Estimation of the Average Treatment Effect with Missing Outcome Data

Feng Han

Teachers' College of Beijing Union University, Beijing, 100011, China

Abstract—**Missing outcome data occurs often in the causal inference of observational studies. For example, in observational study on the safety of a Traditional Chinese Medicine (TCM) injection in market, some patients are missing safety outcome variables. In this paper we proposed a consistent estimator for the average treatment effect (ATE) with partially missing outcome data and derived the asymptotic variance of the proposed ATE estimator under the condition that the missing-data mechanism is missing at random (MAR). We then proceeded to assess the performance of the asymptotic variance estimator via a simulation study. The simulation study showed that the asymptotic variance estimator had good performance in finite sample sizes. This asymptotic variance could then be used to construct a confidence interval for the average treatment effect. We also compared the bias and mean squared error (MSE) of the ATE estimators based on the proposed method-dealing with the missing outcomes with that of the complete-data method, which means that directly deleting the samples with the missing data through a simulation study. The simulation study showed that the MSE and Bias of our method were smaller than the complete-data method under MAR. In addition, we also found that coverage of the confidence interval constructed on the ATE and its asymptotic variance from our method are better than those based on the traditional method.**

Keywords—average treatment effects (ATE); asymptotic variance; delta-method; missing data

I. INTRODUCTION

Recently, there has been a surge in theoretical work focusing on estimating ATE without missing outcome data. For estimation of the ATE, various semi-parametric estimation methods have been proposed, including the regression methods (Heckman et al., 1997; Hahn, 1998; Imbens et al., 2003)methods using the propensity score such as weighting (Hirano et al., 2003), matching on covariates (Rosenbaum, 1995; Abadie and Imbens, 2002), blocking (Rosenbaum and Rubin, 1984), and combinations of those approaches(Graf, 1997).

The propensity score methodology is the most commonly used method among the aforementioned approaches. Hirano et al. (2003) gave the most complete theory on the use of the propensity score methodology when the propensity score is estimated. Montes-Rojas (2009) derived the asymptotic variance of the ATE estimator with the propensity score is estimated by the M-estimator using the Delta method (Wooldridge, 2002). In practice, this asymptotic variance can be used to measure the accuracy of the ATE estimators and to construct confidence interval for the ATE.

There can be a large amount of missing data in observational studies of medicine, however. For example, in the observational study on the safety of a Traditional Chinese Medicine (TCM) injection, some patients are missing safety outcome variables.

In cases where some subject outcomes are missing, the most commonly used technique to estimate the ATE and its associated variance is to exclude subjects with missing outcomes, then perform analysis of the remaining data. This approach is called a complete-case analysis.

When subjects with any missing outcomes are excluded, it is well known that a complete-case analysis can give highly inefficient estimators and potentially biased results. Thus, the method of directly deleting missing data will result in a significant loss of information.

Among various kinds of method for dealing with missing data, one common method is the inverse probability weighting method (IPW). Since Horvitz and Thompson (1952) first proposed to use this method to estimate population mean with nonrandomized missing data mechanisms in 1952, in effect, this approach did not get widely used for a very long time.

It was not until the end of the twentieth Century that this approach has become popular when Robins and Rotnitzky (1995) proved that the inverse probability weighting estimator is a consistent and asymptotically normal estimator in a multivariate regression model and proposed a modified IPW estimator (Little, 1995). Wooldridge (2002) proposed the inverse probability weighted minimization estimator in order to analyze no-response questions in the investigation of a period or multi period (Little and Rubin, 1987). Carpenter et al. (2005) proved that the IPW estimator with double robustness through the empirical analysis (Ralph, 2000). These results make IPW method has received more and more attention.

To increase efficiency and reduce the bias, it is necessary to develop methods that incorporate the partially missing outcome data into the analysis. In this paper, we propose a method by providing formulas that can estimate ATE and its associated variance with missing outcome data. The proposed method is under the missing at random (MAR) assumption, which implies that the missing-data process can depend only on the observed covariates, but not on the missing outcome (Little, 1995; Little and Rubin, 1987; Ralph, 2000; Rubin, 1976). We propose a new ATE estimator under the MAR condition and prove that the new estimator was consistent. We also extend the asymptotic variance formula of the ATE

estimator on complete data by Montes-Rojas (2009) and derived the asymptotic variance formula of the ATE estimator on non-complete data with missing outcomes.

