University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2002 *Paper* 118

An Empirical Study of Marginal Structural Models for Time-Independent Treatment

Tanya A. Henneman* Mark J. van der Laan†

*Department of Biostatistics, School of Public Health, University of California, Los Angeles, thenn@ucla.edu

†Division of Biostatistics, School of Public Health, University of California, Berkeley

An Empirical Study of Marginal Structural Models for Time-Independent Treatment

Tanya A. Henneman and Mark J. van der Laan

Abstract

In non-randomized treatment studies a significant problem for statisticians is determining how best to adjust for confounders. Marginal structural models (MSMs) and inverse probability of treatment weighted (IPTW) estimators are useful in analyzing the causal effect of treatment in observational studies. Given an IPTW estimator a doubly robust augmented IPTW (AIPTW) estimator orthogonalizes it resulting in a more e±cient estimator than the IPTW estimator. One purpose of this paper is to make a practical comparison between the IPTW estimator and the doubly robust AIPTW estimator via a series of Monte- Carlo simulations. We also consider the selection of the optimal choice in the class of IPTW estimators. Also included is an empirical study that examines the range of efficiency across the class of AIPTW estimators. Both the IPTW and AIPTW estimators are applied in the analysis of data collected from The Study of Physical Performance and Age-Related Changes in Sonomans.

1 Introduction

In observational studies when treatment is not randomly assigned, individual characteristics or an individual's degree of well-being can affect the health outcome being studied as well as influence treatment decisions. Even in randomized studies, when the goal is to insure that treatment arms represent similar patient populations, confounding can occur when for example, subjects decide not to comply with their assigned treatment based on how well they are feeling. Accounting for confounders, in randomized and non-randomized studies, in an appropriate manner is an important statistical issue in analyzing data from medical treatment studies. Techniques such as matching, stratification of the data, and adjusting for all possible confounders in a regression model are examples of methods statisticians have used to adjust for measured and known confounders. To estimate causal parameters Robins (1998) proposes a class of inverse probability of treatment weighted (IPTW) estimators when the assumption of no unmeasured confounders is valid. The IPTW estimator has been utilized in treatment studies to determine causal effect. Hernan et al. (2001) used the estimator to estimate the joint effect of zidovudine (AZT) and prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of HIV-positive men. Under certain conditions, the IPTW estimator can produce consistent estimates of causal parameters [5]. However, IPTW estimators require model specification and estimation of the propensity score, and if estimated inconsistently, consistency of the IPTW is not guaranteed. In Robins (1999b) and Robins et al (2000a), a class of doubly robust augmented IPTW (AIPTW) estimators are introduced for marginal structural models (MSMs) that protect against misspecification of the propensity score and makes gains in efficiency over the IPTW estimators.

In this paper we consider the problem of estimating causal parameters in point-treatment studies where all potential confounders are measured and available to the researcher. As of yet, the practical performance of the AIPTW when estimating treatment effect for point treatment studies, has not been evaluated or compared to the IPTW in the literature. We aim to present a practical comparison of the IPTW and the AIPTW estimators via a series of Monte Carlo simulation studies. This paper also includes an empirical study that examines the range of efficiency across the class of AIPTW estimators. Finally, we apply both estimators in the analysis of data collected from The Study of Physical Performance and Age-Related changes in Sonomans.

2 Marginal Structural Models

Let A represent a treatment or exposure variable whose effect on Y is of interest for a given population. Let $\mathcal{A}$ represent the set of all realizations of A. Using Rubin's Causal Model [4] we assume for each subject the existence of treatment specific outcomes Y_a for each $a \in \mathcal{A}$. Each subject in the population is assigned to a specific treatment arm and for a subject who receives treatment a^* we observe Y_{a^*}, the unobserved Y_a's are called counterfactuals or potential outcomes. Let $X = ((Y_a)_{a\in\mathcal{A}}, A, W)$ represent the full data on a subject with W representing a set of measured covariates. Let $O_i = (A_i, Y_i, W_i)$ represent the observed data for $i = 1, \ldots, n$ subjects. The counterfactual response is modeled as

$$Y_a = m(a, V|\beta) + \epsilon,\ E(\epsilon|V) = 0 \tag{1}$$

where $V \subset W$ is a k-dimensional vector of real-valued covariates and where $m(A, V|\beta)$ represents a MSM parameterized by β with $\dim(\beta) = q$. We are focused on the problem of estimating causal parameters for point-treatment studies when the treatment variable is dichotomous, i.e. $\mathcal{A} = \{0, 1\}$, recognizing that many of our results can be extended for ordinal or continuous treatment variables.

Causal inference is based, in part, on the idea that the existence of all possible outcomes is inherent within each subject, and that what determines the outcome is the chance of observing one of these counterfactuals. This makes central to the discussion of causality the propensity score, the probability an individual receives their assigned treatment. Formally, the propensity score is defined as the conditional probability of treatment given all measured confounders, i.e. $P(A = 1|W)$. Note, $g(\cdot)$ will be used to denote the conditional density of A. In addition to identifying a model of the marginal counterfactual distribution, MSMs require specification of the treatment mechanism.

To continue developing the framework necessary for statistical inference, we provide commonly invoked assumptions involving the propensity score. First is the randomization assumption, given formally as,

$$g(A|X) = g(A|W).$$

This means that treatment assignment is determined only by what is observed, and that there are no unmeasured confounders. Secondly, we assume that each subject has a positive probability of receiving any of the proposed treatments, i.e. $g(a|W) > 0$ for all $a \in \mathcal{A}$ and W.

