Published in final edited form as:
Comput Stat Data Anal. 2007 August 15; 51(12): 6114–6122.

Quantile Stratification Based on a Misspecified Propensity Score in Longitudinal Treatment Effectiveness Analyses of Ordinal Doses

Andrew C. Leon, Ph.D.(1) and **Donald Hedeker, Ph.D.**(2)

(1) *Weill Medical College of Cornell University*

(2) *University of Illinois at Chicago*

Summary

The propensity adjustment provides a strategy to reduce the bias in treatment effectiveness analyses that compare non-equivalent groups such as seen in observational studies (Rosenbaum and Rubin, 1983). The objective of this simulation study is to examine the effect of omitting confounding variables from the propensity score on the quintile-stratified propensity adjustment in a longitudinal study. The primary focus was the impact of a misspecified propensity score on bias. Three features of the omitted confounding variables were examined: type of predictor variable (binary vs. continuous), constancy over time (time-varying vs. time-invariant), and magnitude of the association with treatment and outcome (null, small, and large odds ratios). The simulation results indicate that omission of continuous, time-varying confounders that are strongly associated with treatment and outcome (i.e., an odds ratio of 1.75) adversely impacts bias, coverage, and type I error. Omitted time-varying continuous variables had somewhat more effect on bias than omitted binary variables. Time-invariant confounding variables that are not included in the propensity score have a much less effect on results. This evaluation only examined continuous treatment effectiveness outcomes and the propensity scores used for stratification included just four variables. Relative to the use of the propensity adjustment in applied settings that typically comprise numerous potential confounding variables, the impact of one omitted continuous, time-varying confound in this simulation study could be overstated.

Keywords

effectiveness; misspecification; ordinal doses; propensity adjustment; stratification

1. Introduction

An observational design typically includes subjects who are more representative of patients with a particular illness. For instance, observational studies tend to have less restrictive inclusion and exclusion criteria than those used in randomized controlled clinical trials (RCT). As a result the observational design can provide greater generalizability than seen in RCTs. However, because an investigator observes, but does not manipulate treatment, the data analyst

Corresponding Author: Andrew C. Leon, Weill Medical College of Cornell University, Department of Psychiatry, Box 140, 525 East 68th Street, New York, NY 10021, telephone: (212) 746-3872, fax: (212) 746-8754, acleon@med.cornell.edu.

in an observational study is faced with non-equivalent groups. For instance, those who receive more intensive treatment are likely to have had more severe symptoms, a history of failed treatments, or more comprehensive health insurance coverage.

The propensity adjustment is a strategy to compare non-equivalent groups such as in analyses of treatment effectiveness in observational studies through stratification, matching, or covariate adjustment (Rosenbaum and Rubin, 1983). The propensity score, *e(x)*, which represents the conditional probability of group assignment given demographic and clinical characteristics, forms the basis for the propensity adjustment. The propensity adjustment has been shown to reduce bias in the treatment effectiveness estimate relative to unadjusted analyses.

Recent work has examined the application of the propensity for treatment intensity to longitudinal data (Leon and Hedeker, 2005). This approach can incorporate repeated assessments of non-randomized, time-varying, ordinal doses of treatment over the course of a chronic illness. The analyses involve two stages: a propensity model and a treatment effectiveness model. Simulation studies have documented that a vast majority of the bias seen in unadjusted analyses is accounted for with an adjustment for the propensity for treatment intensity when the effectiveness evaluation involves either a mixed-effects grouped-time survival analysis (Leon, Hedeker and Teres, in press) or a mixed-effects linear regression analysis (Leon and Hedeker, in press).

The extent of balance between non-randomized treatment groups on demographic and clinical characteristics that is achieved with the propensity adjustment can be readily evaluated. Yet, it is not feasible to examine balance across treatment groups on confounders that are unknown or unmeasured. Nevertheless, an implicit assumption of the propensity strategy is that the propensity score is adequately specified. Rosenbaum (2002) has described sensitivity analyses that examine the range of bias in propensity-adjusted parameter estimates that could result from a misspecified propensity score. Drake (1993) examined the effect of misspecification of the quintile-stratified propensity adjustment in a simulation study of cross-sectional data in which the outcome variable is either continuous or binary and showed that considerable bias can remain if a key confounder is omitted from the propensity score. Here, that work is extended in several ways.