Our simulation study demonstrates the high accuracy of the proposed method in estimating the ATE and its asymptotic variance. In this paper, we also apply the proposed method to a post-surveillance TCM study. The paper is organized as follows. Section 2 provides notation and derives the ATE estimators with partially missing outcome data. Section 3 derives the asymptotic variance of the ATE estimators under the condition that missing-data mechanism is MAR. Section 4 evaluates the accuracy of the resulting asymptotic variance formula by comparing the empirically estimated variance with the variance estimate derived using the asymptotic formula in a simulation study and compares the MSE of the ATE estimators between the proposed method that addresses the missing outcomes and the conventional method that deletes the samples with the missing data. Then, we also compare the confidence interval coverage constructed by two different methods. We give some remarks and conclusion in Section 5.

II. AVERAGE TREATMENT EFFECT ESTIMATORS WITH MISSING OUTCOME DATA

In this section we derive the ATE estimators with partially missing outcome data. We follow the standard notation in Imbens (2004) and Montes-Rojas (2009). Consider N individuals indexed by $i = 1, \mathrm{K}, N$, who may receive a certain "treatment", indicated by the binary variable $W_i = 0,1$. Each individual has a pair of potential outcomes (Y_{0i}, Y_{1i}) that corresponds to the outcomes that the individual would have been observed with and without the receipt of the treatment, respectively.

The fundamental problem of causal inference is the inability to observe at the same time the outcomes of the same individual both with and without the receipt of treatment effects. That is, we can only observe $Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i}$ and a set of exogenous covariates X_i . We define the propensity score as $p(X_i) = p(W_i = 1 | X_i = x_i)$. We are interested in estimating the ATE, defined as $\delta = E(Y_{1i} - Y_{0i})$. In addition, we let R_i denote a dummy variable such that $R_i = 0$ when the outcome of the ith individual is missing; and $R_i = 1$ otherwise. We extend the idea in Montes-Rojas (2009) and derive the ATE estimators under the condition that the missing-data mechanism is MAR. To derive our estimators, we need to make the following assumptions.

Assumption 1 $0 < p(X_i) < 1$.

This assumption means that each individual in the population has a positive probability of receiving either treatment. In the study, the propensity score is usually obtained from logistic regression. As long as we focus on the

distribution of the propensity score, we may assess that whether the assumption 1 is feasible.

Assumption 2 $R_i \perp W_i \perp (Y_{0i}, Y_{1i}) \big| X_i$.

This assumption says that Y_{1i} and Y_{0i} are independent of W_i, R_i given X_i .This assumption ensures that we can stratified according to known covariates, and the causal effect of each layer is identifiable, similar to the negligible assumption for a complete data (Rosenbaum and Rubin1983). For the actual data, how do we verify that it meets this assumption? This depends on the background of the actual data. We generally based on the background of the data to determine the missing mechanism. As the examples in this article shows, we can consider the missing mechanism as missing at random. Thus, the missing indicator variables are independent with missing outcome variable, given the value of the covariates. This can be expressed as $R_i \perp Y_i | X_i$. Since $Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i}$, so apparently we can get $R_i \perp W_i | X_i$ and $R_i \perp (Y_{0i}, Y_{1i}) \big| X_i$.

Next, as long as we can verify $W_i \perp (Y_{0i}, Y_{1i}) \big| X_i$, we can determine that whether the assumption 2 is satisfied. The verification for this assumption is similar to the negligible assumption for a complete data. The difference between the two is that the data we are concerned is a data with missing outcome variable. The approach used in this paper is the inverse probability weighting method.

Proposition 1. Under Assumptions 1 and 2 and the condition that the missing- data mechanism is MAR, we can write δ as the expected value of a weighted equation.

The average treatment effect in the presence of missing outcome data can be written as follows:

$$\delta = \mathrm{E} \left(\frac{R_i W_i Y_i}{p(X_i) p(R_i = 1 | X_i)} - \frac{R_i (1 - W_i) Y_i}{(1 - p(X_i)) p(R_i = 1 | X_i)} \right).$$

The proof of this proposition is given in Appendix.