3 Inverse Probability of Weighted Treatment Estimator

The motivation behind the IPTW estimator is that weighting observations by their respective propensity score creates a pseudo-population in which treatment assignment is no longer confounded [6]. Consider the following class of estimating functions that is indexed by h, a real-valued function of A and V, and where $\epsilon(\beta) = Y - m(A, V|\beta)$ is the observed residual,

$$\left\{ \theta_h(O|\beta, g) = \frac{h(A,V)\epsilon(\beta)}{g(A|W)} : h \right\}.$$

For all $h \in \mathcal{H} = \{h : h(A, V)\}$ the following identity

$$E\left(\frac{h(A,V)\epsilon(\beta)}{g(A|W)}\right) = 0, \tag{2}$$

holds if 1) the randomization assumption and the 2) nonparametric identifiability assumption that $\max_{a\in\mathcal{A}} \frac{h(a,V)}{g(a|W)} < \infty$, almost everywhere, hold. If this assumption is violated then the data itself does not identify the parameter β, so additional assumptions would be needed. From (2) we get unbiased estimating equations of the form,

$$\frac{1}{n}\sum_{i=1}^{n} \frac{h(A_i, V_i)\epsilon_i(\beta)}{g(A_i|W_i)} = 0. \tag{3}$$

The estimator that solves (3) is referred to as the IPTW estimator and we will represent this estimator by $\hat{\beta}^{1,h}$. $\hat{\beta}^{1,h}$ is asymptotically linear with influence function

$$IC_h(O|\beta, g) = -E\left(\frac{d}{d\beta}\frac{h(A,V)\epsilon(\beta)}{g(A|W)}\right)^{-1} \cdot \frac{h(A,V)\epsilon(\beta)}{g(A|W)}.$$

That is,

$$\begin{aligned} \sqrt{n}(\hat{\beta}^{1,h} - \beta) &= \frac{1}{n}\sum_{i=1}^{n} IC_h(O_i|\beta, g) + o_p(1) \\ &\Rightarrow N(\mathbf{0}, \Sigma), \end{aligned}$$

where $\Sigma = E[IC_h(O|\beta, g)IC_h(O|\beta, g)^\top]$ and $\mathbf{0}$ is the zero vector.

3.1 Selection and estimation of nuisance parameters

As was stated earlier, this is a class of functions that varies over $\mathcal{H}$. Any choice of h can produce a consistent estimate of β because of the identity in (2), therefore selection of a particular h involves determining which is the optimal choice in the efficiency sense. When there is no confounding, the optimal estimating procedure would be a weighted least squares regression with weights given as the inverse of the variance of the residuals. And because there are no confounders a significant association between the treatment and outcome would

imply causation [5, 3]. Considering the MSM in (1) a smart choice for h would be a function such that the resulting IPTW estimator would reduce to the optimal estimator when there is no additional confounding beyond V. One selection of h is [5]

$$h^*(A,V) = g(A|V) \cdot \frac{\frac{d}{d\beta}m(A,V|\beta)}{\mathrm{Var}(\epsilon(\beta)|A,V)}, \tag{4}$$

with corresponding IPTW estimating function,

$$\theta_{h^*}(O|\beta,g) = \frac{g(A|V)}{g(A|W)} \frac{\frac{d}{d\beta}m(A,V|\beta)}{\mathrm{Var}(\epsilon(\beta)|A,V)}\epsilon(\beta). \tag{5}$$

The incorporation of $g(A|V)$ in the numerator of (5) helps to control for observations with very small propensity scores which can result in an estimator with a very large variance [5]. The ratio of weights, $\frac{g(A|V)}{g(A|W)}$, is known as stabilizing weights. Stabilizing weights are useful in reducing the variance of the IPTW estimator [5]. In the case of no additional confounding beyond V, i.e. $\frac{g(A|V)}{g(A|W)} = 1$, (5) reduces to the optimal estimating function for the regression model $E(Y|A,V) = m(A,V|\beta)$. Consequently, the IPTW using estimated weights will be at least as efficient as the standard regression estimator [13].

In van der Laan et al. (2002) an interesting choice of h is the h which gives an optimal covariance matrix $E[IC_h(O \mid \beta,g)IC_h(O \mid \beta,g)^\top]$: i.e. this would be the optimal estimating function in the case where g is known and probably a good choice in general. This alternate index is given by:

$$h^{alt}(A,V) = \frac{d}{d\beta}m(A,V \mid \beta)\frac{E(g^{-1}(A \mid X) \mid A,V)}{E(\epsilon^2(\beta)/g^2(A \mid X) \mid A,V)}.$$