The objective of the current study is to examine the effect of omitting confounders from the propensity score on the quintile-stratified propensity adjustment in a longitudinal study. Initially the two stages of the adjustment procedure are described: the longitudinal propensity and treatment effectiveness models. A simulation study then evaluates the impact of propensity score misspecification.

2. Longitudinal Implementation of the Propensity Adjustment

2.1. Propensity Model for the Longitudinal Study of Ordinal Doses

Elsewhere we have described a modification of the Rosenbaum and Rubin (1983) propensity model to examine the longitudinal study of K time-varying ordinal doses denoted by the variable T_{ij} (Leon and Hedeker, 2005). The Rosenbaum and Rubin (1983) notation is adapted here. The *ordinal propensity score* is specified for subject i (i =1,…N), at time j (j =1,…J_i), and dose k (k = 0,…K − 2). Here, to be consistent with the notation of a dichotomous logistic regression model, the index starts at *0* and goes to K − 2:

$$e_k(x_{ij},v_i)=P(T_{ij}>k|x_{ij},v_i),$$

Where x_{ij} is a vector of covariates hypothesized to be related to intensity of treatment (i.e., ordinal dose) as well as an intercept term. The subject-specific random effect, v_i, is normally distributed in the population with mean *0* and variance σ_v^2.

The propensity score can be estimated using an ordinal mixed-effects logistic regression model (Hedeker and Gibbons, 1994):

$$\ln\left[\frac{P(T_{ij}>k)}{1-P(T_{ij}>k)}\right]=\gamma_k+x'_{ij}\beta+v_i \tag{1}$$

where γ_k represents the threshold for dose k, x_{ij} is the $p \times 1$ vector of covariates and an intercept, and β represents the corresponding regression coefficients. Given subscript j, vector x can include both time-varying and time-invariant covariates, each of which must be assessed prior to the start of a particular course of treatment. Accordingly, this model allows for changes in an individual's propensity score and dose over time. If a temporal trend in repeated dosing is hypothesized, the model can also include time effects. For purposes of identification the first threshold γ_1 is typically set to zero (or the intercept is set to zero). Threshold values γ_k are increasing and reflect the marginal cumulative logits. Specifically, there are $K-1$ cumulative logits (indexed as $k = 0,\ldots K-2$) for the K ordinal doses and, assuming proportional odds, each covariate is assumed to have the same effects across these logits (McCullagh, 1980).

Neither the intercept, which is a constant, nor the threshold is needed for the propensity-based ranking that is used for stratification. It is not necessary to include the thresholds, γ_k, for dose k ($k = 0,\ldots K-2$) in this expression under the proportional odds assumption, because the thresholds do not vary by subject, time, or comparisons among ordinal doses. A logistic response function can be used to express the mixed-effects propensity score for subject i at time j:

$$e(x_{ij},v_i)=\frac{\exp(x'_{ij}\beta+v_i)}{1-\exp(x'_{ij}\beta+v_i)} \tag{2}$$

The propensity score, $e(x_{ij},v_i)$, which ranges from *0* to *1*, represents the probability of receiving a higher dose (T) of treatment based on the contribution of covariates, x, and subject effects, v_i. A high propensity score indicates that the observation has characteristics associated with more intensive doses; whereas a low propensity score indicates that the observation has the characteristics of someone not likely to receive a higher dose at a particular point in time. Based on the propensity score for subject i at time j, each observation is classified into a propensity quintile, $q_{(1)}, \ldots, q_{(5)}$. Quintile-stratified treatment effectiveness analyses are then conducted.

2.2. Longitudinal Treatment Effectiveness Analyses

The k treatment groups are compared on a longitudinal continuous dependent variable, y_{ij}, using a mixed-effects linear regression model. The model is specified as:

$$y_{ij}=\alpha_0+\alpha_1 T_{ij1}\ldots+\alpha_{K-1}T_{ij,K-1}+\theta_i+\varepsilon_{ij} \tag{3}$$

where α_0 is the intercept term, α_1 is the coefficient for dummy-coded treatment dose T_{ij1}, α_{k-1} is the coefficient for dummy coded treatment dose $T_{ij,K-1}$, θ_i is a subject-specific random intercept, distributed as $N(0,\sigma_\theta^2)$, and ε_{ij} is the error term for subject i at time j, distributed independently as $N(0,\sigma_\varepsilon^2)$. The null hypotheses tested below represent the dose-specific effectiveness evaluations, H_{0k}: $\alpha_k=0$ for $k = 0,\ldots K-1$. If a temporal influence on treatment

effectiveness is hypothesized, the model could also include a term representing the slope over time and possibly an interaction of treatment by time.

