SAS® GLOBAL FORUM 2020 #SASGF

Paper 4942-2020

Propensity Score Matching with Survey Data

Bocheng Jing, Northern California Institute for Research and Education
L. Grisell Diaz-Ramirez, University of California, San Francisco
W. John Boscardin, University of California, San Francisco
Michael A. Steinman, University of California, San Francisco and the San Francisco VA Medical Center

ABSTRACT

Clinicians and healthcare researchers often use national-representative surveys to conduct observational studies. Propensity score methods are one of the popular methods to draw estimates for key exposures, treatments, and interventions. It is surprising that propensity score methods have not been ubiquitously incorporated in survey observational studies. Some authors have proposed incorporating sampling weights as a covariate in the propensity score model, while others have used sampling weights as survey weights, that is, a weighted model. Thus, there are still inconclusive results on which method would account best for complex survey design features. In our paper, we used five different methods to calculate propensity scores that account differently for survey design features. We used both matching and weighting methods to adjust the outcome model. It was unclear which method of propensity score calculation was better. For binary or survival outcome, the key exposure seems to be insignificant in both matching and weighting outcome models.

INTRODUCTION

When estimating treatment effects in observational studies, researchers often apply propensity score methods to control for potential confounding and address systematic differences of baseline covariates between treated and control groups. Propensity score (PS) is generally calculated as the probability of treatment assignment conditional on the baseline covariates. It balances the observed baseline covariates between treated and control groups, which mimic the design of a randomized trial. In the field of population health, researchers often collect data through complex survey methods. Complex survey design creates non-independence and disproportional probability of selection in the sample units. Not counting for the complex survey design in statistical analyses would lead to errors when estimating the parameters and their variances. In our study, we focus on obtaining the propensity scores with different regression methods that account differently for the

survey features. Then we estimate the exposure effects on the matched cohort and in the weighted outcome model.

There are many different opinions to obtain propensity scores with survey data. Zanutto et al. (2006) and DuGoff et al. (2014) both suggest that survey weights and design elements can be ignored when calculating propensity scores since the propensity score model does not need to be generalized to the population. However, Ridgeway et al. (2015) suggests adopting sampling weight when calculating the propensity scores. On the other hand, in a simulation study, Austin et. al (2018) indicates inconsistent results on whether to use weighted regression for propensity score estimation. In our study, we adopt both weighted regression and non-weighted regression for propensity scores calculation.

We adjust the key exposure variable in both binary and survival outcomes models. The outcomes models are fitted on the matched cohort and using inverse probability of treatment weights (IPTW). Although Austin et al. previous simulation study suggested that adjusting exposure variable in the matched cohort for binary or survival outcomes could lead to biased effects of the exposure, resulting in poor performance, we only adjust the key exposure variable in the outcome model as the marginal effect of the exposure odds ratio/hazard ratio is our primary effect of interest. Moreover, the current SAS® procedures do not incorporate within-matched cohort correlation with the survey design features on the variance estimation. In our study, we fit the outcome models with different weighted regressions, hoping to bypass this issue.

METHOD

We fitted logistic regression models to calculate the probability of patients experienced the key exposure and incorporated the sampling weights and survey design elements into the logistic regression using five methods:

Method a: We adjusted for all the baseline covariates in the logistic regression model, without specifying survey design features and sampling weights. We used the LOGISTIC procedure to derive the propensity scores. This method follows the rationale of Zanutto et al. (2006) and DuGoff et al. (2014) mentioned in the previous section who considered that propensity score model should not include weights and survey design elements.

Method b: We adjusted for all the baseline covariates in the logistic regression, along with sampling weights. The sampling weight captures the characteristic of the population so adjusting it into the model ensures that our sample is representative of the population. It is difficult to adjust for strata because of the many levels of the strata − the degree of freedoms and limited memory of computation may hinder the estimation of the maximum likelihood. Therefore, we only adjusted for the sampling weight as covariate into the model.