The resulting IPTW estimating function under the randomization assumption is

$$\theta_{h^{alt}}(O|\beta,g) = \frac{E(g^{-1}(A \mid W) \mid A,V)}{E(\epsilon^2(\beta)/g^2(A \mid W) \mid A,V)}\frac{\frac{d}{d\beta}m(A,V \mid \beta)}{g(A|W)}\epsilon(\beta),$$

and is a consistent estimating function for β. $\hat{\beta}^{1,h^{alt}}$ represents the estimator which solves

$$\frac{1}{n}\sum_{i=1}^{n}\theta_{h^{alt}}(O_i|\beta,g),$$

and has influence function

$$\begin{aligned} IC_{h^{alt}}(O|\beta,g) = \quad & -E\left[\frac{E(g^{-1}(A \mid W) \mid A,V)}{E(\epsilon^2(\beta)/g^2(A \mid W) \mid A,V)}\frac{\frac{d}{d\beta}m(A,V \mid \beta)}{g(A|W)}\frac{d}{d\beta}\epsilon(\beta)\right]^{-1} \\ & \times \quad \frac{E(g^{-1}(A \mid W) \mid A,V)}{E(\epsilon^2(\beta)/g^2(A \mid W) \mid A,V)}\frac{\frac{d}{d\beta}m(A,V \mid \beta)}{g(A|W)}\epsilon(\beta). \end{aligned}$$

One thing to observe is that when there is no additional confounding beyond V, $\hat{\beta}^{1,h^{alt}}$ reduces to the optimal estimating function for the regression model $E(Y|A,V) = m(A,V|\beta)$. In order to see that one gains in efficiency using h^{alt} we ran Monte Carlo simulations to compare the limiting variance for $\hat{\beta}^{1,h^*}$ and $\hat{\beta}^{1,h^{alt}}$. In the settings we examined they were comparable and it remains to be seen if there are situations where there is a notable difference.

Generally, the functional form of g is not known to the researcher, and must be estimated. When A is binary a logistic model is commonly used to model g [10, 9, 2]. Similarly a logistic model can be used to estimate $g(A|V)$. If $g(A|W)$ is correctly specified, then the resulting estimator from (5) is consistent, even if the model for $g(A|V)$ is misspecified [3, 6]. However, if $g(A|W)$ is misspecified then the IPTW estimator can be inconsistent.

In addition to g being unknown, $E(\epsilon^2(\beta)|A,V)$ is also usually unknown. The variance component can be estimated by assuming a reasonable model for $E(\epsilon^2(\hat{\beta})|A,V)$, with some initial estimate $\hat{\beta}$. Once the nuisance components of the IPTW estimator have been modelled and estimated, a weighted least squares (WLS) regression with weights for a subject i given as

$$w_i = \frac{g_n(A_i|V_i)}{g_n(A_i|W_i)} \text{ or } \quad w_i = \frac{g_n(A_i|V_i)}{g_n(A_i|W_i)\hat{E}(\epsilon_i^2(\hat{\beta})|A_i,V_i)},$$

and where g_n is an estimate of g, can be performed to calculate $\hat{\beta}^{1,h}$. Alternately, a Newton-Raphson procedure can be employed to determine $\hat{\beta}^{1,h}$ by solving the estimating equation in (3).

There exists a couple of options when considering the calculation of standard errors and confidence intervals. One is to use the resulting standard error estimates from the WLS procedure. However, these estimates will be conservative when using estimated, instead of known, weights. Alternately, bootstrapping methods may be employed or standard error estimates can obtained using the corresponding influence curve for $\hat{\beta}^{1,h}$. The covariance matrix can be estimated as

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} I\hat{C}_h(O_i|\hat{\beta}^{1,h}, g_n)I\hat{C}_h(O_i|\hat{\beta}^{1,h}, g_n)^\top. \tag{6}$$

A 95% confidence interval for $\hat{\beta}_j^{1,h}$, for $j = 1,\ldots,q$, is $\hat{\beta}_j^{1,h} \pm 1.96 \cdot \hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is the appropriate diagonal element in (6).

4 Augmented Doubly Robust IPTW Estimators

In observational studies the mechanism for assigning treatment is often not known. Thus the propensity score must be estimated when implementing the IPTW estimating procedure. Two potential concerns arise in this situation. One is that resulting standard error estimates can be conservative since they do not take into account the use of an estimated propensity score. Secondly, if the specified model for the treatment mechanism is incorrect, then the IPTW estimator could produce inconsistent parameter estimates. One can always improve an estimating function by subtracting its projection on a nuisance tangent space [1], thus efficiency is improved by subtracting from an IPTW estimator its projection on the tangent space of the propensity score. The resulting doubly robust AIPTW estimator [7, 8] is more efficient than the IPTW estimator and under certain conditions can be protected against misspecification of the propensity score.

Let $\theta_h(O|\beta,g)$ be an IPTW estimating function for β. To obtain a more efficient estimator we should subtract from the IPTW estimator its projection onto the space spanned by all scores of g. Consider an extended class of estimating functions

$$\left\{\Psi_{h,\phi}(O|\beta,g) = \frac{h(A,V)\epsilon(\beta)}{g(A|W)} + \phi(A,W) : \phi, h\right\},$$

indexed by $h \in \mathcal{H}$ and ϕ, where ϕ is a real-valued $\dim(\beta)$ function of A and W. Let T_{prop} denote the tangent space of the propensity score, defined as the space of all scores corresponding to the propensity score, i.e. the space of all functions of A and W with conditional mean 0. Then the projection of $\theta_h(O|\beta,g)$ onto T_{prop} is

$$\Pi(\theta_h(O|\beta,g)|T_{\text{prop}}) = E(\theta_h(O|\beta,g)|A,W) - E[E(\theta_h(O|\beta,g)|A,W)|W].$$