As stated earlier, the effectiveness analyses are conducted separately for each quintile. The quintile-specific results can be pooled using the Mantel-Haenszel procedure (as described by Fleiss, 1981) in which each quintile-specific parameter estimate is weighted by the inverse of its squared standard error. In order to pool the results, however, it is assumed that there is not a treatment by propensity interaction. A procedure to test that assumption, based on the likelihood ratio test, is described elsewhere (Leon and Hedeker, 2005).

3. Simulation Study

The impact of propensity score misspecification on the performance of the quintile-stratified, longitudinal propensity adjustment in a mixed-effects linear regression model was examined in a Monte Carlo simulation study. The primary focus was to examine the impact on bias of three features of the omitted confounding variables including:

1. type of predictor variable (binary vs. continuous)
2. constancy over time (time-varying vs. time-invariant)
3. magnitude of the association with treatment and outcome (specified as null, small, and large odds ratios).

In addition, we examined type I error, statistical power and coverage probability.

3.1. Simulation Specifications

This simulation study distinguishes between *true*, $e_T(x_{ij},v_i)$, and *estimated* $e_E(x_{ij},v_i)$, propensity scores. Simulated data were generated based on $e_T(x_{ij},v_i)$, yet those data were subsequently analyzed stratified into quintiles based on $e_E(x_{ij},v_i)$. The data were generated in the following manner. Initially, true propensity scores were calculated for each observation were based on vector x_T,

$$x_T' = [\, x_0 \;\; x_1 \;\; x_2 \;\; x_3 \;\; x_4 \;\; x_5 \;\; x_6 \;\; x_7 \;\; x_8 \,]$$

which includes the intercept, x_0, and eight randomly generated predictor variables:

1. two time-invariant, continuous variables (x_1, x_2)
2. two time-invariant, binary variables (x_3, x_4)
3. two time-varying, continuous variables (x_5, x_6)
4. two time-varying, binary variables (x_7, x_8)

Each of these variables was generated based on an underlying standard normal distribution. In addition, the correlation between all pairs of the propensity model predictor variables, ρ_{x_T}, was set at either 0.20 or 0.40. For each binary variable the underlying standard normal was dichotomized to yield an even split of zeros and ones. The odds ratios for the eight predictors of dose in the propensity model varied (1.0, 1.25, 1.75), representing null, small, and large associations with treatment intensity, respectively. The true propensity score, $e_T(x_{ij},v_i)$, based on those odds ratios, yielded the observed time-varying ordinal doses based on the threshold concept (Agresti, 2002). A sample size of 250 subjects was examined and each subject had eight repeated observations during the hypothetical longitudinal study. The intraclass correlation coefficient (ICC) ρ_x for each time-varying predictor (x_5 - x_8) over time was specified as 0.40.

3.2. Treatment Effectiveness Simulation

The effect of each of three doses, relative to dose 0 (i.e., the control), on the continuous outcome was specified in mixed-effects linear regression model (2), with between dose standardized effects of 0, .22, and .56, representing null, small and moderate treatment effects (in standard deviation units). The association of each covariate ($x_1 - x_8$) on both the continuous effectiveness outcome and dose was specified to be approximately equivalent. The ICC, ρ_k, among the repeated doses was specified to be equal to the ICC, ρ_Y, among the repeated assessments of outcome such that $\rho_k = \rho_Y = .40$. One thousand data sets were generated for each combination of simulation specifications.

3.3. Propensity Score Misspecification

In an effort to evaluate the effect of misspecification, the *estimated propensity score*, $e_E(x_{ij}, v_i)$, was used for quintile stratification in the treatment effectiveness analyses and was calculated based on vector x_E:

$$x'_E = [x_0 \;\; x_1 \;\; x_3 \;\; x_5 \;\; x_7]$$

In this way, four confounds (x_2, x_4, x_6, x_8,) were ignored in the stratification process, even though they were components of the *true propensity score*. (Note that in the simulation when odds ratios of 1.0 are specified for the omitted confounds, the true and estimated propensity scores are equivalent.)