Based on the result in Proposition 1, we propose the following consistent estimator for δ , when the missing outcome data mechanism is MAR:

$$\hat{\delta} = N^{-1} \sum_{i=1}^{N} \left[\frac{R_i W_i Y_i}{p(X_i) p(R_i = 1 | X_i)} - \frac{R_i (1 - W_i) Y_i}{(1 - p(X_i)) p(R_i = 1 | X_i)} \right] \quad (1)$$

When propensity score and missing-data mechanism are known, we can prove that the estimator from equation (1) is an unbiased estimator for the ATE by the property of an identically independent sample.

It is noteworthy that in equation (1) the propensity score and missing-data mechanisms are usually unknown, and they

need to be estimated by means of a parameter model (such as logit or probit). It can be easily shown that the proposed estimator in equation (1) is a consistent estimator of ATE when the parameters in the propensity score are estimated by an M-estimator and the parameters in the missing-data mechanisms are also estimated by an M-estimator.

III. THE ASYMPTOTIC VARIANCE OF AVERAGE TREATMENT EFFECT ESTIMATORS WITH MISSING OUTCOME DATA

In practice, the propensity score $p(X_i)$ is usually unknown and need to be estimated based on data. Here we assume that $p(X_i) = \Phi(X_i; \gamma^*)$,

where γ^* denotes the true value of the parameters γ, in the propensity score.

Here we let $\hat{\hat{\delta}}$ denote a two-step estimator of δ. This means that the parameters γ in the propensity score are estimated using the M-estimator $\hat{\gamma}$. Generally, the M-estimator can be defined to be a zero of an estimating function (Van de Geer, 2000). This estimating function is the derivative of another statistical function. For example, a maximum-likelihood estimate is often defined to be a zero of the derivative of the likelihood function with respect to the parameter. Thus, a maximum-likelihood estimator is often a critical point of the score function (Ferguson, 1982).

Here δ is estimated via equation (1) with $p(X_i)$ being replaced by $\Phi(X_i; \hat{\gamma})$. We also let δ^* denote the true value of δ.

Next, we deal with the situation where the missing-data mechanism is MAR. In this case, we also need to model the missing data mechanism, $p(R_i = 1 | X_i)$.

We assume that $p(R_i = 1 | X_i) = \pi(X_i; \beta^*)$, where β^* represents the true value of β, and β denotes parameters to be estimated in the missing-data mechanism. Let $\hat{\beta}$ be the M-estimator of β, based on the above parametric model for $p(R_i = 1 | X_i)$.

Let $\hat{\hat{\hat{\delta}}}$ denote a three-step ATE estimator of δ^* that is given by formula (1) with both γ and β being replaced by the M-estimators $\hat{\gamma}$ and $\hat{\beta}$, respectively.

A. Theorem 1

When the missing-data mechanism is MAR, under the conditions that $p(X_i)$ can be estimated via a parameter model, we can derive the asymptotic variance of ATE estimator as follows:

$$\lim_{N \to \infty} VAR\left[\sqrt{N}\left(\hat{\hat{\delta}} - \delta^* \right) \right]$$

$$= \lim_{N \to \infty} VAR\left[\xi\left(Y_i, W_i, R_i, X_i; \theta^*, \delta^*\right) + T\left(\theta^*, \delta^*\right) D\left(W_i, R_i, X_i; \theta^*\right) \right]$$

Where

$$\theta^* = \begin{pmatrix} \gamma^* \\ \beta^* \end{pmatrix},$$

$$\xi\left(Y_i, W_i, R_i, X_i; \theta^*, \delta^*\right)$$

$$= \frac{R_i W_i Y_i}{p(X_i) p(R_i = 1 | X_i)} - \frac{R_i (1 - W_i) Y_i}{(1 - p(X_i)) p(R_i = 1 | X_i)} - \delta^*$$

$$= \frac{R_i W_i Y_i}{\Phi(X_i; \gamma^*) \pi(X_i; \beta^*)} - \frac{R_i (1 - W_i) Y_i}{(1 -)\Phi(X_i; \gamma^*) \pi(X_i; \beta^*)} - \delta^*$$

$$T\left(\theta^*, \delta^*\right) = E\left[\nabla_\theta \xi\left(Y_i, W_i, R_i, X_i; \theta^*, \delta^*\right) \right]$$

And $D\left(W_i, R_i, X_i; \theta^*\right)$ is the influence function of an M-estimator for the parameters θ when θ is equal to the true value θ^*.

