

Applying Propensity Score Methods to Complex Survey Data Using PROC PSMATCH

Patrick Karabon, Oakland University William Beaumont School of Medicine

ABSTRACT

A relatively new procedure in SAS/STAT® software, PROC PSMATCH allows users to perform propensity score methods for observational study designs. Complex survey data sets are increasingly being utilized in many fields. Sampling weights, strata, and clusters provided with these data sets are important to include when calculating population-based estimates. Applying propensity score methods to complex surveys is possible with PROC PSMATCH; however, additional steps are required to properly account for design elements with these data sets to generalize the results to the population.

This paper will discuss working with complex survey data sets and propensity score methods together. An illustrative example demonstrates the use of PROC PSMATCH in conjunction with other SAS/STAT® procedures to obtain population-based estimates with propensity score methods. Additional steps needed for variable balance assessment and estimation of treatment effects are highlighted.

INTRODUCTION

Confounding bias is a common bias found in observational, or nonrandomized, data. While adjusting for covariates in a multivariate regression model is one common method to address confounding bias, propensity score methods may have some advantages over traditional regression adjustment (Austin, 2011). The PSMATCH procedure, recently released in SAS/STAT® software 14.2, is the first procedure designed specifically to fit propensity score models from observational data. From PROC PSMATCH, propensity scores can be calculated and numerical/graphical assessments of variable balance between treatment and control groups are output to the results window. A data set with relevant propensity score output can be generated from PROC PSMATCH to use in another procedure for subsequent outcome analysis. Users are able to apply propensity score weighting, propensity score stratification, or propensity score matching methods. Yuan (2017) presents a comprehensive review of propensity score methods using PROC PSMATCH.

Complex survey data sets are often utilized in many fields, including healthcare, policy research, business, and education. Variables for sampling weights, strata, and clusters are included in these data sets and are necessary to include in analysis for appropriate calculation of population-based estimates. A recent study has shown that the majority of analyses using these data sets incorrectly account for sampling weights, strata, and clusters, which can lead to inaccurate and biased results (West, 2016). Propensity score methods can add an extra layer of complexity to an analysis and properly accounting for survey design elements within the propensity score methodology is important.

APPLYING PROPENSITY SCORE METHODS TO COMPLEX SURVEY DATA

Historically, literature addressing propensity score methods in the context of complex survey data has been limited (Zanutto, 2006). DuGoff (2014) proposed a method to account for sampling weights, strata, and clusters located within complex survey data sets. This method allows for population-based estimates that reflect the target population of the complex survey data set instead of estimates that reflect only the survey sample itself.

When combining propensity score methods with complex survey data, the type of outcome variable determines what propensity score methods are appropriate. Table 1 illustrates appropriate methods with a check mark (√) while inappropriate methods are marked with an X. Austin (2018) provides a detailed technical discussion on the pitfalls of propensity score matching for complex survey data.

Propensity Score Method	Continuous Outcome	Binary Outcome	Time-to-Event Outcome
Weighting*	√	√	√
Stratification	√	√	√
Matching**	√	X	X

* Weighting includes both Inverse Probability of Treatment (IPTW)/Average Treatment Effect (ATE) weighting and Average Treatment Effect of the Treated (ATT) weighting

** Variance estimates may be biased as standard errors for matched samples using complex survey data sets have not been developed (Austin, 2018)

Table 1. Propensity Score Methods Applied to Complex Survey Data

BUILDING THE PROPENSITY SCORE MODEL WITH PROC PSMATCH

To begin applying propensity score methods to complex survey data sets, PROC PSMATCH is used to build a propensity score model. The user adds the complex survey sampling weight as a covariate into the PSMODEL statement. Adding strata and cluster variables is recommended; however, this may not be possible when there are a lot of strata or clusters due to limited degrees of freedom and other model overfitting issues (DuGoff, 2014). In certain very large complex survey data sets, memory issues may occur when adding strata or clusters into the PSMODEL statement.

The ASSESS statement in PROC PSMATCH obtains both numeric measures (standardized differences) and plots of balance diagnostics from the propensity score model. These figures and numbers guide a user into refitting a propensity score model over and over again using different methods and variables until the user is satisfied with the fit of the propensity score model (Guo and Fraser, 2015). Users can compare the balance of the survey weight (and strata and clusters if included) between the treatment and control groups just like the other covariates.

The OUTPUT statement is used to obtain and output the relevant information from the propensity score model for subsequent outcome analysis using another procedure in the SAS/STAT® software.

OUTCOME ANALYSIS USING COMPLEX SURVEY DATA PROCEDURES

If performing propensity score weighting, a final weight is obtained by multiplying the sampling weight from the complex survey data set with the propensity score weight from PROC PSMATCH. The outcome regression model is run with the final calculated weight in the WEIGHT statement and the data set provided strata and clusters in the STRATA and CLUSTER statements, respectively (DuGoff, 2014).