3.4. Evaluation of Model Performance

Performance of the models was evaluated using the following criteria: type I error, statistical power, coverage, standardized bias, and root mean square error (RMSE). Type I error and statistical power represent the respective proportions of true and false null hypotheses that were rejected. Coverage is defined as the proportion of simulated data sets for which the 95% confidence interval for the respective parameter estimate included the specified value. Standardized bias, $100 * \frac{E(\widehat{\alpha}) - \alpha}{SE(\widehat{\alpha})}$, was the primary criterion in this study because it expresses bias in units of uncertainty of the parameter estimates across simulated data sets. Demirtas (2004) suggests that if the absolute value of standardized bias exceeds .40, efficiency, coverage and error rates are adversely affected. RMSE combines accuracy and precision and is defined as $\sqrt{E_\alpha\left[(\widehat{\alpha} - \alpha)^2\right]}$. All evaluation criteria were based on the Mantel-Haenszel pooled results. MIXOR software (Hedeker and Gibbons, 1996a) and MIXREG software (Hedeker and Gibbons, 1996b) were used for analyses of the propensity and effectiveness models, respectively. The simulations were designed to emulate the two-staged approach described above such that the results of a propensity model are used to estimate a propensity score which, in turn, is incorporated in the treatment effectiveness analyses. An example of SAS Code to implement this simulation is available from the first author.

3.5. Simulation Results

3.5.1. Standardized Bias—Initially, consider as benchmarks, the performance of the procedure with correctly specified propensity models (two *italicized* models each: odds ratios of 1.25 and 1.75 for x_1, x_3, x_5, x_7) in Table 1 $(\rho_{x_T} = 0.20)$ and Table 3 $(\rho_{x_T} = 0.40)$, in which the omitted confounding variables have a null effect (odds ratios of 1.0 for x_2, x_4, x_6, x_8). The standardized bias in these models is less than 25% of a standard error (median absolute standardized bias: 11.3%), indicating that propensity score-based stratification accounted for substantial bias. The standardized bias, however, is elevated when time-varying confounding variables with associations with treatment of large magnitude (odds ratio of 1.75 for x_6 or x_8)

are omitted from the propensity score, and thus the stratification process. (Standardized bias in excess of 40% is bolded in Tables 1 and 3.) This effect is most prominent for x_6 (the continuous, time-varying confound) and is seen for parameter estimates for each of the three doses, relative to the control, with standardized bias ranging from about 40% to over 180%. Despite the standardization, there is a tendency for somewhat greater standardized bias with larger doses. The largest of these standardized biases arise from the omission of two time-varying confounding variables of large magnitude. In contrast, standardized bias is only slightly elevated with small, time-invariant omitted confounders (odds ratios of 1.25 for x_2 or x_4). The bias increases somewhat with the omission of either larger time-invariant confounders (odds ratios of 1.75 for x_2 or x_4) or small time-varying confounding variables, yet the standardized bias is typically smaller than 35% standard error units, and most often much smaller, unless large time-varying confounding variables are also omitted. Standardized bias is somewhat muted when the correlation among propensity predictors is 0.40 (Table 3) relative to correlations of 0.20 (Table 1), yet the pattern of values in excess of 40% is similar.

3.5.2. Root Mean Squared Error—RMSE is less sensitive than standardized bias to the misspecified propensity score (Tables 1 and 3). Nevertheless, RMSE is somewhat elevated when time-varying confounding variables with odds ratios of 1.75 are omitted, reflecting the pattern described for standardized bias.

3.5.3. Type I Error—The effect of an omitted confounder on type I error mirrors that of bias (Tables 2 and 4). That is, as the odds ratio for the omitted time-varying continuous confound (x_6) increases to 1.75, type I error increases from a nominal level to exceed 10%. An omitted time-invariant confound has little impact on type I error.

3.5.4. Statistical Power—There is increased statistical power (> .90) for the small standardized treatment effects when the continuous, time-varying confounding variable (x_6) of large association (odds ratio: 1.75) is omitted from the propensity score used for stratification. This corresponds to decreased coverage (described below), suggesting increases in both power and precision. Power appears to have reached an asymptote of 1.0 for a moderate treatment effect (.56), given the sample size of 250 with eight repeated observations over time.

3.5.5. Coverage—The simulation-based coverage is adversely affected when the continuous, time-varying confounding variable (x_6), with odds ratio of 1.75, is ignored in the stratification process, resulting in coverage probability of less than 90%. (Coverage that is less than 90% is bolded in the Tables 2 and 4.) The coverage is acceptable for nearly all other simulation specifications.

