HHS Public Access

Author manuscript

Value Health. Author manuscript; available in PMC 2015 March 06.

Published in final edited form as:

Value Health. 2010 ; 13(2): 273–277. doi:10.1111/j.1524-4733.2009.00671.x.

Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals

Stanley Xu, PhD[1], **Colleen Ross, MS**[1], **Marsha A. Raebel, PharmD**[1], **Susan Shetterly, MS**[1], **Christopher Blanchette, PhD**[2], and **David Smith, PhD**[3]

[1]The Institute for Health Research, Kaiser Permanente Colorado, Denver, CO, USA

[2]Division of Clinical and Outcomes Research, Lovelace Respiratory Research Institute, Kannapolis, NC, USA

[3]Kaiser Permanente Center for Health Research, Portland, OR, USA

Abstract

Objectives—Inverse probability of treatment weighting (IPTW) has been used in observational studies to reduce selection bias. To obtain estimates of the main effects, a pseudo data set is created by weighting each subject by IPTW and analyzed with conventional regression models. Currently variance estimation requires additional work depending on type of outcomes. Our goal is to demonstrate a statistical approach to directly obtain appropriate estimates of variance of the main effects in regression models.

Methods—We carried out theoretical and simulation studies to show that the variance of the main effects estimated directly from regressions using IPTW is underestimated, and that the type I error rate is higher due to the inflated sample size in the pseudo data. The robust variance estimator using IPTW often slightly overestimates the variance of the main effects. We propose to use the stabilized weights to directly estimate both the main effect and its variance from conventional regression models.

Results—We applied the approach to a study examining the effectiveness of serum potassium monitoring in reducing hyperkalemia-associated adverse events among 27,355 diabetic patients newly-prescribed a renin-angiotensin-aldosterone system (RAAS) inhibitor. The incidence rate ratio (with monitoring versus without monitoring) and confidence intervals were 0.46 (0.34, 0.61) using the stabilized weights compared to 0.46 (0.38, 0.55) using typical inverse probability of treatment weighting.

Conclusions—Our theoretical, simulation results and real data example demonstrate that the use of the stabilized weights in the pseudo data preserves the sample size of the original data, produces appropriate estimation of the variance of main effect, and maintains an appropriate type I error rate.

Corresponding Author: Stanley Xu, Kaiser Permanente Colorado, Institute for Health Research, 10065 E Harvard Ave., Denver, CO, USA 80111 Tel: 303-614-1210 Fax: 303-614-1205 stan.xu@kp.org.

Keywords

Inverse probability of treatment weighting; stabilized weights; type I error rates; incidence rate ratio; confidence intervals

Introduction

Observational studies have been used by medical researchers seeking to make inference on the effect of treatments on outcomes. Compared to randomized clinical trials, participants' characteristics in an observational study may not be balanced between treated and untreated groups. Consequently, the estimate of a treatment effect may be biased without appropriate adjustment when receipt of treatment is dependent on patients' characteristics (confounders) that also are associated with outcomes. Propensity scores were introduced by Rosenbaum and Rubin [1, 2] and have been used by many researchers to obtain the treatments effects in observational studies [3-8]. A propensity score is the probability of receiving treatment given a set of known covariates and can be used to balance covariates between treated and untreated to obtain an unbiased estimate of treatment effects. Typically, propensity scores in an observational study can be obtained from ordinary logistic regressions if the treatment is binary.

The simplest use of propensity scores is to include them as covariates in outcome modeling. One can first fit a propensity score model that includes many potential covariates, and then the outcome model only has to include the propensity score and a few covariates that have no association with treatment [3, 21]. But this approach can perform poorly if the sample linear discriminant based on covariates is not a monotone function of propensity score [1]. There are three additional strategies that use propensity scores to reduce selection bias: matching, stratification, and inverse probability of treatment weighting (IPTW). Matching subjects in treated groups with those in untreated groups with similar propensity scores can balance the known covariates and reduce selection bias. But it can also result in significant loss of observations of treated subjects, particularly if the untreated pool is small. Stratification places subjects into several mutually-exclusive groups or strata. Based on their propensity scores treatment effects are estimated from each stratum and averaged across strata to estimate the overall treatment effect [3, 9]. The limitation of stratification is that one overall treatment effect may not be interpretable when the treatment effects of strata are very different in scale especially in direction. In addition, subjects in different strata may not separate into distinguishable groups that are meaningful to clinicians. The third propensity score approach is to use IPTW weighted estimators to obtain treatment effects adjusting for known confounders [6, 10, 11]. This approach can incorporate time-dependent covariates and deal with censored data and produce one overall estimate of treatment effect.