Method c: We used the weighted logistic regression to calculate the propensity scores, adjusting for all the baseline characteristics. We specified the WEIGHT

statement in the logistic regression. Weighted logistic regression applies weighted maximum likelihood method to estimate the coefficients; therefore, the predicted probability from the weighted logistic regression will differ from the one without specifying the weight regression.

Method d: We used the robust variance estimator in the logistic regression to obtain the propensity scores, adjusting for the baseline covariates. Survey data are highly correlated within the strata and clusters. Therefore, the assumption of homogenous variance for the logistic regression may not be valid when using survey data. The robust variance estimator accounts for the heterogeneity of the variance and handle the correlation within clusters and strata. We created the variable, clu_strata_ID, combining both stratum and cluster together so each observation in the sample can be denoted by one variable. For example, if an observation is in cluster 2, strata 24. We created a variable that counts both cluster and stratum: 2×100 + 24 = 224. We assumed that within each ID level, the correlation structure is exchangeable. We also specified the sampling weight as well in the regression. We used the GENMOD procedure to run the logistic regression as shown in the code below:

```
proc genmod data=data.analy_dataset1;
   class clu_strata_ID gender2  race4  smoken0 hbp0 stroke0 heartfail0
heartprobl0 angina0 falls0 meds0
            lungoxy0 canceract0 cognitive0 frailscore4g0 rural0
degree4 pain0 cesbin0 diab0
            totass02g  mst0 lung0 sight0;

   model treatvar(Event='1')=age0 gender2  race4  smoken0  hbp0
stroke0  bmi0 heartfail0 heartprobl0 angina0 falls0
                                        meds0 lungoxy0 canceract0
cognitive0 frailscore4g0 rural0 degree4 pain0 cesbin0 diab0
                                        totass02g mst0 lung0 sight0
weight0 /link=logit;
   repeated subject=clu_strata_ID /type=exch;
   output out=ps_method_e p=pscore_e;
   weight pre_wgtr;
run;
```

Method e: We used the SURVEYLOGISTIC procedure to run the survey weighted logistic regression model. Sampling weight, cluster, and strata were specified to obtain the propensity scores, which generalize the probability of the key exposure to the population.

After obtaining five sets of propensity scores, we used the PSMATCH procedure to conduct the matching and weighting. For propensity score matching, we used 1:1 greedy matching algorithm, with exact matching on observation's gender. We used the 0.2 standard deviations of the logit of the propensity score as the matching standard. The standardized mean difference of baseline variables was outputted to assess the matching balance, with standardized mean differences less than 0.10 considered as evidence of acceptable balance. For weighting, we output the inverse

probability of treatment weights (IPTW) from the OUTPUT statement to estimate the average treatment effect (ATE). In the outcome model, we multiplied the IPTW weight by the sampling weight to obtain a compound weight. Then, we normalized the compound weight and included the normalized compound weight into the outcome models.

EXAMPLE

Coronary Artery Bypass Surgery (CABG) procedure and procedure of Cardiac Catheterization (PCI) are two common procedures that repair and restore normal blood flow to an obstructed heart vessel. It is meaningful to study whether one of the two procedures has a higher hazard ratio for post-surgery survival or a higher odds ratio of mortality. We created a nationally representative cohort of 1,680 community-dwelling seniors enrolled in the Health and Retirement Study (HRS) at age 65 and above from waves 1992 to wave 2012 who had gone through CABG or PCI procedures. HRS is a longitudinal study that measures the health and economic circumstances changes within aging Americans. It is a nationally representative sample of participants over the age of 50. It started in 1992 and new participants are added in the study every 6 years so that the sample remains representative of the population over age 50. There are two Primary Sampling Units and 56 strata in HRS.

In our study, we adjusted for 26 baseline covariates, most of which are questions in HRS answered by respondents prior to their procedures.