4.0. Discussion

The sensitivity of a longitudinal, quintile-stratified, propensity adjustment to incomplete specification of the propensity model has been evaluated in this simulation study. The results indicate that time-varying confounders can play a critical role in bias reduction in longitudinal studies, apparently more so than confounders that do not change over follow-up. Although omission of time-varying confounders with a small association with treatment and outcome (i.e., odds ratio of 1.25) had minimal impact on the model performance, those with a larger association clearly had an impact on bias, coverage, and type I error. Omitted continuous time-varying variables had somewhat more effect on bias than did omitted binary time-varying variables. In contrast, omitted variables that were more highly correlated with variables that were included in the propensity score had less impact on bias than omitted variables with lower correlations.

This evaluation examined continuous treatment effectiveness outcomes and included only four variables in the estimated propensity score, which served as the basis for stratification. We acknowledge that this is somewhat oversimplified relative to the number of variables included when the propensity adjustment is applied in practice; and thus, the impact of just one omitted confound could be overstated. For example, studies of cardiovascular disease have included considerably more variables in the propensity score: 18 variables (Grzybowski et al., 2003), 34 variables (Gum et al., 2001) and 102 variables (Normand 2001). Nevertheless, when one large, time-varying confounding variable was omitted from the propensity score in the simulation study, the parameter estimate resulted in substantial bias, reduced coverage probability, and inflated type I error rates. The impact of an omitted time-varying confounding variable would likely be mitigated by including a term for time in the mixed-effects outcome analyses, to the extent that this omitted variable is associated with time.

The simulation results underscore the importance of conducting comprehensive assessments over the course of follow-up in a longitudinal study. Most importantly, the selection of assessments must be guided by clinicians and other researchers with expertise in the substantive area of focus. Whether bias is introduced because a variable is inadvertently excluded from analyses or not collected during assessment is immaterial once the analyses have been completed. The building of a propensity model is not simply a data analytic exercise, but instead must be an active collaboration among researchers. The availability of the variables for the propensity adjustment, through a well-guided choice of assessments at the design stage of an observational study, plays a critical role in bias reduction. Of course, there is a risk of a tradeoff between in depth, time-consuming assessments and retention in a longitudinal study.

Rosenbaum and Rubin (1983) described three approaches to implementing the propensity adjustment: stratification, matching and covariate adjustment. The use of the latter approach is typically discouraged. The simulation study described here was limited to stratification. It is not clear how these results compare to the impact of propensity score misspecification on matching. Furthermore, this simulation-based evaluation of misspecification does not examine sensitivity to hidden bias in the manner proposed by Rosenbaum (2002) in which one estimates a range of change in the magnitude of treatment effectiveness estimates based on the strength of the association of hypothetical omitted confounders with treatment assignment.

Observational studies, by design, will seldom, if ever, provide the complete data needed to calculate the true propensity score; instead an estimated score will be used to implement the propensity adjustment. This simulation study has shown that neglecting to include in the estimated propensity score time-varying confounds that are strongly associated with treatment and outcome had much greater impact on the performance of a quintile-stratified propensity adjustment than omission of time-invariant confounds. In conclusion, careful propensity model building and evaluation of group balance are essential when the adjustment is applied.

References

1. Agresti, A. Categorical Data Analysis. 2nd. Hoboken, NJ: John Wiley and Sons; 2002.
2. Demirtas H. Simulation-driven inferences for mutiply imputed longitudinal datasets. Statistica Neerlandica 2004;58:466–482.
3. Drake C. Effects of misspecification of the propensity score on estimators of the treatment effect. Biometrics 1993;49:1231–1236.
4. Fleiss, JL. Statistical Methods for Rates and Proportions. John Wiley and Sons; 1981.
5. Grzybowski M, Clements EA, Parsons L, Welch R, Tintinalli AT, Ross MA, Zalenski RJ. Mortality benefit of immediate revascularization of acute ST-segment elevation myocardial infarction in patients with contraindications to thrombolytic therapy: a propensity analysis. JAMA 2003;290:1891–8. [PubMed: 14532318]