For continuous outcome variables, there are three unbiased estimators for treatment effects [9, 11] based on the inverse probability of treatment weighting which have shown consistency but with different variance estimators. However, these variance estimators are large-sample based and they may produce large variance estimates and decrease efficiency of the estimators [9]. Estimators and variance estimates are less developed for discrete

outcome variables. Accurate variance estimation of the treatment effect is critical to testing hypotheses. Underestimation of the variance produces inappropriately narrow confidence intervals and leads to falsely rejecting the null hypothesis. In addition to the large-sample based variance estimators, others have suggested the use of the bootstrap method to obtain the variance of treatment effects [12, 13], which can be used for medium or large samples, and for different effect measures, e.g., difference for continuous outcomes, incidence rate ratio for count data, and odds ratio for dichotomous outcomes. However, the bootstrap method is not suitable for small datasets as there are few values to select from and involves complex programming [14, 15]. A robust variance estimator [12, 19, 20] has also been used to obtain standard error of the treatment effect. This approach adjusts for the lack of independence in replications of records for a subject in the pseudo data and is available in common statistical software packages such as the SAS PROC GENMOD. There are also a variety of weights developed based on sampling designs in survey studies to accurately compute estimates of population statistics and their standard errors from a small sample [22].

The aim of this study is to evaluate the use of stabilized weights to obtain both the treatment effects and their appropriate confidence intervals in the presence of confounders directly from conventional regression in observational studies. In addition, we provide some comparisons of type I error rates using stabilized weights to the robust variance estimator.

Statistical Methods

Let z be an indicator of binary treatment with 1 for treated and 0 for untreated, X be a row vector of confounders for the probability of treatment and outcome, π be the propensity score, and y be the outcome variable. Suppose that there are N subjects in a dataset, with n_1 subjects who received the treatment and n_0 subjects who did not, $N=n_0+n_1$. The probability of treatment without considering covariates is $p=n_1/N$, and the probability of no treatment is 1-p. The propensity score π_i= prob ($z=1|X_i$) is the probability of treatment given the observed covariates X_i. The propensity score can be estimated with a logistic regression model $\pi_i=\frac{exp(X_i\beta)}{1+exp(X_i\beta)}$ where β is a vector of parameters to be estimated from data. With the covariates X in the propensity score model using IPTWs as weights, $W_i=\frac{1}{\pi_i}$ if $z_i=1$ and $W_i=\frac{1}{1-\pi_i}$ if $z_i=0$ where W_i denote the IPTW for subject i.

In the pseudo data using IPTWs, the number of observations is the sum of weights

$$N_w=\sum_{i=1}^{N} w_i.$$

N_w is always greater than N, the sample size of the original data. To examine this further, assume that there is only one covariate, x_1, which is dichotomous and associated with the probability of being treated with a coefficient β_{x1z}. For subjects with $x_1=0$, let m_1 be the number of treated subjects and m_0 be the number of untreated subjects, $M=m_1+m_0$, and e_0 is the probability of being treated when $x_1=0$. For subjects with $x_1=1$, let l_1 be the number of

treated subjects and l_0 be the number of untreated subjects, and $L=l_1+l_0$, and e_1 is the probability of being treated when $x_1=1$. The sample size of the pseudo data with IPTWs is

$$N_w=\frac{m_1}{e_0}+\frac{m_0}{1-e_0}+\frac{l_1}{e_1}+\frac{l_0}{1-e_1} \quad (1)$$

where e_0 and e_1 are estimated from data, $\hat{e}_0=\frac{m_1}{M}$ and $\hat{e}_1=\frac{l_1}{L}$. Substituting $\hat{e}_0$ and $\hat{e}_1$ into equation (1):

$$N_w=M+M+L+L=2N \quad (2)$$

Thus, the sample size doubles in the pseudo data. This is also true when there are other categorical variables that are associated with the probability of being treated. Consequently, regression estimates with IPTWs tend to reject the null hypothesis too frequently because of inflated sample sizes.