With propensity score stratification, an outcome model is first run for each individual propensity score strata to assess heterogeneity in treatment effect across the strata. This can be done using a DOMAIN statement. Then, the estimates in each strata are combined by adjusting for the strata as a covariate in the outcome model (DuGoff, 2014; Roberts, 2010). The survey weight is used in the WEIGHT statement while strata and clusters are also included into the outcome model.

In every outcome model, sampling weights, strata, and clusters are included in the model; therefore, SAS/STAT® survey analysis procedures are necessary. The SURVEYREG procedure is the appropriate procedure for continuous outcomes while the SURVEYLOGISTIC procedure and SURVEYPHREG procedure are used for binary and time-to-event outcomes, respectively. Doubly robust methods can be used.

ILLUSTRATIVE EXAMPLE

BACKGROUND

The Behavioral Risk Factor Surveillance System (BRFSS) is an annual survey conducted by the United States Centers for Disease Control and Prevention (CDC). In this survey, a stratified sample of Americans is contacted and asked about their use of preventative health services, in addition to other questions (US CDC, 2017). BRFSS contains variables for sampling weights (_LLCPWT) along with the design element variables for strata (_STSTR) and clusters (_PSU). A unique de-identified variable is provided as a respondent ID (SEQNO).

For this illustrative example, the BRFSS 2016 data set is used to determine whether married respondents have significantly greater odds of receiving their flu shot in the past 12 months. Previous research has demonstrated that married respondents are significantly more likely to receive a flu shot; however, other confounders are present (Lochner, 2011).

The binary outcome variable is whether a respondent received their flu shot in the past year (FLUSHOT) while the binary treatment variable is whether or not the respondent is married (MARRIED). Binary confounders are chosen from the literature (Lochner, 2011) and include being Non-Hispanic Black (BLACK) or Hispanic (HISPANIC), High School Graduate (HSGRAD), Low Income (POVERTY), Current Smoker (SMOKER), Having a Primary Care Physician (DOCTOR), Health Insurance (INSURANCE), and Fair/Poor Health Status (BADHEALTH). All variables are coded as 1 for Yes and 0 for No for simplicity.

UNADJUSTED ANALYSIS AND REGRESSION ADJUSTMENT

PROC SURVEYLOGISTIC is used to examine the univariate association between being married and receiving a flu shot in the past year while accounting for the sampling weights, strata, and clusters provided with the BRFSS. The code for the unadjusted analysis is as follows:

```
PROC SURVEYLOGISTIC DATA = BRFSS;  
  WEIGHT _LLCPWT;  
  STRATA _STSTR;  
  CLUSTER _PSU;  
  CLASS FLUSHOT(REF = "0") MARRIED(REF = "0") / PARAM = GLM;  
  MODEL FLUSHOT = MARRIED;  
RUN;
```

Regression adjustment, also known as multivariate logistic regression for survey data in this case, examines the association between marriage and receiving a flu shot in the past year while adjusting for all potential confounding variables. The code for the regression adjustment is as follows:

```
PROC SURVEYLOGISTIC DATA = BRFSS;
  WEIGHT _LLCPWT;
  STRATA _STSTR;
  CLUSTER _PSU;
  CLASS FLUSHOT(REF = "0") MARRIED(REF = "0") BLACK HISPANIC HSGRAD POVERTY
    SMOKER DOCTOR INSURANCE BADHEALTH / PARAM = GLM;
  MODEL FLUSHOT = MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR
    INSURANCE BADHEALTH;
RUN;
```

The table below shows the results, which indicate that married respondents have significantly higher odds of receiving a flu shot in the past year as compared to the unmarried respondents.

Analysis	Odds Ratio	95% Confidence Limits		P-Value
Unadjusted Model	1.379	1.347	1.413	< 0.0001
Regression Adjustment	1.179	1.149	1.211	< 0.0001

Table 2. Unadjusted and Regression Adjustment Results

COMPARING BASELINE MEASUREMENTS BETWEEN GROUPS

Before drawing a definitive conclusion from the results of the table above, a quick comparison of baseline characteristics between groups is necessary. As shown in the Table below, there are significant differences in all the baseline characteristics between the treated (respondents who are married) and control (respondents who are not married) groups (all $P < 0.0001$). Standardized Differences in the furthest right column were obtained from PROC PSMATCH for the unadjusted data. The Standardized Differences indicate that there is some imbalance between the groups, especially where Standardized Differences are greater than 0.25. Since inadequacies in variable balance cannot be addressed with traditional regression adjustment, applying propensity score methods may have benefit (Austin, 2011).