6. Gum PA, Thamilarasan M, Watanabe J, Blackstone EH, Lauer MS. Aspirin use and all-cause mortality among patients being evaluated for known or suspected coronary artery disease: A propensity analysis. JAMA 2001;286:1187–94. [PubMed: 11559263]
7. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. Biometrics 1994;50:933–944. [PubMed: 7787006]
8. Hedeker D, Gibbons RD. MIXOR: a computer program for mixed-effects ordinal regression analysis. Computer Methods and Programs in Biomedicine 1996;49:157–176. [PubMed: 8735023]
9. Hedeker D, Gibbons RD. MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors. Computer Methods and Programs in Biomedicine 1996;49:229–252. [PubMed: 8800609]
10. Leon AC, Hedeker D. A Mixed-Effects Quintile-Stratified Propensity Adjustment for Effectiveness Analyses of Ordered Categorical Doses. Statistics in Medicine 2005;24:647–658. [PubMed: 15678413]
11. Leon AC, Hedeker D. A Comparison of Mixed-Effects Quantile Stratification Propensity Adjustment Strategies for Longitudinal Treatment Effectiveness Analyses of Continuous Outcomes. Statistics in Medicine. in press
12. Leon AC, Hedeker D, Teres JJ. Bias Reduction in Effectiveness Analyses of Longitudinal Ordinal Doses with a Mixed-Effects Propensity Adjustment. Statistics in Medicine. in press
13. McCullagh P. Regression models for ordinal data (with discussion). J Roy Statist Soc 1980;B 42:109–142.
14. Normand SLT, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol 2001;54:387–98. [PubMed: 11297888]
15. Rosenbaum, PR. Observational Studies. 2nd. New York: Springer; 2002.
16. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.

Table 1

Standardized Bias and Root Mean Squared Error in a Simulation Study of Quintile-Stratification based on a Misspecified Propensity Score: Correlation among propensity covariates, ρ_{x_T} =0.20

Time Invariant Confounds				Time-Varying Confounds				Standardized bias			RMSE		
X_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$
*													
1.25	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	11.66	8.91	9.42	0.08	0.08	0.09
	1.25		1.00		1.00		1.00	11.49	7.72	6.66	0.08	0.09	0.09
	1.00		1.25		1.00		1.00	9.08	8.56	8.05	0.08	0.09	0.09
	1.00		1.00		1.25		1.00	23.34	24.22	32.91	0.09	0.09	0.09
	1.00		1.00		1.00		1.25	13.29	16.32	16.65	0.08	0.09	0.09
	1.25		1.25		1.00		1.00	7.70	8.50	5.00	0.08	0.09	0.09
	1.00		1.00		1.25		1.25	27.93	31.55	42.36	0.09	0.09	0.09
	1.25		1.00		1.25		1.00	22.11	22.77	28.80	0.08	0.09	0.09
	1.00		1.25		1.00		1.25	11.81	15.91	15.98	0.09	0.09	0.09
	1.25		1.00		1.00		1.25	14.56	15.26	14.73	0.08	0.09	0.09
	1.00		1.25		1.25		1.00	22.21	25.57	34.02	0.09	0.09	0.09
1.75	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	21.96	15.84	10.90	0.09	0.09	0.09
	1.75		1.00		1.00		1.00	33.02	20.17	24.61	0.10	0.10	0.10
	1.00		1.75		1.00		1.00	29.93	21.26	15.97	0.10	0.10	0.10
	1.00		1.00		**1.75**		1.00	**82.12**	**93.88**	**141.05**	0.12	0.13	0.16
	1.00		1.00		1.00		**1.75**	**48.51**	**44.05**	**59.28**	0.10	0.11	0.11
	1.75		1.75		1.00		1.00	34.82	29.60	26.82	0.10	0.10	0.10
	1.00		1.00		**1.75**		**1.75**	**102.67**	**122.21**	**187.63**	0.13	0.16	0.21
	1.75		1.00		**1.75**		1.00	**84.18**	**102.82**	**159.09**	0.12	0.14	0.18
	1.00		1.75		1.00		**1.75**	**47.08**	**49.46**	**63.33**	0.11	0.12	0.12
	1.75		1.00		1.00		**1.75**	**50.12**	**55.52**	**68.42**	0.11	0.12	0.12
	1.00		1.75		**1.75**		1.00	**81.86**	**93.59**	**136.93**	0.13	0.14	0.17

Notes:

1) *x1*, *x2*, *x5*, and *x6* are continuous variables

2) *x3*, *x4*, *x7*, and *x8* are binary variables

3) *x2*, *x4*, *x6*, and *x8* have been ignored in stratification

4) Treatment effects (α_k) for dose *k*, are expressed in standard deviation units (i.e., effect size).