Table 1 Baseline Covariates for PS Matching and Weighting

Variables	PCI (n=1015)	CABG (n=665)
age0: R age pre procedure		
Mean (SD)	74.0 (6.60)	73.7 (6.02)
gender2: R gender. 0.female, 1.male		
0	459 (45.2%)	219 (32.9%)
1	556 (54.8%)	446 (67.1%)
race4: R race/ethnicity. 0.white, 1.black, 2.hispanic, 3.Other		
0	829 (81.7%)	572 (86%)
1	98 (9.7%)	41 (6.2%)
2	61 (6%)	45 (6.8%)
3	27 (2.7%)	7 (1.1%)
smoken0: R smoked pre procedure. 0.no, 1.yes		
0	878 (86.5%)	575 (86.5%)
1	137 (13.5%)	90 (13.5%)
hbp0: R ever had hbp pre procedure. 0.no, 1.yes		
0	339 (33.4%)	207 (31.1%)
1	676 (66.6%)	458 (68.9%)
stroke0: R ever had stroke pre procedure. 0.no, 1.yes		
0	869 (85.6%)	577 (86.8%)
1	146 (14.4%)	88 (13.2%)

bmi0: R BMI pre procedure		
Mean (SD)	28.0 (5.27)	27.7 (4.93)
heartfail0: R had history of heart failure pre procedure. 0.no, 1.yes		
0	952 (93.8%)	628 (94.4%)
1	63 (6.2%)	37 (5.6%)
heartprobl0: R had history of heart problem pre procedure. 0.no, 1.yes		
0	539 (53.1%)	365 (54.9%)
1	476 (46.9%)	300 (45.1%)
angina0: R had angina pre procedure. 0.no, 1.yes		
0	878 (86.5%)	582 (87.5%)
1	137 (13.5%)	83 (12.5%)
falls0: R fall in the last two years pre procedure. 0.no, 1.yes		
0	658 (64.8%)	467 (70.2%)
1	357 (35.2%)	198 (29.8%)
meds0: R had difficulty taking meds pre procedure. 0.no, 1. yes		
0	980 (96.6%)	648 (97.4%)
1	35 (3.4%)	17 (2.6%)
lungoxy0: R had severe lung disease requiring oxygen. 0.no, 1. yes		
0	987 (97.2%)	660 (99.2%)
1	28 (2.8%)	5 (0.8%)
canceract0: R had active malignancy pre procedure. 0.no, 1.yes		
0	993 (97.8%)	648 (97.4%)
1	22 (2.2%)	17 (2.6%)
cognitigve0: R had problem in cognitive dom pre procedure. 0.no, 1.yes		
0	949 (93.5%)	638 (95.9%)
1	66 (6.5%)	27 (4.1%)
frailscore4g0: R frailty score pre procedure:0-3+		
0	296 (29.2%)	216 (32.5%)
1	393 (38.7%)	250 (37.6%)
2	257 (25.3%)	168 (25.3%)
3	69 (6.8%)	31 (4.7%)
rural0: R lived in rural or urban pre procedure. 0.urban, 1. rural		
0	637 (62.8%)	407 (61.2%)
1	378 (37.2%)	258 (38.8%)
degree4: R highest education with 4 categories.		
0. <high school/GED	329 (32.4%)	220 (33.1%)
1. high school	343 (33.8%)	200 (30.1%)
2. some college	192 (18.9%)	113 (17%)
3. college and above	151 (14.9%)	132 (19.8%)
pain0: R pain group pre procedure. 0.no pain/mild, 1. pain		
0	713 (70.2%)	494 (74.3%)
1	302 (29.8%)	171 (25.7%)
cesbin0: R CESD score pre procedure. 0.no depressed, 1. depressed		
0	746 (73.5%)	522 (78.5%)