An improvement to the inverse probability of treatment weighting is the use of stabilized weights (SW). Stabilized weights have been proposed in modeling time-varying treatment status in reducing selection bias in observational studies [16, 17]. The purpose of using SW in these studies is to reduce the weights of either those treated subjects with low propensity scores or those untreated subjects with high propensity scores. For this paper, we only considered constant treatment status, if $z_i=1$ then $SW_i=\frac{p}{\pi_i}$, and if $z_i=0$ then $SW_i=\frac{1-p}{1-\pi_i}$ where p is the probability of treatment without considering covariates. We will show that the use of SW reduces the type I error by preserving the sample sizes in pseudo datasets. Again assuming that there is only one dichotomous predictor for the probability of being treated, x_1, p can be estimated from data as $\hat{p}=\frac{m_1+l_1}{M+L}$. Using the stabilized weights,

$$N_{SW}=\frac{m_1\hat{p}}{\hat{e}_0}+\frac{m_0(1-\hat{p})}{1-\hat{e}_0}+\frac{l_1\hat{p}}{\hat{e}_1}+\frac{l_0(1-\hat{p})}{1-\hat{e}_1}=m_1+l_1+m_0+l_0=N \quad (3)$$

Equation (3) demonstrates that using SWs in observational studies will result in a pseudo data with sample size that is the same as that of original data. Thus the variance estimate of treatment effect is appropriate directly from conventional regression with SWs. This is also true when other categorical variables exist that are associated with the probability of being treated. The impact of continuous variables on sample size in the pseudo data can not be revealed in closed forms and will be evaluated by simulations in the next section.

Simulation Studies and Results

The simulations were designed to evaluate the use of stabilized weights to estimate the effect of treatment and its variance in the presence of confounders and to obtain appropriate confidence intervals using conventional regressions analyzing data from observational studies. Specifically, we examined the sample sizes in the pseudo datasets and type I error rates when confounders in the propensity score and outcome models were dichotomous, categorical, and continuous.

Simulation Algorithm

Probability model for treatment, z—The treatment indicator variable, z, was simulated according to model (4)

$$\pi = prob\,(z=1|X,\alpha_z,\beta_{xz}) = \frac{exp\,(\alpha_z + X\beta_{xz})}{1+exp\,(\alpha_z + X\beta_{xz})} \quad (4)$$

where α is the intercept and is equal to 0.69, X is a row vector of dichotomous, categorical, or continuous independent variables (confounders). We report the results with independent variables in model (4) being dichotomous, or dichotomous and continuous variables. However, results were similar when categorical variables were included in the model (4).

For simulations with only a dichotomous variable x_1, distributions of the dichotomous variable x_1 were either 50%=0 and 50%=1 or 66.6%=0 and 33.3%=1. For simulations with a dichotomous variable, x_1, and a continuous variables, x_2, when x_1=0, the mean of x_2 was either 1 or −1 and the variance was held constant at 1; when x_1=1, the mean of x_2 ranged from −4 to 4 by increments of 1 and the variance was held constant at 4. We also evaluated different values of the coefficients β_{x1z}, β_{x2z}, β_{x1y}, and β_{x2y} to reflect differing strengths of association with treatment and outcome. For dichotomous x_1, we evaluated positive and negative values of 0.69, 1.39 and 1.79 which correspond to odds ratios of 2, 3, and 4 when positive. For the continuous variable x_2, simulations used values of 0.3, 0.6, and 1.2 for β_{x2z} and β_{x2y}.

We then generated the dichotomous treatment variable z_i based on the treatment probability model (4), i=1 to 500.