The proof of this theorem is given in the Appendix.

In practice, a consistent estimator of this asymptotic variance can be obtained by the sample version:

$$Est.VAR\left[\sqrt{N}\left(\hat{\hat{\delta}} - \delta^* \right) \right]$$

$$= N^{-1} \sum_{i=1}^{N} \left[\xi\left(Y_i, W_i, R_i, X_i; \hat{\theta}, \hat{\hat{\delta}}\right) + T\left(\hat{\theta}, \hat{\hat{\delta}}\right) D\left(W_i, R_i, X_i; \hat{\theta}\right) \right]^2 \quad (2)$$

In the actual analysis, if the propensity score is estimated by the logistic regression, the estimated value of the propensity score can be expressed as $p(X) = \dfrac{e^{\gamma X}}{1 + e^{\gamma X}}$; the missing mechanism can also be estimated by the logistic regression that can be expressed as $p(R = 1 | X) = \dfrac{e^{\beta X}}{1 + e^{\beta X}}$. Then, we can get that

$$\xi\left(Y_i, W_i, R_i, X_i; \hat{\theta}, \hat{\hat{\delta}}\right)$$

$$= \frac{R_i W_i Y_i}{p(X_i)\, p(R_i = 1 | X_i)} - \frac{R_i(1-W_i)Y_i}{(1-p(X_i))\, p(R_i=1|X_i)} - \hat{\hat{\delta}}$$

$$= \frac{R_i W_i Y_i}{\dfrac{e^{\gamma X_i}}{1+e^{\gamma X_i}} \dfrac{e^{\beta X_i}}{1+e^{\beta X_i}}} - \frac{R_i(1-W_i)Y_i}{\left(1 - \dfrac{e^{\gamma X_i}}{1+e^{\gamma X_i}}\right)\dfrac{e^{\beta X_i}}{1+e^{\beta X_i}}} - \hat{\hat{\delta}}$$

$$T\left(\theta, \hat{\delta}\right) = E\left[\nabla_\theta \xi\left(Y_i, W_i, R_i, X_i; \hat{\theta}, \hat{\delta}\right)\right].$$

Here putting

$$\hat{\theta} = \begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix}$$

$$D\left(W_i, R_i, X_i; \hat{\theta}\right)$$

$$= \begin{pmatrix} \nabla_\gamma \left[W_i \ln \dfrac{e^{\gamma X_i}}{1+e^{\gamma X_i}} + (1-W_i)\ln\left(1 - \dfrac{e^{\gamma X_i}}{1+e^{\gamma X_i}}\right) \right]\Big|_{\gamma=\hat{\gamma}} \\ \nabla_\beta \left[W_i \ln \dfrac{e^{\beta X_i}}{1+e^{\beta X_i}} + (1-W_i)\ln\left(1 - \dfrac{e^{\beta X_i}}{1+e^{\beta X_i}}\right) \right]\Big|_{\beta=\hat{\beta}} \end{pmatrix}$$

And into the equation (2), we can get the asymptotic variance of the ATE estimation constructed by the formula (1). With estimator and its asymptotic variance, we can construct confidence intervals. The confidence interval of the real value of the ATE for the nominal level of $100 \times \alpha\%$ confidence can be expressed as $\hat{\hat{\delta}} \pm Z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{N}}$, where σ represents the estimated standard deviation for ATE estimation.

IV. SIMULATION STUDY

A. Simulation Setup

We conducted a simulation study to assess the performance of the proposed ATE estimator, called Method I, and its asymptotic variance formula in finite sample sizes.

We also compared the performance of Method I with the complete-data method, which deleted subjects with the missing data and then applied the methods of Hirano et al. (2003) and Montes-Rojas (2009) with the remaining complete data to estimate the ATE and computed the associated variance estimate. We call the conventional method Method II.

The simulation study used six covariates and 1000 randomly generated data sets for each of the three parameter configurations to be given in Table 1 below. The sample size for each simulated data set was set to be 1000, 2000 and 5000,

respectively. All statistical analyses were conducted using SAS software, version 9.2.