1	269 (26.5%)	143 (21.5%)
diab0: R ever had diabetes pre procedure. 0.no, 1.yes		
0	701 (69.1%)	454 (68.3%)
1	314 (30.9%)	211 (31.7%)
totass02g: R total wealth<=wgt median: 168,274 pre procedure. 0. no, 1.yes		
0	503 (49.6%)	360 (54.1%)
1	512 (50.4%)	305 (45.9%)
mst0: R is married or partnered pre procedure. 0.no, 1.yes		
0	368 (36.3%)	186 (28%)
1	647 (63.7%)	479 (72%)
lung0: R ever had lung disease pre procedure. 0.no, 1.yes		
0	880 (86.7%)	611 (91.9%)
1	135 (13.3%)	54 (8.1%)
sight0: R eyesight pre procedure.		
1. excellent	87 (8.6%)	44 (6.6%)
2. very good	212 (20.9%)	168 (25.3%)
3. good	430 (42.4%)	293 (44.1%)
4. fair	201 (19.8%)	115 (17.3%)
5. poor	81 (8%)	43 (6.5%)
6. blind	4 (0.4%)	2 (0.3%)
weight0: R weight in kg pre procedure		
Mean (SD)	80.6 (16.76)	81.5 (16.78)

We calculated the propensity scores for PCI and CABG and matched the patients who had PCI procedure to patients who had CABG procedure. The two outcomes were time to death during the post-procedure follow-up time and binary mortality during the follow-up. For the outcome model, we only adjusted for the exposure variable (PCI vs. CABG) in the matched cohort or in the IPTW weighted model in the full cohort. For the binary outcome, we used both PROC LOGISTIC and PROC SURVEYLOGISTIC to estimate the treatment effect as they represent the sample ATE and the population ATE, respectively. Similarly, for the survival outcome, we used PROC PHREG and PROC SURVEYPHREG, and in addition, we also fitted a model that uses the robust sandwich estimator to estimate the treatment effect.

RESULTS

We examined the standardized mean difference for key baseline covariates for PS score matching and weighting using the five propensity score calculation methods.

Table 2 Standardized Mean Difference Heatmap

Variable	StdMeanDiff_a	StdMeanDiff_b	StdMeanDiff_c	StdMeanDiff_d	StdMeanDiff_e
age0	0.03480	0.02196	0.00120	0.00526	0.00120
angina0	-0.04737	-0.02874	-0.05153	0.00935	-0.05153
bmi0	-0.04019	0.03606	0.02172	-0.02985	0.02172

canceract0	0.02097	0.01060	0.00000	0.00000	0.00000
cesbin0	-0.00747	0.02644	0.02216	0.01844	0.02216
cognitive0	-0.01426	-0.02163	0.03526	0.09153	0.03526
diab0	0.01373	0.00000	0.01358	0.03390	0.01358
falls0	0.01362	0.04133	0.05390	0.07063	0.05390
gender2	0.00000	0.00000	0.00000	0.00000	0.00000
hbp0	-0.00341	-0.03446	-0.01685	0.00673	-0.01685
heartfail0	-0.05413	-0.03421	0.01338	0.06013	0.01338
heartprobl0	-0.04793	0.00969	0.03792	0.02524	0.03792
lung0	-0.01033	-0.02090	-0.01022	-0.01531	-0.01022
lungoxy0	-0.01216	0.00000	-0.02405	0.01201	-0.02405
meds0	-0.02800	0.00000	0.05539	0.09217	0.05539
mst0	0.02054	0.01731	0.00339	0.01690	0.00339
pain0	-0.00356	0.06121	-0.05282	-0.01055	-0.05282
rural0	0.00656	0.00000	0.00324	-0.03887	0.00324
smoken0	0.02794	-0.01884	-0.07370	-0.03679	-0.07370
stroke0	0.00462	-0.04668	0.02739	0.03191	0.02739
totass02g	0.02871	0.02581	0.00947	-0.00315	0.00947
weight0	-0.03144	0.01961	0.00356	-0.03953	0.00356