Probability model for the outcome, y—The dichotomous outcome variable, y, was simulated based on the following model,

$$prob\,(y=1|z,\alpha_y,\beta_{zy},\beta_{xy}) = \frac{exp\,(\alpha_y + z\beta_{zy} + X\beta_{xy})}{1+exp\,(\alpha_y + z\beta_{zy}, X\beta_{xy})} \quad (5)$$

where α_y is the intercept and equals to 0.69, β_{zy} is the coefficient for the association between treatment and outcome and is assigned zero to assess the type I error rates. X are confounders and β_{xy} are the corresponding coefficients and their values are the same as those of β_{xz} in (4). The dichotomous outcome variable y_i was generated based on the outcome probability model (5), i=1 to 500.

Analysis of each simulated dataset—For each dataset we fit the propensity score models, obtained the inverse probability of treatment weighting and stabilized weights, and then calculated the sample sizes in the pseudo data and fit outcome model. 5000 datasets were simulated and analyzed for each combination of parameters.

Evaluation Measures

Mean sample sizes and standard deviations from 5000 simulated datasets were estimated. Type I error rates were computed as the proportion of p-values less than 0.05 under a null

hypothesis of no treatment effect (β_{zy} =0) based on Wald tests. In addition to IPTW and SW methods, type I error rates using robust variance estimator with IPTWs are also reported.

Simulation Results

Sample sizes and type I error rates when there is only a dichotomous confounder x_1 and β_{zy}=0—We first evaluated the use of stabilized weights when there is only a dichotomous confounder, x_1, and there is no treatment effect, β_{zy} =0. Under a variety of conditions, the IPTW method clearly doubled the sample sizes in the pseudo dataset and inflated the type I error rates (Table 1). Stabilized weights preserved the sample sizes and had type I error rates that were close to 5% (Table 1). The standard deviations of sample sizes in the pseudo datasets were small, indicating that the samples sizes of these 5000 pseudo datasets were all about 500, the original simulated sample size. The level of imbalance of the dichotomous confounding covariate between treated and untreated groups had no impact on the sample sizes of the pseudo data sets and type I error rates with the SW method. Compared to stabilized weights, the robust variance estimator method consistently produced lower than 5% type I error rates because of slightly larger variance estimates. This is consistent with previous studies [19, 20].

Sample sizes and type I error rates when there are a dichotomous confounder x_1 and a continuous confounder x_2 and β_{zy}=0—Sample sizes with stabilized weights remained similar to the original simulated sample size with small standard deviations in most of cases (Table 2). Larger differences emerged when the confounding effect of the continuous variable is strong (β_{x2z} =β_{x2y} =1.2). In those simulations, standard deviations became relatively large, implying greater deviation of some pseudo dataset sample sizes from the original, although the average sample size still remained about 500. In addition, type I error rates became as high as 12%. Also the level of imbalance of the continuous confounding covariate between treated and untreated groups has no impact on the sample sizes of the pseudo data sets and type I error rates with the SW method. Again Table 2 showed that on average using IPTW doubled sample sizes in the pseudo data with the type I error rates reaching as high as 44.0%. For most of cases with continuous confounding covariate the robust variance estimator method produced lower than 5% type I error rates because of slightly larger variance estimates.

An Example

In a recent study examining the effectiveness of serum potassium monitoring in reducing hyperkalemia-associated adverse events during the first year of therapy, 27,355 diabetic patients newly-prescribed a renin-angiotensin-aldosterone system (RAAS) inhibitor between January 1, 2001 and December 31, 2006 were retrospectively identified. Table 3 shows that the patients with and without serum potassium monitoring in the original cohort were significantly different on many demographic and clinical characteristics. Nearly three-fourths of this cohort had serum potassium monitoring during their study follow-up period. This study is an example of when matching by propensity scores would not be optimal as the majority of those with serum potassium monitoring would be omitted due to a smaller number of those without serum potassium monitoring.

We fit a logistic regression model to obtain the propensity scores and included the following variables: use of digoxin, use of diuretic, use of potassium supplements, study site, gender, drug groups of RAAS inhibitor, age, kidney transplant, a drug-dispensing based chronic disease score based on a modification of the method of Clark et al. [18], potassium monitoring within 6 months prior to study entry, diagnosis of hyperkalemia within 6 months prior to study entry, inpatient hospitalization or emergency department visit within 6 months prior to study entry, presence of heart failure, and presence of chronic kidney disease. The SW adjusted results of characteristics comparisons are presented in Table 3 as well. All covariates except age, chronic disease score, and use of potassium supplements became comparable after SW adjustment between the potassium monitored and not monitored (see Table 3). Although age, chronic disease score, and use of potassium supplements remained statistically different between groups, the magnitudes of difference were markedly reduced.