We choose ϕ so that $\Psi_{h,\phi}(O|\beta,g)$ corresponds to the IPTW estimator minus it's projection on the tangent space of the propensity score. Let $\perp$ denote the orthogonal complement and let ϕ^* be such that

$$\phi^*(A,W) = -\Pi(\theta_h(O|\beta,g)|T_{\text{prop}}).$$

Under the nonparametric identifiability assumption, the proposed AIPTW estimator has as an estimating function

$$\begin{aligned}\Psi_{h,\phi^*}(O|\beta,g) &= \Pi(\theta_h(O|\beta,g)|T^{\perp}_{prop}) \\ &= \frac{h(A,V)\epsilon(\beta)}{g(A|W)} - \frac{h(A,V)}{g(A|W)}E(\epsilon(\beta)|A,W) + \sum_{\mathcal{A}} h(a,V)E(\epsilon(\beta)|A=a,W).\end{aligned}$$

This proposed estimator retains the consistency property of the IPTW estimator since $\Pi(\theta_h(O|\beta,g)|T_{\text{prop}})$ has conditional mean zero. Let $\hat{\beta}^{2,h}$ denote the AIPTW estimator that solves the following estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_{h,\phi^*}(O_i|\beta,g) = 0. \tag{7}$$

Under standard regularity conditions, $\hat{\beta}^{2,h}$ is a consistent and asymptotically linear estimate of β with influence function

$$IC_{h,\phi^*}(O|\beta,g) = -E\left[\frac{d}{d\beta}\Psi_{h,\phi^*}(O|\beta,g)\right]^{-1}\cdot\Psi_{h,\phi^*}(O|\beta,g).$$

Thus,

$$\sqrt{n}(\hat{\beta}^{2,h}-\beta) \Rightarrow N(\mathbf{0},\Sigma),$$

where $\Sigma = E[IC_{h,\phi^*}(O|\beta,g)IC_{h,\phi^*}(O|\beta,g)^{\top}]$. One thing to observe is

$$\frac{d}{d\beta}\Psi_{h,\phi^*}(O|\beta,g) = \frac{d}{d\beta}E\left[\frac{h(A,V)\epsilon(\beta)}{g(A|W)}\right].$$

Therefore the influence curve for $\hat{\beta}^{2,h}$ can be expressed as,

$$IC_{h,\phi^*}(O|\beta,g) = E\left[\frac{d}{d\beta}\frac{h(A,V)\epsilon(\beta)}{g(A|W)}\right]^{-1}\cdot\Psi_{h,\phi^*}(O|\beta,g).$$

The Newton-Raphson procedure can be employed to solve the equations in (7) and produce AIPTW estimates of causal parameters. After obtaining an estimate of β we can calculate standard error estimates using an estimate of the influence function. The empirical covariance matrix is estimated as

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\hat{IC}_{h,\phi^*}(O_i|\hat{\beta}^{2,h})\hat{IC}_{h,\phi^*}(O_i|\hat{\beta}^{2,h})^{\top},$$

and a 95% confidence interval for $\beta_j^{2,h}$, for $j=1,\ldots,q$ is computed by $\hat{\beta}_j^{2,h} \pm 1.96\cdot\hat{\sigma}/\sqrt{n}$ where $\hat{\sigma}$ is the appropriate diagonal element of $\hat{\Sigma}$.

An important note is that the AIPTW estimator is dependent on the specification of a subclass of parameter models. We have already discussed the selection of ϕ^* to correspond to the projection of the IPTW estimator on T_{prop}. Two additional nuisance parameters of concern are g and $E(\epsilon(\beta)|A,W)$. A desirable property of the AIPTW is that it is doubly robust in the sense that only one of the models for g or $E(\epsilon(\beta)|A,W)$ needs to be correctly specified to achieve consistency [7].

4.1 Efficiency of AIPTW estimators

We performed a simulation exercise to compare the range of limiting variances for $\hat{\beta}^{2,h}$ for different choices of h. Consider the data given by $O=(Y,A,W)$ where W is a univariate real valued variable that represents a potential confounder. We are interested in the following linear MSM, $E(Y_a)=\beta_0+\beta_1 a$ and our parameter

Table 1: Empirical variance calculations for $\hat{\beta}_1^{2,h}$ with $\tilde{\alpha} = 0$.

(h_0, h_1)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	-	617.73	106.76	32.48	13.35	7.63	6.22	6.39	7.16	8.14	9.16
0.1	542.38	-	197.01	27.07	8.39	6.17	7.12	8.70	10.29	11.74	13.02
0.2	77.67	156.00	-	14.57	6.95	10.78	13.89	16.13	17.79	19.05	20.04
0.3	18.81	15.16	7.92	-	70.43	46.41	39.65	36.50	34.68	33.50	32.68
0.4	7.39	6.17	12.22	98.14	-	364.58	145.44	94.99	74.00	62.76	55.84
0.5	6.30	8.80	20.02	68.85	426.65	-	897.03	304.04	176.81	126.40	100.38
0.6	7.98	12.66	25.11	60.33	185.06	993.45	-	1667.77	522.22	285.11	193.69
0.7	10.35	16.18	28.55	56.31	127.13	360.84	1798.54	-	2676.80	799.97	419.88
0.8	12.77	19.15	31.00	53.97	102.40	220.40	596.20	2841.93	-	3924.12	1137.29
0.9	15.03	21.64	32.83	52.44	88.92	163.39	340.15	891.12	4123.61	-	5409.74
1.0	17.09	23.73	34.24	51.36	80.50	133.40	239.26	486.38	1245.62	5643.58	-

of interest is β_1. Since A is binary, $h(A)$ is a two-dimensional vector that is completely determined by two numbers, say h_0 and h_1. Define

$$h(A) = \begin{pmatrix} 1 \\ h_0 I(A=0) + h_1 I(A=1) \end{pmatrix}.$$

A collection of AIPTW estimators was created by considering all parings of h_0 and h_1 where $h_0 \neq h_1$ and $h_0, h_1 \in \{0.0, 0.1, 0.2, \ldots, 1.0\}$. Data was generated in the following manner:

$$\begin{aligned} P(W=1) &= 0.5 \\ \text{logit}(P(A=1|W)) &= -0.5 + \tilde{\alpha} W \\ Y_A &= 1 + A + W + \epsilon, \epsilon \sim N(0,1). \end{aligned}$$