The color scale in Table 2 represents the magnitude of the standardized mean difference: if the difference is less than 0, it was presented in red color scale. As the magnitude of the difference is higher, the color scale turns redder. The blue color scale applied to the difference greater than 0, as the difference is higher, the color scale is bluer. When examining the standardized mean differences across different methods in Table 2, we did not find one method that balanced the variables significantly better than the other. All the standardized mean differences for the baseline covariates were well-balanced as they were all less than 0.10 in absolute values. We noticed that method c (weighted regression by PROC LOGISTIC) and method e (Specified survey design by PROC SURVEYLOGISTIC) have produced the same standardized mean difference. This was not surprising as both methods applied sampling weight in the weighted logistic regression, resulting in the same propensity score estimations.

Table 3.1 Binary Outcome Model for Matched Cohort

	Proc logistic		Proc surveylogistic	
method	OR 95% CI	P-value	Odds 95% CI	P-value
a	0.696 (0.556, 0.870)	0.0015	0.707 (0.540, 0.926)	0.0127
b	0.799 (0.638, 1.001)	0.051	0.853 (0.639, 1.138)	0.2725
c	0.765 (0.611, 0.954)	0.0175	0.788 (0.583, 1.064)	0.1178
d	0.784 (0.627, 0.979)	0.0316	0.809 (0.598, 1.094)	0.1648
e	0.764 (0.611, 0.954)	0.0175	0.788 (0.583, 1.064)	0.1178

Table 3.2 Survival Outcome Model for Matched Cohort

	proc phreg		proc phreg (sandwich estimator)		proc surveyphreg	
method	HR 95% CI	P-value	HR 95% CI	P-value	HR 95% CI	P-value
a	0.813 (0.701, 0.943)	0.0062	0.813 (0.689, 0.959)	0.0143	0.812 (0.679, 0.972)	0.0242
b	0.878 (0.758, 1.017)	0.0817	0.878 (0.745, 1.035)	0.1204	0.901 (0.735, 1.104)	0.3078
c	0.865 (0.747, 1.001)	0.0519	0.865 (0.740, 1.011)	0.0682	0.878 (0.729, 1.057)	0.1642
d	0.871 (0.752, 1.008)	0.0639	0.871 (0.747, 1.015)	0.0775	0.869 (0.732, 1.032)	0.108
e	0.865 (0.747, 1.001)	0.0519	0.865 (0.740, 1.011)	0.0682	0.878 (0.729, 1.057)	0.1642

When using the matched cohort to estimate the treatment effect, without adjusting for survey features (proc logistic), we obtained significant difference of treatment effect in the binary outcome model (except method b). That is, the PCI procedure was more protective compared to the CABG procedure. However, as Table 3.1 indicates, using weighted regression plus specifying survey feature (proc surveylogistic) altered the standard error estimation, which leads to insignificant treatment effect. The exception was method a, where we did not incorporate survey features and sampling weights into the propensity score calculation. For all five methods, PROC SURVEYLOGISTIC produced odds ratio higher than those obtained in PROC LOGISTIC, as well as wider confidence intervals (larger standard error) after incorporating survey features.

For the survival outcome model, as shown in Table 3.2, PCI procedure had significant protective effect compared to CABG procedure only in method a, and this was consistent regardless of the survey features or probability weights included. On the other hand, no significant treatment effects were obtained in method b through e, regardless of the survey features and sampling weight used.