The sample size in the pseudo data using the stabilized weights was 27,407 compared to 54,891 using inverse probability of treatment weighting. The sample size in the pseudo data using the stabilized weights was only slightly larger than the original 27,355 and the impact on variance estimate of treatment effect was minimal. The incidence rate ratio and confidence intervals were 0.46 (0.34, 0.61) using stabilized weights compared to 0.46 (0.38, 0.55) using typical inverse probability of treatment weighting. While adjusting for age, use of potassium supplements, and chronic disease score using SWs, the incidence rate ratio was 0.49 (0.37, 0.66) which was very close to the results without adjustment of these covariates in outcome model, indicating that the balance of age, use of potassium supplements and chronic disease score between the two groups with SWs was sufficient. Comparison of these two weights from this example showed that IPTWs have larger standard deviations and wider ranges than SWs (Table 4).

Discussion

In this paper we demonstrate several advantages of SWs over IPTWs in analyzing data obtained from observational studies. First, using SWs can reduce the weights of either those treated subjects with low propensity scores or those untreated subjects with high propensity scores in the pseudo data sets. Our serum potassium monitoring example showed that IPTWs have larger standard deviation and wider range than SWs (Table 4). Thus results using SWs are robust even with few observations with extreme IPTWs. Second, unlike variance estimators, no additional steps are needed when SWs are used because the SW approach provides appropriate variance estimates and confidence intervals of treatment effect from conventional regression models for fitting the outcome variables. Third, computer programming is simple for one to use SWs to obtain the effect of treatment effects and confidence intervals as compared to the bootstrap approach. One only needs to calculate the weights differently. Fourth, in our simulation studies and example, outcome variables are dichotomous. Unlike those developed estimators, the SW approach is applicable to outcome variables (e.g., dichotomous, continuous and count data) that have a finite distribution. Our simulation results also show that SW is a reasonable alternative to the robust variance estimator and has the advantage of reducing influential weights.

The limitation of the SW approach is the uncertainty of the influence of continuous confounders when their association with the probability of being treated and outcome is very strong. As shown in simulation studies, the sample size in some of the pseudo data sets can be different from the original data set when the confounding effect is strong. However, it is uncommon as our simulation results showed that the mean sample size approximated the original sample size. It is recommended that one always examine the difference between sample sizes in the original cohort and the pseudo cohort. When there is evidence that the sample size of the pseudo data is different from that of the original dataset, one can use the robust variance estimator with IPTWs although this latter method can produce slightly larger standard errors.

We conclude that our theoretical, simulation results and the real data example demonstrate that the use of the stabilized weights in the pseudo data preserves the sample size close to the original data. In addition, we conclude that use of stabilized weights produces the appropriate estimation of the variance of the main effect and maintains an appropriate type I error rate. Stabilized weights may be a useful tool to balance confounders between groups in observational studies.

Acknowledgements

We thank Gwyn Saylor, BS, Xiuhai Yang, MS and Junling Ren, MEd, for programming efforts, and Leslie A Wright, MA, and Jill Mesa for project management.

Source of financial support: Diabetes and Drug-Associated Hyperkalemia: Effect of Laboratory Monitoring is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) under the following grant number: R21DK075076.