We generated three data sets, each consisting of one million observations, with varying degrees of confounding. The level of confounding was determined by the parameter $\tilde{\alpha}$, and we considered $\tilde{\alpha} = 0.0, 0.7$ and 1.1. All necessary components for the influence curve of each AIPTW estimator were calculated explicitly using the known data generating mechanism. For each estimator, $\hat{\beta}^{2,h}$, the empirical covariance matrix was calculated, and we report the limiting variance for $\hat{\beta}_1^{2,h}$, the second diagnal element of the covariance matrix.

In Tables 1-3 we report empirical variance calculations for $\hat{\beta}_1^{2,h}$ for each pairing of h_0 and h_1. To quantify the range in efficiency over the collection of estimators we compute the ratio of the variance of each pairing to the minimum variance within each data set. We report quantile information for this ratio in Table 4, and observe that the range in relative efficiency is quite large, with the maximum greater than 300 for all data sets. We also observe that the estimator producing the smallest variance is at least as four times as efficient for half of the estimators ($\tilde{\alpha}$=0.7 and 1.1), and at least eight times as efficient as half of the estimators when there is no confounding. It appears that the selection of h can indeed have a significant effect on the efficiency of the AIPTW estimator. For an implicit characterization of the optimal solution h_{opt}, we refer the reader to Robins (1998).

5 Simulation Study

As of yet, no practical comparisons have been made between the IPTW and AIPTW estimators. We consider estimation of causal parameters of linear MSMs and present three different simulation studies to illustrate the applicability of these estimators. For all simulations the observed data is given as $O = (Y, A, W = (V, W_1))$, where V and W_1 are both univariate real-valued covariates and the MSM is given by (1) with $m(A, V|\beta) = \beta_0 + \beta_1 A + \beta_2 V$. In this scenario, β_1 measures the causal effect of treatment and is the primary parameter of interest.

Table 2: Empirical variance calculations for $\hat{\beta}_1^{2,h}$ with $\tilde{\alpha} = 0.7$.

(h_0, h_1)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	-	1165.51	226.57	76.86	32.63	15.88	8.80	5.74	4.54	4.25	4.43
0.1	1139.25	-	638.07	113.65	35.10	13.87	6.91	4.67	4.26	4.59	5.23
0.2	215.16	618.71	-	269.11	40.36	10.95	5.02	4.28	4.95	6.01	7.11
0.3	70.41	105.70	256.64	-	58.61	6.68	4.40	6.08	7.98	9.63	10.99
0.4	28.65	30.95	35.85	53.05	-	6.59	12.62	15.46	17.03	18.03	18.71
0.5	13.39	11.62	9.09	5.62	7.93	-	113.04	58.18	44.13	37.90	34.41
0.6	7.30	5.80	4.49	4.84	15.02	121.28	-	377.97	143.36	90.41	68.66
0.7	4.95	4.32	4.54	7.27	18.21	64.03	393.10	-	801.36	268.16	154.29
0.8	4.28	4.45	5.75	9.63	19.96	49.18	152.65	823.40	-	1383.23	432.57
0.9	4.40	5.19	7.18	11.58	21.06	42.54	97.75	280.90	1412.17	-	2123.58
1.0	4.91	6.14	8.57	13.15	21.81	38.82	75.04	163.94	448.76	2159.42	-

Table 3: Empirical variance calculations for $\hat{\beta}_1^{2,h}$ with $\tilde{\alpha} = 1.1$.

(h_0, h_1)	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	-	1622.67	326.46	115.00	50.63	25.25	13.84	8.39	5.79	4.63	4.26
0.1	1632.22	-	1008.52	191.41	63.19	26.11	12.51	7.05	4.91	4.28	4.35
0.2	330.73	1016.05	-	540.21	92.81	27.58	10.70	5.61	4.31	4.41	5.04
0.3	117.51	194.66	545.72	-	217.75	30.68	8.18	4.41	4.54	5.62	6.88
0.4	52.26	65.02	95.06	221.23	-	41.12	5.01	4.99	7.23	9.30	10.98
0.5	26.35	27.23	28.74	31.91	42.57	-	10.33	15.79	18.00	19.17	19.90
0.6	14.59	13.21	11.32	8.67	5.23	9.76	-	125.38	63.04	47.21	40.22
0.7	8.89	7.46	5.90	4.52	4.80	15.00	122.79	-	386.27	146.74	92.63
0.8	6.09	5.12	4.38	4.42	6.84	17.13	61.23	381.65	-	793.01	266.91
0.9	4.79	4.33	4.33	5.36	8.78	18.27	45.67	143.93	786.36	-	1345.58
1.0	4.31	4.29	4.85	6.51	10.38	18.98	38.82	90.41	263.08	1336.90	-

Table 4: Quantile information for $\hat{\mathrm{Var}}(\hat{\beta}_1^{2,h}) / \min_{h_0,h_1} \hat{\mathrm{Var}}(\hat{\beta}_1^{2,h})$.