Table 2

Type I Error, Statistical Power, and Coverage Probability in a Simulation Study of Quintile-Stratification based on a Misspecified Propensity Score: Correlation among propensity covariates, ρ_{x_T} =0.20

Time Invariant Confounds				Time-Varying Confounds				Type I Error	Statistical Power		Coverage		
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$
1.25	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	0.06	0.88	1.00	0.94	0.95	0.95
	1.25		1.00		1.00		1.00	0.06	0.86	1.00	0.94	0.95	0.95
	1.00		1.25		1.00		1.00	0.06	0.87	1.00	0.94	0.95	0.95
	1.00		1.00		1.25		1.00	0.07	0.90	1.00	0.93	0.95	0.93
	1.00		1.00		1.00		1.25	0.05	0.88	1.00	0.95	0.94	0.95
	1.25		1.25		1.00		1.00	0.05	0.86	1.00	0.95	0.95	0.95
	1.00		1.00		1.25		1.25	0.07	0.91	1.00	0.93	0.94	0.92
	1.25		1.00		1.25		1.00	0.06	0.90	1.00	0.94	0.95	0.95
	1.00		1.25		1.00		1.25	0.05	0.87	1.00	0.95	0.94	0.94
	1.25		1.00		1.00		1.25	0.06	0.88	1.00	0.94	0.94	0.94
	1.00		1.25		1.25		1.00	0.05	0.91	1.00	0.95	0.95	0.93
1.75	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	0.07	0.84	1.00	0.93	0.95	0.96
	1.75		1.00		1.00		1.00	0.07	0.82	1.00	0.93	0.94	0.93
	1.00		1.75		1.00		1.00	0.06	0.83	1.00	0.94	0.94	0.94
	1.00		1.00		**1.75**		1.00	0.13	0.94	1.00	**0.87**	**0.86**	**0.72**
	1.00		1.00		1.00		**1.75**	0.09	0.87	1.00	0.91	0.92	0.91
	1.75		1.75		1.00		1.00	0.07	0.83	1.00	0.93	0.94	0.94
	1.00		1.00		**1.75**		**1.75**	0.18	0.96	1.00	**0.82**	**0.77**	**0.53**
	1.75		1.00		**1.75**		1.00	0.14	0.95	1.00	**0.86**	**0.85**	**0.67**
	1.00		1.75		1.00		**1.75**	0.08	0.85	1.00	0.92	0.91	**0.89**
	1.75		1.00		1.00		**1.75**	0.09	0.88	1.00	0.91	0.91	**0.89**
	1.00		1.75		**1.75**		1.00	0.16	0.92	1.00	**0.84**	**0.83**	**0.69**

Notes:

1) *x1*, *x2*, *x5*, and *x6* are continuous variables

2) *x3*, *x4*, *x7*, and *x8* are binary variables

3) *x2*, *x4*, *x6*, and *x8* have been ignored in stratification

4) Treatment effects (α_k) for dose *k*, are expressed in standard deviation units (i.e., effect size).

Table 3

Standardized Bias and Root Mean Squared Error in a Simulation Study of Quintile-Stratification based on a Misspecified Propensity Score: Correlation among propensity covariates, ρ_{x_T} =0.40

Time Invariant Confounds				Time-Varying Confounds				Standardized bias			RMSE		
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$
*													
1.25	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	2.57	3.41	1.69	0.08	0.08	0.09
	1.25		1.00		1.00		1.00	2.85	−2.59	4.08	0.08	0.09	0.09
	1.00		1.25		1.00		1.00	5.61	3.35	1.57	0.08	0.09	0.09
	1.00		1.00		1.25		1.00	10.82	16.29	19.46	0.08	0.09	0.09
	1.00		1.00		1.00		1.25	5.80	7.56	9.30	0.08	0.09	0.09
	1.25		1.25		1.00		1.00	8.32	6.14	2.24	0.08	0.09	0.09
	1.00		1.00		1.25		1.25	18.73	19.90	28.73	0.08	0.09	0.09
	1.25		1.00		1.25		1.00	15.69	15.98	19.42	0.08	0.09	0.09
	1.00		1.25		1.00		1.25	8.52	8.62	9.83	0.08	0.09	0.09
	1.25		1.00		1.00		1.25	10.80	7.27	8.86	0.08	0.09	0.09
	1.00		1.25		1.25		1.00	12.18	16.70	18.29	0.08	0.09	0.09
1.75	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	23.03	14.38	13.89	0.09	0.10	0.09
	1.75		1.00		1.00		1.00	23.64	15.91	17.59	0.09	0.10	0.10
	1.00		1.75		1.00		1.00	29.80	22.20	18.11	0.10	0.10	0.10
	1.00		1.00		**1.75**		1.00	**63.50**	**74.70**	**108.23**	0.11	0.13	0.15
	1.00		1.00		1.00		**1.75**	**40.27**	**42.24**	**51.33**	0.10	0.11	0.11
	1.75		1.75		1.00		1.00	33.64	27.68	27.20	0.10	0.11	0.10
	1.00		1.00		**1.75**		**1.75**	**83.90**	**106.00**	**161.16**	0.13	0.15	0.19
	1.75		1.00		**1.75**		1.00	**73.11**	**84.47**	**118.09**	0.12	0.14	0.16
	1.00		1.75		1.00		**1.75**	**45.35**	**47.15**	**57.46**	0.10	0.11	0.12
	1.75		1.00		1.00		**1.75**	**44.34**	**44.23**	**62.83**	0.10	0.12	0.12
	1.00		1.75		**1.75**		1.00	**66.53**	**79.04**	**113.49**	0.12	0.13	0.16