The parameters were set up as follows:

First, we generated the six covariates

$X_1, X_2, X_3, X_4, X_5, X_6$. The covariate X_1 and X_2 were drawn from the normal distribution with the mean and variance, $(\mu_1, \ \sigma_1^2)$ and $(\mu_2, \ \sigma_2^2)$, respectively. The covariates X_3, X_4, X_5, X_6 were drawn from a Bernoulli distribution with p_3, p_4, p_5, p_6 , respectively.

We chose three different parameter settings for the parameters, $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, p_3, p_4, p_5, p_6$,

And those values are listed in Table I.

TABLE I. PARAMETER SETTINGS FOR SIMULATION

Settings	μ_1	σ_1^2	μ_2	σ_2^2	p_3	p_4	p_5	p_6
1 0.600	1	0.010 0.600	0.600	0.040		0.200	0.800	0.300
2 0.400	1	0.010 0.400	0.600	0.040		0.200	0.600	0.250

Second, we generated two potential outcomes Y_1 and Y_0 for each subject in the sample according to the following model:

$$Y_j = \alpha_{0(j)} + \alpha_{1(j)} X_1 + \alpha_{2(j)} X_2 + \alpha_{3(j)} X_3 + \alpha_{4(j)} X_4 + \alpha_{5(j)} X_5 + \alpha_{6(j)} X_6 + \varepsilon_j, j = 0, 1,$$

Where

$$[\alpha_{0(0)}, \alpha_{0(1)}] = [0,0]; [\alpha_{1(0)}, \alpha_{1(1)}] = [0.02, 0.03]; [\alpha_{2(0)}, \alpha_{2(1)}] = [0.1, 0.1]; [\alpha_{3(0)}, \alpha_{3(1)}] = [0.01, 0.02]; [\alpha_{4(0)}, \alpha_{4(1)}] = [0.2, 0.2]; [\alpha_{5(0)}, \alpha_{5(1)}] = [0.02, 0.02]$$

And ε_j was drawn from the normal distribution with $\mu = 0$ and $\sigma^2 = 1$

Third, the treatment assignment W for each subject was generated by the following model:

$$p(W = 1 | X) = \frac{e^{\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 + \gamma_6 X_6}}{1 + e^{\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 + \gamma_4 X_4 + \gamma_5 X_5 + \gamma_6 X_6}},$$

Where

$$\gamma_0 = 0, \gamma_1 = 0.2, \gamma_2 = 0.2, \gamma_3 = 3, \gamma_4 = 0.4, \gamma_5 = 0.2, \gamma_6 = 0.3$$

The value of the generated W for a subject determines which of the two potential outcomes, Y_1 and Y_0, is observed for this subject. If W=1, Y_1 is observed, and if W=0, Y_0 is observed.

Up to this point, we have randomly generated complete data sets under the three different parameter configurations.

Fourth, we generated the missing outcomes, according to the MAR missing data. Under the MAR missing data mechanism, we generated a missing-data indicator for each subject, R , from a Bernoulli distribution with the following event probability with covariates $X_1, X_2, X_3, X_4, X_5, X_6$:

$$p(R=1|X) = \frac{e^{\eta_0 + \eta_1 X_1 + \eta_2 X_2 + \eta_3 X_3 + \eta_4 X_4 + \eta_5 X_5 + \eta_6 X_6}}{1 + e^{\eta_0 + \eta_1 X_1 + \eta_2 X_2 + \eta_3 X_3 + \eta_4 X_4 + \eta_5 X_5 + \eta_6 X_6}},$$

Where
$$\eta_0 = 0, \eta_1 = 0.02, \eta_2 = 0.6, \eta_3 = 1, \eta_4 = 0.05, \eta_5 = 0.2, \eta_6 = 0.1$$

For each of the three parameter configurations, we repeated the above process 1000 times. Then, we got 1000 randomly generated data sets under each of the three parameter settings.

To evaluate the effect of the sample size on the accuracy of the estimated variance of the ATE estimator, we further considered four different sample sizes, 1000, 2000 and 5000, for each of the three parameter settings in Table I.

B. Performance Measures

For comparison of point estimators for the ATE, we used bias and mean squared error (MSE). We define these quantities below.