Table 3.3 Binary Outcome Model for PS Weighting

	Proc logistic		Proc surveylogistic	
method	OR 95% CI	P-value	Odds 95% CI	P-value
a	0.907 (0.748, 1.099)	0.3202	0.907 (0.659, 1.248)	0.5424
b	0.928 (0.766, 1.125)	0.4475	0.928 (0.674, 1.278)	0.6418
c	0.916 (0.756, 1.111)	0.3726	0.916 (0.676, 1.243)	0.5669
d	0.946 (0.780, 1.146)	0.5705	0.946 (0.668, 1.340)	0.7498
e	0.916 (0.756, 1.111)	0.3726	0.916 (0.676, 1.243)	0.5669

Table 3.4 Survival Outcome Model for PS Weighting

	proc phreg		proc phreg (sandwich estimator)		proc surveyphreg	
method	HR 95% CI	P-value	HR 95% CI	P-value	HR 95% CI	P-value
a	0.908 (0.799, 1.033)	0.1431	0.908 (0.779, 1.059)	0.2206	0.908 (0.769, 1.073)	0.2518
b	0.909 (0.799, 1.034)	0.1455	0.909 (0.778, 1.061)	0.2253	0.909 (0.769, 1.074)	0.2545

c	0.916 (0.806, 1.042)	0.1833	0.916 (0.789, 1.064)	0.2522	0.916 (0.779, 1.077)	0.2839
d	0.920 (0.809, 1.047)	0.2082	0.920 (0.791, 1.071)	0.2836	0.920 (0.781, 1.085)	0.3157
e	0.916 (0.806, 1.042)	0.1833	0.916 (0.789, 1.064)	0.2522	0.916 (0.779, 1.077)	0.2839

When using the compound weights or the compound normalized weights (derived from the IPTW weights) to estimate the treatment effect and only adjusting for the exposure variable in the model), both binary outcome model (Table 3.3) and survival outcome model (Table 3.4) did not have significant treatment effects. Given that both method c and e used the weighted regression to produce the same propensity score estimations, plus the IPTW method retained the full cohort, the treatment effects from method c and method e are identical.

DISCUSSION

In this study, we applied different methods to obtain the propensity score for the exposure variable and used both matching and weighting method to estimate the treatment effects. All propensity score methods achieved good balance for baseline covariates. In general, the treatment effect in the outcome models was not significant when survey features and sampling weights were specified − in both binary and survival outcome models− and this was consistent in both cohorts matching and PS weighting. On the other hand, when we obtained the propensity score without specifying survey features, the treatment effect was significant in the matched cohort for both outcome models. The unweighted regression point estimates are smaller and have narrower confidence intervals compared to the weighted regression results.

None of the model demonstrates significant treatment effect in the IPTW methods. However, we only estimated the marginal effect of the exposure variable. The conditional effect of the exposure variable in IPTW outcome models will likely differ from the marginal effect.

Another important aspect to consider is that the variance of the treatment effect in the matching cohort outcome model needs to consider both within-matched sets correlation and survey features. In our study, we ignored the within-matched cohort correlation, that is we only applied the survey features when estimating the variance of the treatment effect. Currently, there is limited research on this topic, and we have limited knowledge on how the within-matched cohort correlation would impact the treatment effect. Thus, further research should be conducted in this area.

Based upon our study results, it seems not adjusting for survey features and incorporate the sampling weights as survey weights into the model to estimate the propensity scores. This has aligned with Zanutto et al. (2006) and DuGoff et al. (2014) conclusions. For the outcome models, adjusting for the survey design elements and sampling weights in PROC SURVEYLOGISTIC or PROC SURVEYPHREG on the matched cohort to obtain the marginal effect estimation of the key exposure

seems to yield significant results. For IPTW weighting method, we suggest that incorporating the survey design features and the compound weight in the outcome model. Additionally, we suggest looking into the conditional effects of the exposure in the outcome models.