$\tilde{\alpha}$	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	2.38	8.41	64.35	34.77	914.68
0.7	1.47	4.26	38.92	24.40	508.10
1.1	1.37	4.12	33.90	21.78	383.15

The specific estimators we consider are $\hat{\beta}^{1,h^*}$ and $\hat{\beta}^{2,h^*}$. For both estimators and for all simulations $g_n(A|V)$ and $g_n(A|W)$ were computed using a logistic regression model. To calculate an estimate of $E[\epsilon^2(\beta)|A,V]$, $\epsilon(\hat{\beta})$ was first computed with an initial estimate of β obtained by performing a WLS regression of Y on A and V with weights $g_n(A|V)/g_n(A|W)$. Next, $E[\epsilon^2(\hat{\beta})|A,V]$ was estimated using a generalized additive model. $\hat{\beta}^{1,h^*}$ was computed using the WLS procedure described in Section 3.1 with standard errors calculated empirically from the estimated influence function. To compute $\hat{\beta}^{2,h^*}$ a Newton-Raphson procedure was used with $E[\epsilon(\beta)|A,W]$ estimated using the linear model $\gamma_0+\gamma_1 A+\gamma_2 V+\gamma_3 W_1$. For all simulations the number of iterations was one-thousand. The measures of performance reported in the following examples are bias, mean squared error(mse), relative mean squared error, given by $\text{mse}(\hat{\beta}^{1,h^*})/\text{mse}(\hat{\beta}^{2,h^*})$, and 95% coverage probabilities.

Example 1 demonstrates the gains in efficiency made by the AIPTW estimator over the IPTW when applied to a data set of a moderate size. In Example 2 we consider causal parameter estimation when the treatment mechanism is misspecified. We see in this scenario that not only is the IPTW less efficient, but inconsistent. Furthermore, the AIPTW proves to be consistent. Example 3 inspects the performance of the estimators, in particular the AIPTW, when the treatment mechanism is modelled correctly but, the model for $E(\epsilon(\beta)|A,W)$ is misspecified. Here we see that the AIPTW retains its consistency properties.

Example 1: Varying the Level of Confounding

The data was generated so that both V and W_1 confounded the effect of treatment. The data generating distribution has

$$\begin{aligned} P(V=1) &= 0.5 \\ W_1 &\sim \text{Unif}(1,3) \\ \text{logit}(P(A=1|W)) &= -3+2V+W_1 \\ Y_A &= 4+2A+V+\tilde{\beta}W_1+\epsilon, \quad \epsilon \sim N(0,1). \end{aligned}$$

We conducted simulations for a fixed sample size of $n=5000$ while varying the parameter $\tilde{\beta}=(0.5,1.0,2.0,4.0,6.0)$. The results, presented in Table 5, validate the theory that while both estimators are unbiased, the AIPTW is more efficient than the IPTW, this is reflected in the coverage probabilities and the relative mean squared error, which increases when the effect of W_1 on the outcome becomes stronger.

Table 5: Linear MSM, varying the level of confounding.

	IPTW			Augmented IPTW		
$\tilde{\beta}$	RMSE	Bias	Cov(%)	$n\cdot$MSE	Bias	Cov(%)
0.5	1.006	-0.001	94.1	6.112	-0.001	92.8
1.0	1.009	-0.001	98.3	5.814	-0.001	95.2
2.0	1.095	0.002	99.6	5.414	0.0004	96.0
4.0	1.353	0.003	100.0	5.280	0.001	96.1
6.0	1.761	0.003	100.0	5.812	0.002	94.1

Example 2: Misspecified Model for Propensity Score

Data for this example was generated in such a manner that estimating the treatment mechanism with a logistic model would be incorrect. The data generating distribution has

$$\begin{aligned} P(W_1=1) &= 0.5 \\ P(A=1|W) &= 0.1+0.25VW_1+0.6W_1, \\ Y_A &= 1+2A+V+7W_1+\epsilon, \quad \epsilon \sim N(0,1), \end{aligned}$$

with V generated in the same manner as the previous example. The results are presented in Table 6 are for a range of sample sizes ($n = 500, 1000, 2000, 5000,$ and 7000). For all sample sizes the IPTW is clearly biased, and is outperformed by the AIPTW estimator in terms of bias and efficiency.

Table 6: Linear MSM, misspecification of the propensity score.

	IPTW			Augmented IPTW		
n	RMSE	Bias	Cov(%)	n·MSE	Bias	Cov(%)
500	4.780	0.230	99.8	12.640	-0.002	94.1
1000	6.272	0.223	99.7	13.082	0.003	95.0
2000	9.650	0.207	99.6	11.826	-0.00003	95.4
5000	18.976	0.197	95.5	11.722	0.003	95.7
7000	21.897	0.193	91.2	13.312	-0.002	94.1

Example 3: Misspecification of $E(\epsilon|A, W)$

For this example an interaction term between A, V and W_1 was incorporated into the mean model generating distribution for Y. For the AIPTW estimator the model for $E(\epsilon|A, W)$ was specified in the same manner as for the first two examples, meaning no interaction term was included. We wanted to validate the assertion that the AIPTW would prove to still be a consistent estimator if g was correctly specified while $E(\epsilon|A, W)$ was not. The data generating distribution has

$$\begin{aligned} Y_A &= 4 + 2A + 0.5V + W_1 + AV(W_1 - 2) + \epsilon, \epsilon \sim N(0, 1) \\ \text{logit}(P(A = 1|W)) &= -2 + 2V + W_1, \end{aligned}$$

with V and W_1 generated in the same manner as Example 1. Table 7 presents the results for various sample sizes, and indeed the augmented IPTW estimator proved to be consistent in spite of the model for $E(\epsilon|A, W)$ being misspecified.

