Simulation Study of Estimators for Treatment Effect in a Pretest-Posttest Study with Posttest Response Missing at Random

Marie Davidian, Anastasios A. Tsiatis, and Selene Leon

Department of Statistics

North Carolina State University

1 INTRODUCTION

This paper reports the results of selected simulations carried out to assess the performance of various estimators for treatment effect in a pretest-posttest study when the posttest response is missing at random (MAR), including several suggested by the results in Davidian, Tsiatis, and Leon (2004). The presentation assumes the reader is familiar with the material in the main narrative of Davidian et al. (2004).

We describe the results of simulations performed under two different simulation scenarios, each involving 1000 Monte Carlo replications, $\delta = 0.5$, $Y_1 \sim \mathcal{N}(\mu_1 = 0, \sigma_1^2 = 1)$, and binary X_2 with $P(X_2 = 1|Y_1, Z) = (1-Z)/[1+\exp\{-(\kappa_0^{(0)} + \kappa_1^{(0)} Y_1)\}] + Z/[1+\exp\{-(\kappa_0^{(1)} + \kappa_1^{(1)} Y_1)\}]$, with $(\kappa_0^{(0)}, \kappa_1^{(0)}) = (1, -0.1)$ and $(\kappa_0^{(1)}, \kappa_1^{(1)}) = (-0.9, 0.1)$. For simplicity, no baseline covariate X_1 was included. For each scenario, $P(R = 1|Y_1, X_2, Z) = Z\pi^{(1)}(Y_1, X_2) + (1 - Z)\pi^{(0)}(Y_1, X_2) = (1-Z)/[1+\exp\{-(\gamma_0^{(0)} + \gamma_1^{(0)} Y_1 + \gamma_2^{(0)} X_2 + \gamma_3^{(0} X_2 Y_1)\}] + Z/[1+\exp\{-(\gamma_0^{(1)} + \gamma_1^{(1)} Y_1 + \gamma_2^{(1)} X_2 + \gamma_3^{1)} X_2 Y_1)\}]$, with $\gamma^{(0)} = (\gamma_0^{(0)}, \gamma_1^{(0)}, \gamma_2^{(0)}) = (0.2, 2.0, 0.1, 0.1)$ and $\gamma^{(1)} = (\gamma_0^{(1)}, \gamma_1^{(1)}, \gamma_2^{(1)}) = (1.0, -1.0, -0.1, 0.1)$, resulting in 34% and 20% of the cases missing follow-up response for treatment and control.

For each data set, β was estimated several ways. Naive inference was obtained based on complete cases only using several popular approaches: the two-sample t-test estimator; the paired t-test estimator $\sum_{i=1}^n R_i Z_i (Y_{2i} - Y_{1i})/n_{R1} - \sum_{i=1}^n R_i (1-Z_i)(Y_{2i} - Y_{1i})/n_{R0}$; the estimator ANCOVA I obtained as the coefficient of Z from ordinary least squares (OLS) regression of Y_2 on $(Y_1, Z)^T$; and the estimator ANCOVA II obtained as the coefficient of $Z - \widetilde{Z}$ from the OLS regression of $Y_2 - \widetilde{Y}_2$ on $\{(Y_1 - \widetilde{Y}_1), (Z - \widetilde{Z}), (Y_1 - \widetilde{Y}_1)(Z - \widetilde{Z})\}^T$, where $\widetilde{}$ denotes sample average over the complete cases. We also used the GEE estimator discussed by Yang and Tsiatis (2001, sec. 2.5), where the "response vector" is either $(Y_2, Y_1)^T$ or Y_1 depending on whether Y_2 is observed. For all of these estimators, standard errors were obtained using conventional formulæ. Estimation of β was also carried out by several methods suggested by the preceding development: using the simple IWCC estimator with $\pi^{(c)}$ correctly specified, $c = 0, 1$, and the $\gamma^{(c)}$ estimated via ML; based on the efficient influence function and parametric modeling of $E(Y_2|Y_1, X_2, Z)$ and $E(Y_2|Y_1, Z)$, denoted REG; again based on the efficient influence function, with these conditional expectations estimated by locally weighted polynomial smoothing, denoted LOESS; and by the ideal, practically unachievable efficient estimator with these conditional expectations known, denoted BENCHMARK. More detail on the last three approaches is given shortly. Standard errors for these methods were obtained using the variance of the influence function given in equation (19) of Davidian et al. (2004) and by the sandwich technique (see Section 6 of Davidian et al., 2004). The latter performed better for $n = 250$ and the two were virtually identical otherwise, so sandwich estimates are used in results below. Nominal 95% Wald confidence intervals for β were constructed as estimate ± 1.96 times standard error.