References

1. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.
2. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. J Am Stat Assoc. 1984; 79:516–24.
3. D'Agostino RB. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998; 17:2265–81. [PubMed: 9802183]
4. Shepardson LB, Youngner SJ, Spero T, Rosenthal GE. Increased risk of death in patients with donot-resuscitate orders. Med Care. 1999; 37:727–37. [PubMed: 10448716]
5. Perkins SM, Tu W, Underhill MG, Zhou XH, Murray MD. The use of propensity scores in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf. 2000; 9:93–101. [PubMed: 19025807]
6. Hirano K, Imbens G. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Serv Outcomes Res Methodol. 2001; 2:259–78.
7. Allen-Ramey FC, Duong PT, Goodman DC, et al. Treatment effectiveness of inhaled corticosteroids and leukotriene modifiers for patients with asthma: An analysis from managed care data. Allergy Asthma Proc. 2003; 24:43–51. [PubMed: 12635577]
8. Lipkovic I, Adams DH, Mallinckrodt C, et al. Evaluating dose response from flexible dose clinical trials. BMC Psychiatry. 2008; 8:1–9. [PubMed: 18173833]
9. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat Med. 2004; 23:2937–60. [PubMed: 15351954]

10. Cassel, CM.; Sarndal, CE.; Wretman, JH. Some uses of statistical models in connection with the nonresponse problem In Incomplete Data in Sample Surveys III.. In: Madow, WG.; Olkin, I., editors. Symposium on Incomplete Data, Proceedings. Academic Press; New York: 1983.
11. Rosenbaum PR. Model-based direct adjustment. Journal of the American Statistical Association. 1987; 82:387–394.
12. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000; 11:561–70. [PubMed: 10955409]
13. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. Chapman and Hall; New York, NY: 1993.
14. Mooney, CZ.; Duval, RD. Bootstrapping: a non-parametric approach to statistical inference. Sage; Newbury Park CA: 1993.
15. Campbell MK, Torgerson DJ. Confidence intervals for cost-effectiveness ratios: the use of bootstrapping. JHSRP. 1997; 2:253–5.
16. Robins, JM. 1997 Proceedings of the Section on Bayesian Statistical Science. American Statistical Association; Alexandria, VA: 1998. Marginal structural models.; p. 1-10.
17. Robins JM, Hernan M, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiol. 2000; 11:550–60.
18. Clark DO, Von Korff M, Saunders K, et al. A chronic disease score with empirically derived weights. Med Care. 1995; 33:783–95. [PubMed: 7637401]
19. Huber, PJ. The behavior of maximum likelihood estimation under nonstandard conditions.. In: LeCam, LM.; Neyman, J., editors. 11 Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,1. University of California Press; 1967.
20. White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982; 50:1–25.
21. Wu AH, Aaronson KD, Bolling SF, et al. Impact of mitral valve annuloplasty on mortality risk in patients with mitral regurgitation and left ventricular systolic dysfunction. J Am Coll Cardiol. 2005; 45:381–7. [PubMed: 15680716]
22. Lohr, S. Sampling: Design and Analysis. Duxbury Press; Pacific Grove, CA: 1999.

Table 1

Sample sizes (standard deviations) and type I error rates based on 5000 replications when there is only a dichotomous variable, x_1, $\alpha_z=\alpha_y=0.69$, $\beta_{zy}=0$.

		Sample size (std)		Type I error rate (%)		
x_1=1(%)	**β_{x1z}/β_{x1y}**	**IPTW**	**SW**	**IPTW**	**SW**	**Robust variance estimator**
33.33	0.69	1000 (1.1)	500 (0.3)	21.2	4.6	4.2
	1.39	1000 (3.4)	500 (0.8)	23.1	5.2	4.2
	1.79	1000 (5.5)	500 (1.3)	22.8	5.0	3.4
50	0.69	1000 (0.9)	500 (0.2)	22.3	4.7	4.3
	1.39	1000 (3.2)	500 (0.6)	25.7	5.2	4.1
	1.79	1000 (5.6)	500 (1.0)	26.2	4.6	3.6
33.33	−0.69	1000 (0.2)	500 (0.1)	17.7	4.6	4.2
	−1.39	1000 (0.3)	500 (0.2)	16.6	5.2	4.0
	−1.79	1000 (0.7)	500 (0.4)	17.5	5.1	3.1
50	−0.69	1000 (0.2)	500 (0.1)	16.3	4.8	4.3
	−1.39	1000 (0.2)	500 (0.1)	16.1	5.2	3.8
	−1.79	1000 (0.6)	500 (0.2)	17.7	5.4	3.3