Let $\hat{\hat{\delta}}_m$ and $\hat{\delta}_{Z_m}$ be the estimates for theATE, δ^* , derived using the proposed method and the complete case method, respectively, using the mth simulated data set. Then, the biases of these ATE estimators are given, respectively, as follows:

$$Bias = \frac{1}{1000}\sum_{m=1}^{1000}\left(\hat{\hat{\delta}}_m - \delta^*\right) \quad Bias = \frac{1}{1000}\sum_{m=1}^{1000}\left(\hat{\delta}_{Z_m} - \delta^*\right).$$

The MSE of the ATE estimators are given, respectively, as follows:

$$MSE = \frac{1}{1000}\sum_{m=1}^{1000}\left(\hat{\hat{\delta}}_m - \delta^*\right)^2,$$
$$MSE = \frac{1}{1000}\sum_{m=1}^{1000}\left(\hat{\delta}_{Z_m} - \delta^*\right)^2.$$

We assessed the performance of the asymptotic variance estimator of the ATE, by comparing it with the empirical variances estimators.

For the estimator for the ATE proposed in this paper, the empirical variance is defined as follows:

$$S^2 = \frac{1}{999}\sum_{m=1}^{1000}\left(\hat{\hat{\delta}}_m - \overline{\hat{\delta}}_m\right)^2,$$

The derived asymptotic variance formula in this article is as follows:

$$Est.VAR\left(\hat{\hat{\delta}}\right) = \frac{1}{1000}\sum_{m=1}^{1000}VAR\left(\hat{\hat{\delta}}_m\right)$$
$$= \frac{1}{1000}\sum_{m=1}^{1000}\left\{N^{-2}\sum_{i=1}^{N}\left[\xi\left(Y_i, W_i, R_i, X_i; \theta, \hat{\hat{\delta}}_m\right) + T\left(\theta, \hat{\hat{\delta}}_m\right)D\left(W_i, R_i, X_i; \theta\right)\right]^2\right\}$$

C. Simulation Results

The simulation results are summarized in Table II and Table III.

TABLE II. SIMULATION RESULTS: EMPIRICAL VARIANCE VS. VARIANCE USING THE FORMULA

Parameter setting	Sample Size	Empirical variance(S2)	Variance using the formula
1	1000	0.025297	0.027354
	2000	0.011545	0.011372
	5000	0.004210	0.004283
2	1000	0.021957	0.022679
	2000	0.009939	0.009979
	5000	0.003447	0.003763

TABLE III. SIMULATION RESULTS: BIAS AND MSE OF ATE ESTIMATOR

Parameter setting	Sample Size	Method	Bias	MSE
1	1000	I	-0.006010	0.025308
		II	-0.008010	0.033745
	2000	I	0.002377	0.011539
		II	0.003405	0.015930
	5000	I	0.000463	0.004206
		II	0.000545	0.005628
2	1000	I	-0.000850	0.021935
		II	-0.002990	0.031240
	2000	I	0.002734	0.009936
		II	0.004447	0.013400
	5000	I	0.000445	0.003443
		II	0.000990	0.004453

TABLE IV. COVERAGE PROBABILITIES OF THE 95% CONFIDENCE INTERVALS

Parameter setting	Sample Size	Method	Coverage probabilities
1	1000	I	0.946
		II	0.893
	2000	I	0.948
		II	0.904
	5000	I	0.954
		II	0.904
2	1000	I	0.951
		II	0.890
	2000	I	0.955
		II	0.885
	5000	I	0.960
		II	0.897

Note: *Method I, calculated with the formula constructed in this paper;

#Method II, calculated by the method of directly deleting samples with missing outcomes.

By comparing the empirical variance with the variances obtained with the formula in Table II, we can conclude that the asymptotic variance formula of the proposed ATE

estimator under MAR missing data provides an accurate variance estimation of the estimated ATE.

In addition, from the bias and MSE results of the ATE estimators in Table III, we can conclude that the proposed method in this paper (Method I) outperforms the complete case method (the method of directly deleting patients with missing outcomes).We can also see from Table IV that coverage of the confidence interval constructed on the ATE and its asymptotic variance from method I with 95% confidence level is greater than that based on the traditional method. Moreover, the explicit variance estimator for the proposed ATE estimator also performs well in the sample sizes considered in the simulation study.