2 SCENARIO 1

In the first scenario, follow-up responses were generated according to

$$Y_{2i} = \alpha_0 Z_i + [\alpha_1 + \alpha_2\{X_{2i} - E(X_{2i})\}](Y_{1i} - \mu_1) + [\alpha_3 + \alpha_4\{X_{2i} - E(X_{2i})\}]\{(Y_{1i} - \mu_1)^2 - \sigma_1^2\} + \alpha_5\{X_{2i} - E(X_{2i})\} + \epsilon_i, \tag{1}$$

where $\epsilon_i \sim \mathcal{N}(0, 1)$, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.5, 0.5, 0.3, 0.4, 0.3, 0.4)$, which yields $\beta = 0.336$ (computed from 100,000 simulated data sets). Here, then, $E(Y_2|Y_1, X_2, Z = c)$, $c = 0, 1$, are linear in functions of (Y_1, X_2), and, as X_2 is binary, $E(Y_2|Y_1, Z = c) = E(Y_2|Y_1, X_2 = 0, Z = c)P(X_2 = 0|Y_1, Z = c) + E(Y_2|Y_1, X_2 = 1, Z = c)P(X_2 = 1|Y_1, Z = c)$. For

BENCHMARK, we substituted these expressions for each i, evaluated at the true values of all parameters, for $\widehat{e}_{q(c)i}$ and $\widehat{e}_{h(c)i}$ in $\widehat{\beta}$ in Section 6 of Davidian et al. (2004). For REG, we fitted the models $E(Y_2|Y_1, X_2, Z = c) = \alpha_0^{(c)} + \alpha_1^{(c)} Y_1 + \alpha_2^{(c)} X_2 Y_1 + \alpha_3^{(c)} Y_1^2 + \alpha_4^{(c)} X_2 Y_1^2 + \alpha_5^{(c)} X_2$, $c = 0, 1$, by OLS on complete cases, corresponding to the correct model (1), to obtain the $\widehat{e}_{q(c)i}$ and then obtained the $\widehat{e}_{h(c)i}$ two ways: fitting the true logistic regression model $P(X_2 = 1|Y_1, Z)$ by ML for $c = 0, 1$ and then integrating the fit of $E(Y_{2i}|Y_{1i}, X_{2i}, Z_i = c)$ for each i, as above, denoted REG-QUAD-TRUE; and fitting the models $E(Y_2|Y_1, Z = c) = \alpha_0^{(c)} + \alpha_1^{(c)} Y_1 + \alpha_2^{(c)} Y_1^2$, $c = 0, 1$, by OLS based on the complete cases, denoted REG-QUAD-DIRECT. For each, the $\widehat{e}_{q(c)i}$ and $\widehat{e}_{h(c)i}$ were substituted in $\widehat{\beta}$. For LOESS, we fit $E(Y_2|Y_1, X_2 = x, Z = c)$ for $x = 0, 1$, $c = 0, 1$ using locally weighted polynomial smoothing (Cleveland et al. 1993) with polynomials of degree 2 and span 0.7, obtaining the $\widehat{e}_{q(c)i}$ for each i as predicted values, and derived $\widehat{e}_{h(c)i}$ in two ways, LOESS-TRUE and LOESS-DIRECT analogous to REG-QUAD above. For all of BENCHMARK, the REG-QUAD methods, and the LOESS methods, we took $\pi^{(c)}$, $c = 0, 1$, to be correctly specified and estimated $\gamma^{(c)}$ by ML. For REG-QUAD, we also estimated β by incorrectly modeling $\pi^{(c)}$, $c = 0, 1$, as constants, estimated by the sample proportions n_{Rc}/n_c, to assess the potential for "double robustness" when these probabilities are mismodeled, denoted by MIS. Finally, to evaluate "double robustness" in "the other direction," we took $\pi^{(c)}$, $c = 0, 1$, to be correct but fitted the incorrect model $E(Y_2|Y_1, X_2, Z = c) = \alpha_0^{(c)} + \alpha_1^{(c)} Y_1 + \alpha_2^{(c)} X_2 Y_1 + \alpha_3^{(c)} X_2$, $c = 0, 1$, by OLS on complete cases; REG-LIN-TRUE-MIS corresponds to then integrating these fits to obtain $E(Y_1|Y_1, Z = c)$ as above and REG-LIN-DIRECT-MIS to modeling $E(Y_1|Y_1, Z = c)$ instead by linear functions of Y_1.