SAMPLE CODES

```
Method e: logistic model with all survey features specified;

proc surveylogistic data=data.analy_dataset1;
class gender2  race4  smoken0 hbp0 stroke0 heartfail0 heartprobl0
angina0 falls0 meds0
            lungoxy0 canceract0 cognitive0 frailscore4g0 rural0
degree4 pain0 cesbin0 diab0
            totass02g  mst0 lung0 sight0;
cluster cluster;
strata stratum;
 model treatvar(Event='1')=age0 gender2  race4  smoken0  hbp0  stroke0
bmi0 heartfail0 heartprobl0 angina0 falls0 meds0 lungoxy0 canceract0
cognitive0 frailscore4g0 rural0 degree4 pain0 cesbin0 diab0
totass02g mst0 lung0 sight0 weight0;
output out=ps_method_f p=pscore_f;

weight pre_wgtr;
run;

ods graphics on;
ods trace on;
ods output StdDiff=stdDiff_f;
proc psmatch data=ps_method_f region=cs;
   class treatvar gender2  race4  smoken0 hbp0 stroke0 heartfail0
heartprobl0 angina0 falls0 meds0
            lungoxy0 canceract0 cognitive0 frailscore4g0 rural0
degree4 pain0 cesbin0 diab0
            totass02g  mst0 lung0 sight0;
   psdata treatvar=treatvar(Treated='1') ps=pscore_f;
   match method=greedy(K=1 order=RANDOM(Seed=13)) exact=gender2
stat=lps caliper=0.20;
   assess ps var=(age0 gender2  race4  smoken0  hbp0  stroke0  bmi0
heartfail0 heartprobl0 angina0 falls0
                                        meds0 lungoxy0 canceract0
cognitive0 frailscore4g0 rural0 degree4 pain0 cesbin0 diab0
                                        totass02g  mst0 lung0 sight0
weight0) / weight=none;
   output out(obs=match)=out_f lps=_Lps matchid=_MatchID;
run;


**binary and survival outcome model for matching cohort.;

proc logistic data=out_f;
```

```
      class treatvar(ref="0") /param=ref;
      model death=treatvar;
      roc;
run;

proc surveylogistic data=out_f;
      class  treatvar(ref="0") /param=ref;
      strata stratum;
      cluster cluster;
      model death=treatvar;
      weight pre_wgtr;
run;

proc phreg data=out_a  ;
      class treatvar(ref="0") /param=ref;
      model mortality*death(0)=treatvar;
      HAZARDRATIO treatvar;
run;

proc phreg data=out_a covsandwich(aggregate) ;
      id clu_strata_ID;
      class treatvar(ref="0") /param=ref;
      model mortality*death(0)=treatvar;
      HAZARDRATIO treatvar;
run;

proc surveyphreg data=out_a;
      class treatvar(ref="0") /param=ref;
      strata stratum;
      cluster cluster;
      weight pre_wgtr;
      model mortality*death(0)=treatvar;

run;


*IPTW weighting methods;
proc psmatch data=ps_method_f region=cs;
   class treatvar gender2  race4  smoken0 hbp0 stroke0 heartfail0
heartprobl0 angina0 falls0 meds0
            lungoxy0 canceract0 cognitive0 frailscore4g0 rural0
degree4 pain0 cesbin0 diab0
            totass02g  mst0 lung0 sight0;
   psdata treatvar=treatvar(Treated='1') ps=pscore_f;
   match method=greedy(K=1 order=RANDOM(Seed=13)) exact=gender2
stat=lps caliper=0.20;
   assess lps var=(age0 gender2  race4  smoken0  hbp0  stroke0  bmi0
heartfail0 heartprobl0 angina0 falls0
                                        meds0 lungoxy0 canceract0
cognitive0 frailscore4g0 rural0 degree4 pain0 cesbin0 diab0
                                        totass02g  mst0 lung0 sight0
weight0) /  varinfo plots=all weight=none;
```

```
   output out(obs=all)=out_f_wt lps=_Lps atewgt=ate_f;
run;

proc sql;
      create table out_f_wt1 as
      select *, pre_wgtr*ate_f as att_wt_f,
pre_wgtr*ate_f/mean(pre_wgtr*ate_f) as att_wt_f_norm
      from out_f_wt;
quit;

***** applying weighting to outcome models;

proc logistic data=out_f_wt1;
      class  treatvar(ref="0") /param=ref;
      model death= treatvar;
      weight att_wt_f_norm;
run;

proc surveylogistic data=out_f_wt1;
      class  treatvar(ref="0") /param=ref;
      model death=treatvar;
      strata stratum;
      cluster cluster;
      weight pre_wgtr ;
run;

proc phreg data=out_f_wt1  ;
      class treatvar(ref="0") /param=ref;
      model mortality*death(0)=treatvar;
      HAZARDRATIO treatvar;
      weight att_wt_f_norm;
run;

proc phreg data=out_f_wt1 covsandwich(aggregate) ;
      id clu_strata_ID;
      class treatvar(ref="0") /param=ref;
      model mortality*death(0)=treatvar;
      HAZARDRATIO treatvar;
      weight att_wt_f_norm;
run;

proc surveyphreg data=out_f_wt1;
      class treatvar(ref="0") /param=ref;
      model mortality*death(0)=treatvar;
      strata stratum;
      cluster cluster;
      weight pre_wgtr;

run;
```