Table 7: Linear MSM, misspecification of $E(\epsilon|A, W)$.

	IPTW			Augmented IPTW		
n	RMSE	Bias	Cov(%)	n·MSE	Bias	Cov(%)
500	1.012	-0.002	97.0	6.877	-0.004	94.8
1000	1.015	-0.001	97.8	6.271	-0.003	96.0
2000	0.986	0.0002	98.3	6.118	-0.00004	96.0
5000	1.006	0.001	97.1	6.455	0.001	95.7
7000	1.000	-0.0002	97.5	6.333	-0.0004	95.6

6 Exercise and the risk of cardiovascular conditions in older adults

6.1 SPPARCS Data

We apply our methods to data collected from The Study of Physical Performance and Age-Related Changes in Sonomans (SPPARCS). This data arises from a study designed to examine the relationship between physical functionality and aging in a cohort of individuals over the age of fifty living in and around Sonoma, California. Specifically, we are interested in the relationship between exercise and the development of a

cardiac or non-cardiac cardiovascular condition. More details regarding the study design can be found elsewhere [11, 12].

Data was collected on 2092 individuals in a baseline interview and a follow-up interview given approximately two years later. A cardiovascular condition was defined as having any one of the following: a myocardial infarction, stroke, heart bypass, angioplasty of the heart, cartiod artery surgery, congestive heart failure, surgery to repair a blocked artery, angioplasty of a blocked artery or an aortic aneurysm. Cardiac conditions involve only the heart while non-cardiac cardiovascular conditions include all other diseases related to arteriosclerosis of blood vessels, such as strokes. Each subject was determined to belong to one of four levels of physical activity. The lowest level consists of anyone who engaged in moderate activities, biking on level ground, intramural activities, leisurely walking, light gardening. The second level consists of anyone who might be engaged in moderate activities, but is also engaged in brisk walking and/or digging in the garden and/or dancing, level three may or may not include any activities in the previous two levels but requires participation in one vigorous activity such as swimming, jogging, aerobics. The last level is the same as the previous level, except the requirement is participation in at least two vigorous activities.

There were several time-independent covariates measured at the baseline interview. Measured covariates included a self-reported measure of difficulty in performing everyday tasks (NRB), smoking status, if they were an ex-smoker, body mass index (BMI), age, indicator of a dramatic drop in activity in the past 5-10 years prior to baseline (decline), gender, a measure of consistent level of activity in a person's recollected history of their physical activity (past activity history), indicator of whether the individual participated in high school sports (team), self-perception of overall health, and the presence of a cardiovascular condition reported at baseline. Table 8 presents a list of frequencies of covariates by activity level. For the binary and ordinal variables p-values were calculated using a χ^2 test of independence. For continuous covariates, association with activity level was determined using standard linear regression.

6.2 Analysis

Let Y be the indicator of an individual first reporting a cardiovascular condition at the follow-up interview. Let A be an indicator of engaging in vigorous levels of activity (level three or four). The observed data is represented by $O = (Y, A, W)$, where W =(NRB, smoker, ex-smoker, BMI, age, decline, gender, past history, team, perception, baseline condition). The model we are interested in fitting is a logistic MSM represented by

$$E(Y_a) = [1 + \exp(-(\beta_0 + \beta_1 a))]^{-1}.$$

Our causal parameter of interest is the causal odds ratio $\exp(\beta_1)$.

We considered two estimators of β_1, $\hat{\beta}^{1,h^*}$ an IPTW estimator, and an AIPTW estimator $\hat{\beta}^{2,h^*}$, where from (4)

$$h^*(A) = \frac{g(A)}{\mathrm{Var}(\epsilon(\beta)|A)} \frac{d}{d\beta}[1 + \exp(-(\beta_0 + \beta_1 A))]^{-1}.$$

For both estimators, $g(A|W)$ was estimated using a logistic regression model, while $g(A)$ was estimated empirically from the data. For the AIPTW estimator, $E(\epsilon(\beta)|A, W)$ was estimated using a linear model. Both estimates of β were calculated using a Newton-Raphson algorithm with standard errors calculated from their respective estimated influence function.

6.3 Results

Subjects with missing data (N=76) were removed from the analysis, so calculations were performed on data from 2016 men and women ranging in ages from 53 to 96. Nearly, 60% of the subjects were women. At the baseline interview almost 19% of the subjects reported having at least one cardiovascular condition. Nearly 5% of subjects reported having a cardiovascular condition at the second interview.

The results are presented in Table 9. We observe that engaging in a vigorous level of exercise appears to have a protective effect against the risk of developing a cardiovascular condition, however, the effect was not shown to be statistically significant. We also computed odds ratios within different strata of the data.

Table 8: Individual characteristics by exercise level.