Results are shown in Table 1 for $n = 250$ and 1000. Estimators using the TRUE and DIRECT approaches showed almost identical performance in every case, so we report the latter for brevity. As expected, all estimators based on the theory are approximately unbiased, including those denoted MIS, which involve mismodeling of the $\pi^{(c)}$ or the conditional expectations, demonstrating "double robustness." In contrast, the "popular" estimators exhibit nonnegligible bias and sampling variances in most cases at least as large as those for the estimators in Section 6 of Davidian et al. (2004), with the exception of LOESS. This estimator, not unexpectedly, shows substantial variation for $n = 250$, reflecting degraded performance of smoothing with the small sample sizes in each group (≈ 100 or 80 for $Z = 0$ or 1, respectively); this vanishes for $n = 1000$. Regression modeling based on the correct model, REG-QUAD, is only mildly inefficient relative to the ideal BENCHMARK for $n = 250$, and all estimators based on using correct regression relationships or smoothing are fully efficient relative to the ideal for $n = 1000$. The IWCC estimator, which does not exploit auxiliary covariates, suffers considerable efficiency loss, demonstrating how the "augmentation" used by the estimators based on the Robins et al. theory can dramatically increase precision. Overall, confidence intervals for the latter estimators achieve coverage close to the nominal level in most cases.

3 SCENARIO 2

In the second scenario, with $\epsilon_i \sim \mathcal{N}((0,1)$, follow-up responses were generated from

$$(2) \qquad Y_{2i} = \alpha_0 Z_i + [\alpha_1 + \alpha_2\{X_{2i} - E(X_{2i})\}]e^{\alpha_3(Y_1-\mu_1)} + \alpha_4\{X_{2i} - E(X_{2i})\} + \epsilon_i$$

and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.5, 0.3, 0.1, 1.0, 2.0)$, leading to $\beta = -0.449$. With the exception of BENCHMARK, for which $E(Y_2|Y_1, X_2, Z = c)$ and $E(Y_2|Y_1, Z = c)$ were determined from (2), the remaining estimators were as before. For REG, we used quadratic functions in Y_1 and X_2 to approximate the exponential relationship in (2), and, instead of linear functions, we also considered cubic functions in Y_1 for $E(Y_2|Y_1, X_2, Z = c)$ and $E(Y_2|Y_1, Z = c)$. Thus, note in this setting, all REG estimators involve mismodeling of the regression functions implied by (2), and, in Table 2, MIS denotes in addition mismodeling of the $\pi^{(c)}$, as before.

The results in Table 2 again show that the estimators in Section 6 of Davidian et al. (2004) are unbiased. LOESS shows substantial relative variation for $n = 250$; here, the REG estimators using quadratic models are 90% efficient, while using more complicated cubic models degrades performance. At $n = 1000$, however, both the LOESS and REG methods attain precision approaching that of the ideal BENCHMARK, showing that regression modeling that is “close” to representing the true relationships may be sufficient to achieve almost fully efficient performance. Moreover, despite the fact that these models are “wrong,” the “double robustness” property appears to hold. Confidence intervals for all of these estimators achieve nominal coverage for both sample sizes in almost every case. Again, the IWCC estimator is inefficient relative to the others, and the “popular” estimators show overall poor performance.

4 CONCLUSIONS

From these results and others not shown, we conclude that the methods in Section 6 of Davidian et al. (2004) based on modeling the conditional expectations parametrically or nonparametrically (for larger n) offer the analyst a reliable set of techniques for taking into account missing follow-up response that moreover achieve considerable gains in precision over simpler such approaches by exploiting relationships between follow-up and baseline response and other covariates.

REFERENCES

Cleveland, W.S., Grosse, E. and Shyu, W.M. (1993). Local regression models. In *Statistical Models in S* (J.M. Chambers and T.J. Hastie, eds.) 309–376. Chapman and Hall, New York.

Davidian, M., Tsiatis, A.A., and Leon, S. (2004). Semiparametric estimation of treatment effect in a pretest-posttest study with missing data. *Statistical Science*, to appear.