REFERENCE

1. Austin, P.C., "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies", Multivariate Behavioral Research, 46: 399-424, 2011.
2. Austin, P.C., Jembere, N., Chiu, M., "Propensity Score Matching and Complex Surveys", Statistical Methods in Medical Research, Vol. 27(4) 1204-1257, 2018.
3. Health and Retirement Study, (RAND HRS Data, Version P) public use data set. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI, (2016).
4. RAND HRS Data, Version P. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA (August 2016).
5. Karabon, P., "Applying Propensity Score Methods to Complex Survey Data Using PROC PSMATCH", Proceedings of the SAS® Global Forum 2019, Paper 3634-2019, https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3634-2019.pdf.
6. Yuan, Y., Yung, Y., Stokes, M., "Propensity Score Methods for Causal Inference with the PSMATCH Procedure", Proceedings of the SAS® Global Forum 2017, Paper 332-2017, https://support.sas.com/resources/papers/proceedings17/SAS0332-2017.pdf.
7. Austin, P.C., Grootendorst, P., Normand, ST., Anderson, GM., "Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: A Monte Carlo study". Statistics in Medicine, 2007, 26: 754 -768.
8. Gail, MH., Wieand, S., Piantadosi, S., "Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates", Biometrika, 1984, 71: 431-444.
9. Zanutto EL. "A comparison of propensity score and linear regression analysis of complex survey data". J Data Sci, 2006, 4: 67–91.
10. DuGoff, EH., Schuler, M., and Stuart, EA., "Generalizing observational study results: applying propensity score methods to complex surveys", Health Service Research, 2014, 49: 284–303.
11. Austin, PC., "The performance of different propensity score methods for estimating marginal hazard ratios", Statistics in Medicine, 2013, 32: 2837-2849.
12. Austin, PC., "The performance of different propensity score methods for estimating marginal odds ratios", Statistics in Medicine, 2007, 26: 3078-3094.
13. Kuss, O., "How to Use SAS for Logistic Regression with Correlated Data", SUGI 27, Paper261-27, https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p261-27.pdf
14. Austin, PC., "Balance diagnositics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples", Statistics in Medicine, 2009, 28: 3083-3107.

ACKNOWLEDGMENTS

Our work is supported by the NIA grant and University of California, San Francisco Pepper Center (P30 AG044281), which promotes promising research aimed at better understanding and addressing late-life disability in vulnerable populations.

CONTACT INFORMATION

Bocheng Jing
NCIRE-The Veterans Health Research Institute
bocheng.jing@ncire.org

L. Grisell Diaz-Ramirez
University of California, San Francisco
grisell.diaz-ramirez@ucsf.edu

W. John Boscardin
University of California, San Francisco
john.boscardin@ucsf.edu

Michael A. Steinman
University of California, San Francisco and the San Francisco VA Medical Center
mike.steinman@ucsf.edu