Notes:

1) *x1*, *x2*, *x5*, and *x6* are continuous variables

2) *x3*, *x4*, *x7*, and *x8* are binary variables

3) *x2*, *x4*, *x6*, and *x8* have been ignored in stratification

4) Treatment effects (α_k) for dose *k*, are expressed in standard deviation units (i.e., effect size).

Table 4

Type I Error, Statistical Power, and Coverage Probability in a Simulation Study of Quintile-Stratification based on a Misspecified Propensity Score: Correlation among propensity covariates, ρ_{x_T} =0.40

Time Invariant Confounds				Time-Varying Confounds				Type I Error	Statistical Power		Coverage		
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$	$\alpha_1 = 0$	$\alpha_2 = .22$	$\alpha_3 = .56$
1.25	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	*1.25*	*1.00*	0.04	0.86	1.00	0.96	0.95	0.95
	1.25		1.00		1.00		1.00	0.07	0.84	1.00	0.93	0.95	0.95
	1.00		1.25		1.00		1.00	0.04	0.86	1.00	0.96	0.94	0.94
	1.00		1.00		1.25		1.00	0.06	0.87	1.00	0.94	0.95	0.93
	1.00		1.00		1.00		1.25	0.06	0.86	1.00	0.94	0.94	0.95
	1.25		1.25		1.00		1.00	0.06	0.86	1.00	0.94	0.95	0.94
	1.00		1.00		1.25		1.25	0.06	0.87	1.00	0.94	0.95	0.93
	1.25		1.00		1.25		1.00	0.06	0.88	1.00	0.94	0.94	0.93
	1.00		1.25		1.00		1.25	0.05	0.85	1.00	0.95	0.94	0.94
	1.25		1.00		1.00		1.25	0.05	0.85	1.00	0.95	0.95	0.94
	1.00		1.25		1.25		1.00	0.05	0.87	1.00	0.95	0.94	0.94
1.75	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	*1.75*	*1.00*	0.06	0.82	1.00	0.94	0.94	0.95
	1.75		1.00		1.00		1.00	0.06	0.79	1.00	0.94	0.93	0.95
	1.00		1.75		1.00		1.00	0.06	0.81	1.00	0.94	0.94	0.95
	1.00		1.00		**1.75**		1.00	0.12	0.92	1.00	**0.88**	**0.87**	**0.78**
	1.00		1.00		1.00		**1.75**	0.06	0.86	1.00	0.94	0.93	0.91
	1.75		1.75		1.00		1.00	0.05	0.81	1.00	0.95	0.93	0.93
	1.00		1.00		**1.75**		**1.75**	0.15	0.94	1.00	**0.85**	**0.81**	**0.65**
	1.75		1.00		**1.75**		1.00	0.11	0.92	1.00	**0.89**	**0.86**	**0.77**
	1.00		1.75		1.00		**1.75**	0.07	0.84	1.00	0.93	0.92	0.91
	1.75		1.00		1.00		**1.75**	0.06	0.84	1.00	0.94	0.91	**0.90**
	1.00		1.75		**1.75**		1.00	0.11	0.91	1.00	**0.89**	**0.86**	**0.78**

Notes:

1) *x1*, *x2*, *x5*, and *x6* are continuous variables

2) *x3*, *x4*, *x7*, and *x8* are binary variables

3) *x2*, *x4*, *x6*, and *x8* have been ignored in stratification

4) Treatment effects (α_k) for dose *k*, are expressed in standard deviation units (i.e., effect size).