Table 2

Sample sizes (standard deviations) and type I error rates based on 5000 replications when there are a dichotomous and a continuous variable, $\alpha_z=\alpha_y=\beta_{x1z}=\beta_{x1y}=0.69$, variance($x_2$) =1 for $x_1=0$ and equal to 4 for $x_1=1$

		Means of x_2		Sample size (std)		Type I error rate (%)		
$x_1=1$(%)	β_{x2z}/β_{x2y}	$x_1=0$	$x_1=1$	IPTW	SW	IPTW	SW	Robust variance estimator
50	0.3	1	1	999 (9.7)	500 (2.0)	25.9	5.1	4.6
	0.6	1	1	999(35.9)	500 (6.7)	28.4	6.3	4.0
	1.2	1	1	995 (196.7)	499 (36.7)	35.5	12.0	3.4
	0.6	1	2	999(35.1)	500 (5.1)	32.1	5.1	4.0
	0.6	1	3	998 (38.4)	500 (4.4)	36.5	5.0	5.6
	0.6	1	4	999 (102.3)	500 (4.6)	41.5	6.0	5.8
33.33	0.3	1	1	999 (8.8)	500 (1.9)	24.0	5.0	4.3
	0.6	1	1	999 (31.1)	500 (6.1)	28.5	6.2	4.5
	1.2	1	1	996 (152.0)	499 (29.0)	32.8	9.0	3.0
	0.6	1	2	1000 (31.1)	500 (4.4)	32.9	5.0	4.4
	0.6	1	3	999 (35.7)	500 (3.6)	39.0	4.5	6.2
	0.6	1	4	1000 (49.3)	500 (3.8)	43.9	4.5	7.7
	0.6	−1	−1	999 (13.6)	500 (5.9)	17.2	5.6	3.3
	0.6	−1	−2	999 (12.9)	500 (6.8)	17.8	6.0	2.8
	0.6	−1	−3	999 (13.2)	500 (8.0)	20.4	6.8	2.2
	0.6	−1	−4	999 (16.5)	500 (9.0)	22.0	6.2	1.4

Table 3

Characteristics of patients in the original study cohort and in the pseudo cohort with Stabilized Weights

	Original cohort (n=27,355)			Pseudo cohort (n=27,407)		
Characteristic	**monitored**	**Not monitored**	**p-values**	**monitored**	**Not monitored**	**p-values**
Mean age in years (std)	60.4 (13.0)	55.5 (13.2)	<0.001	59.0 (13.1)	59.3 (13.8)	0.054
Male gender (%)	50.8	53.4	<0.001	51.4	50.7	0.30
Drug groups (%)			<0.001			0.98
ACEi	91.9	93.1		92.5	92.3	
ARB	5.70	5.5		5.4	5.6	
spironolactone	1.90	1.2		1.6	1.6	
combinations	0.50	0.2		0.5	0.5	
Kidney transplant during or prior to study entry (%)	0.30	<0.1	<0.001	0.2	0.2	0.39
Prior potassium monitoring (%)	0.90	0.93	0.67	0.9	0.9	0.89
Prior hyperkalemia diagnosis (%)	0.57	0.38	0.05	0.5	0.5	0.79
Hospitalization or emergency department visit(s) within 6 months prior to study entry (%)	23.50	19.1	<0.001	22.6	22.2	0.39
Heart failure diagnosis (%)	8.9	3.5	<0.001	7.4	7.6	0.66
Chronic kidney disease stage 3 or 4 (%)	10.0	3.0	<0.001	8.0	8.4	0.28
Median chronic disease score (5th, 95th percentile)	6 (3,11)	6 (3,9)	<0.001	6 (3,10)	6 (3,11)	0.03
Digoxin therapy (%)	4.4	1.6	<0.001	3.6	3.8	0.40
Diuretic therapy (%)	37.1	19.9	<0.001	32.2	32.8	0.27
Potassium supplement therapy (%)	13.9	4.7	<0.001	11.3	12.3	0.02

Table 4

Comparison of distribution characteristics between IPTW and SW in serum potassium monitoring example

Distribution characteristics	IPTW	SW
Mean	2.01	1.00
Median	1.49	0.95
Standard deviation	1.59	0.40
Minimum	1.01	0.42
maximum	42.59	12.39