	Level of Exercise				
CHARACTERISTIC	1	2	3	4	p
Smoker					
No	268 (14.5)	538 (29.1)	616 (33.4)	425 (23.0)	0.0006
Yes	32 (18.9)	60 (35.5)	61 (36.1)	16 (9.5)	
Ex-smoker					
No	147 (14.4)	325 (31.9)	342 (33.5)	206 (20.2)	0.096
Yes	153 (15.4)	273 (27.4)	335 (33.6)	235 (23.6)	
Drop in activity					
No	97 (7.8)	379 (30.4)	430 (34.5)	339 (27.2)	<0.0001
Yes	203 (26.3)	219 (28.4)	247 (32)	102 (13.2)	
Gender					
Female	196 (16.3)	360 (30.0)	402 (33.5)	241 (20.1)	0.034
Male	104 (12.7)	238 (29.1)	275 (33.7)	200 (24.5)	
History of activity					
Not much	65 (21.3)	99 (32.5)	96 (31.5)	45 (14.8)	<0.0001
Some	87 (21.8)	123 (30.8)	132 (33)	58 (14.5)	
Habitual	148 (11.3)	376 (28.7)	449 (34.2)	338 (25.8)	
High school sports					
Did not participate	182 (16.5)	342 (31.1)	377 (34.2)	200 (18.2)	<0.0001
Participated	118 (12.9)	256 (28)	300 (32.8)	241 (26.3)	
Perception of health					
Excellent/good	189 (11.3)	508 (30.4)	579 (34.6)	396 (23.7)	<0.0001
Fair/poor	111 (32.3)	90 (26.2)	98 (28.5)	45 (13.1)	
Cardiovascular					
None reported	223 (13.4)	502 (30.2)	561 (33.8)	374 (22.5)	0.001
At least one	77 (21.6)	96 (27.0)	116 (32.6)	67 (18.8)	
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	
NRB score	0.67 (0.32)	0.87 (0.19)	0.86 (0.21)	0.90 (0.12)	<0.0001
BMI	27.37 (5.20)	26.42 (4.42)	26.87 (4.65)	26.74 (4.26)	0.03
Age	73.86 (8.88)	69.98 (8.15)	69.74 (8.25)	67.95 (8.06)	<0.0001

Table 9: The effect of exercise on the risk of developing a cardiovascular condition.

Estimator	OR	CI (95%)
All subjects (N=2016)		
IPTW	0.80	(0.52, 1.23)
AIPTW	0.80	(0.52, 1.22)
Men (N=817)		
IPTW	0.92	(0.49, 1.73)
AIPTW	0.93	(0.50, 1.72)
Women (N=1199)		
IPTW	0.71	(0.39, 1.28)
AIPTW	0.71	(0.40, 1.27)
No condition reported at baseline (N=1660)		
IPTW	0.78	(0.47, 1.31)
AIPTW	0.78	(0.47, 1.30)
Condition reported at baseline (N=356)		
IPTW	0.89	(0.40, 1.96)
AIPTW	0.89	(0.40, 1.97)

It appears that women receive a greater benefit than men by engaging in moderately vigorous exercise, although again the protective effect of exercise is not statistically significant. We also compared odds ratios between individuals who reported having a cardiovascular condition at baseline, and those who did not. As may be expected, those not reporting a condition at baseline appear to benefit slightly more from moderately vigorous exercise, but in both stratum the effect was not statistically significant. There were no remarkable differences in estimates produced by the IPTW and the AIPTW estimators.

Acknowledgements

The authors wish to thank Ira Tager for providing the data set.

References

[1] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner. *Efficient and Adaptive Inference in Semiparametric Models*. Baltimore, Maryland:Johns Hopkins University Press, 1993.

[2] C. Drake. Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics*, 49:1231–1236, 1993.

[3] M. A. Hernan, B. Brumback, and J. M. Robins. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448, 2001.

[4] P. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81:945–970, 1986.

[5] J. Robins. Marginal structural models. In *1997 Proceedings of the American Statistical Association on Bayesian Statistical Science*, pages 1–10, 1998.

[6] J. Robins. Marginal structural models versus structural nested models as tools for causal inference. Technical report, Technical Report Series; Spring 1998 Symposium on Prospects for a Common Sense Theory of Causation. Stanford, CA, 1998.

[7] J. Robins. Robust estimation in sequentially ignorable missing data and causal inference models. In *1999 Proceedings of the American Statistical Association on Bayesian Statistical Science*, 1999.

[8] J. Robins, A. Rotnitzky, and M. van der Laan. Comment on "Profile Likelihood". *Journal of the American Statistical Association*, 95(450):477–482, 2000.

[9] J. M. Robins, S. D. Mark, and W. K. Newey. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48:479–495, 1992.

[10] P. R. Rosenbaum and D. B. Rubin. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524, 1984.

[11] W. Satariano, J. Smith, A. Swanson, and T. IB. A census-based design for recruitment of community sample of older residents: Efficacy and costs. *Annals of Epidemiology*, 8:278–82, 1998.

[12] I. Tager, M. Hollenberg, and W. Satariano. Self-reported leisure time physical activity and measures of cardiorespiratory fitness in an elderly population. *American Journal of Epidemiology*, 147:921–931, 1998.

[13] M. J. van der Laan and J. Robins. *Unified Methods for Censored Longitudinal Data and Causality*. To appear. Springer-Verlag, 2002.