V. CONCLUSION AND DISCUSSION

For the analysis of observational data with a partially missing outcome variable, in this article we proposed an estimator for the ATE when the missing outcome is MAR, and also derived the associated variance formula. Furthermore, we verified the accuracy of the variance formula via the simulation study. The advantage of having a variance formula, such as this, is simple to calculate the variance estimate in practice. We compared the proposed method with the commonly used complete-case method and showed our method outperforms the commonly used method.

In this paper, we used a parametric model to estimate the propensity score and the missing-data mechanism. The disadvantages of using the parametric model are that modeling needs to be based on certain assumptions. A further research idea might be to relax these parametric model assumptions by using non-parametric models for the propensity score with the missing-data outcome data.

In other words, we first estimated propensity score with missing outcome data, and then get the estimator of ATE by inverse probability weighting. Furthermore, we can also use variance formula given in this paper to estimate the variance estimators of the ATE. So that we can make a reasonable statistical inference and this inference has very important reference value in practical research. It is worth mentioning that the assumption required by using the method of this article is relatively relaxed. From the practical examples given in this paper, we know that in the observational data analysis, assumption 1 generally can be met and assumption 2 is also possible to meet depends on the missing data mechanism. We can also say that whether the missing data mechanism satisfies MAR is important for assumption 2. Of course, we need to make a judgment based on a medical background in the actual study. In short, the method presented in this paper has a certain value.

REFERENCES

[1] Abadie, A. and Imbens, G. (2002).Simple and Bias-Corrected Matching Estimators for Average Treatment Effects. NBER technical working paper, No.283.

[2] D'Agostino, R. B. and Rubin, D. B. (2000).Estimating and Using Propensity Scores with Partially Missing Data. Journal of the American Statistical Association 95(451), 749-759.

[3] Ferguson, T. S. (1982).An inconsistent maximum likelihood estimate. Journal of the American Statistical Association 77 (380), 831–834.

[4] Graf, E. (1997).The Propensity Score in the Analysis of Therapeutic Studies. Biometrical Journal 39(3), 297–307.

[5] Hahn, J. (1998).On the Role of the Propensity Score in Efficient Semi-parametric Estimation of Average Treatment Effects. Econometrica 66,315–331.

[6] Heckman, J., Ichimura, H. and Todd, P. (1997).Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. Review of Economic Studies 64,605-654.

[7] Hirano, K., Imbens G. and Ridder G. (2003).Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica 71, 1161–1189.

[8] Horvitz, D. and Thompson, D. (1952).A Generalization of Sampling without Replacement from a Finite Population. Journal of the American Statistical Association 47,663 - 685.

[9] Imbens, G. W. (2004).Nonparametric estimation of average treatment effects under exogeneity: a review. Review of Economics and Statistics 86(1), 4-29.

[10] Imbens, G. W., Newey, W. and Ridder, G. (2003). Mean-Squared-Error Calculations for Average Treatment Effects. Department of Economics, UC Berkeley, unpublished manuscript.

[11] Little, R. J. (1995).Modeling the dropout mechanism in repeated-measures studies. Journal of the American Statistical Association 90, 1112-1121.

[12] Little, R. J. and Rubin, D. B. (1987).Statistical Analysis with Missing Data. J. Wiley & Sons, New York.

[13] Montes-Rojas, G. (2009).A note on the variance of average treatment effects estimators. Economics Bulletin 29(4), 2937-2943.

[14] Robins, J. M. and Rotnitzky, A. (1995).Semiparametric Efficiency in Multivariate Regression Models with Missing Data. Journal of the American Statistical Association 90,122-129.

[15] Rosenbaum, P. (1995).Observational studies. New York: Springer-Verlag.

[16] Rosenbaum, P. and Rubin, D. (1984).Reducing the Bias in Observational Studies Using Sub-classification on the Propensity Score. Journal of the American Statistical Association 79,516–524.

[17] Rubin, D. B. (1976).Inference and missing data (with discussion).Biometrika 63,581–592.

[18] Van de Geer, S.A. (2000).Empirical Processes in M-estimation: Applications of empirical process theory, volume 6 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[19] Wooldridge, J. M. (2002).Econometric analysis of cross section and panel data. The MIT Press, Cambridge, Massachusetts.