Table 1

Simulation results for true quadratic relationship (1), 1000 Monte Carlo data sets, true $\beta = 0.336$. *Estimators are described in the text. MC Mean is Monte Carlo average, RB is relative bias (%), MC SD is Monte Carlo standard deviation, SE is the average of estimated standard errors (see text), MSE Ratio is Mean Square Error (MSE) for BENCHMARK divided by MSE of the indicated estimator, CP is empirical coverage probability of 95% confidence interval*

Estimator	**MC Mean**	**RB**	**MC SD**	**SE**	**MSE Ratio**	**CP**
			$n = 250$			
BENCHMARK	0.341	1.4	0.163	0.159	1.00	0.94
LOESS-DIRECT	0.340	1.4	0.216	0.167	0.57	0.91
REG-QUAD-DIRECT	0.341	1.4	0.168	0.155	0.95	0.94
REG-QUAD-DIRECT-MIS	0.341	1.4	0.168	0.153	0.94	0.94
REG-LIN-DIRECT-MIS	0.335	−0.3	0.188	0.178	0.75	0.94
IWCC	0.334	−0.6	0.200	0.188	0.66	0.95
GEE	0.265	−21.1	0.186	0.183	0.66	0.93
ANCOVA II	0.279	−17.0	0.181	0.177	0.73	0.93
ANCOVA I	0.281	−16.4	0.181	0.178	0.74	0.93
Paired t-test	0.303	−9.9	0.196	0.192	0.67	0.94
Two sample t-test	0.256	−23.9	0.199	0.193	0.57	0.93
			$n = 1000$			
BENCHMARK	0.333	−0.8	0.079	0.078	1.00	0.94
LOESS-DIRECT	0.333	−0.9	0.080	0.078	0.99	0.93
REG-QUAD-DIRECT	0.333	−0.8	0.079	0.078	1.00	0.94
REG-QUAD-DIRECT-MIS	0.333	−0.9	0.079	0.077	1.00	0.93
REG-LIN-DIRECT	0.335	−0.3	0.088	0.089	0.81	0.95
IWCC	0.334	−0.6	0.094	0.095	0.72	0.95
GEE	0.265	−21.1	0.089	0.084	0.50	0.90
ANCOVA II	0.276	−17.9	0.087	0.089	0.57	0.91
ANCOVA I	0.277	−17.5	0.087	0.089	0.57	0.91
Paired t-test	0.299	−11.0	0.094	0.096	0.62	0.94
Two sample t-test	0.253	−24.6	0.094	0.097	0.40	0.88

Table 2

Simulation results for true exponential relationship (2), 1000 Monte Carlo data sets, true $\beta = -0.449$. Entries are as in Table 1.

Estimator	MC Mean	RB	MC SD	SE	MSE Ratio	CP
			$n = 250$			
BENCHMARK	−0.446	0.7	0.198	0.197	1.00	0.96
LOESS-DIRECT	−0.446	0.7	0.258	0.204	0.59	0.93
REG-QUAD-DIRECT	−0.448	0.4	0.210	0.196	0.90	0.94
REG-QUAD-DIRECT-MIS	−0.448	0.3	0.210	0.194	0.89	0.94
REG-CUB-DIRECT	−0.444	1.2	0.240	0.209	0.68	0.94
REG-CUB-DIRECT-MIS	−0.444	1.2	0.240	0.209	0.68	0.93
IWCC	−0.449	0.0	0.225	0.214	0.78	0.94
GEE	−0.798	−77.7	0.213	0.204	0.23	0.58
ANCOVA II	−0.785	−74.8	0.205	0.206	0.25	0.61
ANCOVA I	−0.784	−74.5	0.205	0.206	0.26	0.62
Paired t-test	−0.761	−69.3	0.217	0.219	0.27	0.68
Two sample t-test	−0.806	−79.5	0.224	0.218	0.22	0.60
			$n = 1000$			
BENCHMARK	−0.448	0.1	0.100	0.098	1.00	0.95
LOESS-DIRECT	−0.447	0.2	0.101	0.097	0.98	0.95
REG-QUAD-DIRECT	−0.448	0.0	0.102	0.099	0.96	0.95
REG-QUAD-DIRECT-MIS	−0.448	0.0	0.102	0.098	0.96	0.95
REG-CUB-DIRECT	−0.448	0.0	0.103	0.100	0.95	0.95
REG-CUB-DIRECT-MIS	−0.448	0.0	0.103	0.099	0.95	0.94
IWCC	−0.447	0.3	0.109	0.108	0.85	0.95
GEE	−0.798	−77.9	0.108	0.107	0.08	0.10
ANCOVA II	−0.786	−75.1	0.106	0.103	0.08	0.10
ANCOVA I	−0.786	−75.1	0.106	0.103	0.08	0.10
Paired t-test	−0.765	−70.4	0.112	0.110	0.09	0.18
Two sample t-test	−0.808	−80.0	0.112	0.110	0.07	0.10