outliers. As a result, the MSE estimates based on DEB may not reflect, for example, the true gains in variance achieved by truncation. Thus we view DEB as a source of additional diagnostic information under a specified treatment mechanism, rather than as a formal method for data-adaptive selection of truncation level or of covariates for inclusion. Further work is needed to examine the best way to integrate the additional information provided by DEB into the practical choices of IPTW implementation. It is worth noting, however, that currently these choices are made in the absence of any quantitative information. Thus DEB represents a useful addition to the tools currently available.

The current article has focused on the estimation of the causal effects of point treatment. However, the same diagnostic approach can be readily applied to the longitudinal data setting. Indeed, potential for ETA violations becomes of even greater concern as the number of time points, and thus a subject's covariate and treatment history, increases. In addition, the approach readily generalizes to situations in which the aim is not to estimate a causal effect, but simply a specific parameter of the data-generating distribution, such as a treatment-specific mean adjusted for covariates W.

It is increasingly apparent that ETA violations, both practical and theoretical, represent a common source of significant bias when estimating causal parameters from real data. We have provided a simple diagnostic to evaluate the extent of bias in the IPTW estimator due to ETA violations. DEB can be be used not only to evaluate the validity of the IPTW estimate, but also to help identify situations in which the G-computation and DR estimates rely on extrapolation to areas of the data with little support. In addition, when the G-computation and IPTW estimators provide discrepant results, an estimate of the finite sample ETA bias can help to guide decisions regarding the reliability of the IPTW estimate. The approach presented is available as a public R routine, is easily implemented, and provides readily interpretable results about the extent of bias due to ETA assumptions. As such, it is a valuable addition to the statistical toolbox for causal inference.

References

D. R. Bangsberg, F. M. Hecht, E. D. Charlebois, A. R. Zolopa, M. Holodniy, L. Sheiner, J. D. Bamberger, M. A. Chesney, and A. Moss. Adherence to protease inhibitors, HIV-1 viral load, and development of drug resistance in an indigent population. *AIDS*, 14(4):357–366, 2000.

R. Neugebauer and M. J. van der Laan. Locally efficient estimation of nonparametric causal effects on mean outcomes in longitudinal studies. *Journal of Statistical Planning and Inference*, In press, 2006.

R. Neugebauer and M.J. van der Laan. Why prefer double robust estimates. *Journal of Statistical Planning and Inference*, 129(1-2):405–426, 2005.

J.M. Robins. Robust estimation in sequentially ignorable missing data and causal inference models. In *Proceedings of the American Statistical Association: Section on Bayesian Statistical Science*, pages 6–10, 1999.

J.M. Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials (Minneapolis, MN, 1997)*, pages 95–133. Springer, New York, 2000.

J.M. Robins, M.A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000.

D.B. Rubin. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics*, 6:34–58, 1978.

S. E. Sinisi and M. J. van der Laan. Deletion/Substitution/Addition algorithm in learning with applications in genomics. *Statistical Applications in Genetics and Molecular Biology*, 3(1), 2004. URL `http://www.bepress.com/sagmb/vol3/iss1/art18`. Article 18.

M.J. van der Laan and J.M. Robins. *Unified methods for censored longitudinal data and causality*. Springer, New York, 2002.

Appendix.

R routine `bias.ETA()`.

The following is an example of the R routine `bias.ETA()`, available at http://www.stat.berkeley.edu/ laan/Software/index.html:

```
bias.ETA(y=y, a=a, v=w1, w=w, data=obs.data,
      yfamily='gaussian', afamily='binomial',
      model.msm = list(Model="a+w1+a:w1"),
      model.aw  = list(Model="w1+w2+w1:w3"),
      model.yaw = list(Model="a+w1+a:w1+w2+w1:w3"),
      model.av  = NULL, stabilized.wt=F,
      n.sim = 3000, index.v.inW = c(1))
```