	Married (n = 127,548,853)	Not Married (n = 126,602,283)	P-Value	Standardized Difference (SD)
Non-Hispanic Black	7.23%	15.98%	< 0.0001	0.238
Hispanic	14.60%	18.58%	< 0.0001	0.076
HS Graduate	88.20%	83.08%	< 0.0001	0.181
Poverty	12.91%	34.30%	< 0.0001	0.621
Current Smoker	7.89%	13.27%	< 0.0001	0.198
Has a PCP	83.56%	71.91%	< 0.0001	0.194
Health Insurance	91.69%	83.55%	< 0.0001	0.182
Bad Health	15.23%	20.61%	< 0.0001	0.223

Table 3. Baseline Characteristics between Treated and Control Groups

In this illustrative example, we will explore three different propensity score methods: propensity score IPTW/ATE weighting, propensity score ATT weighting, and propensity score stratification. Propensity score matching is not feasible because the outcome variable in this study is binary.

PROPNESITY SCORE IPTW/ATE WEIGHTING

First, PROC PSMATCH specifies the propensity score model:

```
PROC PSMATCH DATA = BRFSS REGION = ALLOBS;  
  CLASS MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR INSURANCE  
    BADHEALTH;  
  PSMODEL MARRIED(TREATED = "1") = BLACK HISPANIC HSGRAD POVERTY SMOKER  
    DOCTOR INSURANCE BADHEALTH _LLCPWT;  
  ASSESS PS VAR = (BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR INSURANCE  
    BADHEALTH _LLCPWT) / VARINFO PLOTS = ALL WEIGHT = ATEWGT;  
  OUTPUT OUT(OBS = ALL) = ATE ATEWGT = _ATE_;  
RUN;
```

The REGION = ALLOBS option is used in the PROC PSMATCH statement because all observations are included to accurately reflect the entire target population instead of a subset of that population.

The PSMODEL statement runs the logistic regression used to calculate the propensity scores. Since this is a complex survey data set, the sampling weight (_LLCPWT) is added as a covariate. Strata and cluster variables are not included here because there are too many strata and clusters relative to total observations in this data set.

The ASSESS statement produces tables and figures that assess variable balance between the treated and control groups. The PLOTS = ALL option produces all plots; however, if the data set is very large, run times may be long. The option WEIGHT = ATEWGT indicates that the treatment effect of interest is the ATE.

Finally, the OUTPUT statement produces a dataset called ATE with ATE weights as the variable _ATE_. The figure below shows the graphical variable balance provided by the procedure.

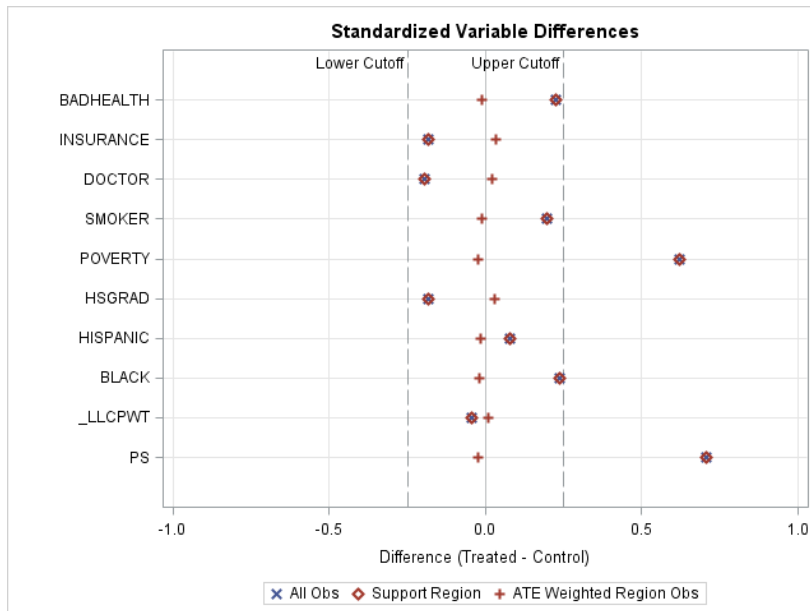


Figure 1. IPTW/ATE Weighting Variable Balance Assessment

All standardized differences are between -0.25 and 0.25, which indicates good variable balance between the treated and control groups.

A subsequent DATA step merges the original BRFSS dataset with the IPTW/ATE weights generated from PROC PSMATCH. The IPTW/ATE weights that were output from PROC

PSMATCH (`_ATE_`) are multiplied by the BRFSS-provided sampling weights (`_LLCPWT`) to obtain a final weight (ATE) to use in the subsequent outcome analysis. The code is as follows:

```
DATA BRFSS;
  MERGE BRFSS ATE;
  BY SEQNO;

  ATE = _ATE_*_LLCPWT;
RUN;
```

An outcome model is specified using PROC SURVEYLOGISTIC. This doubly robust model includes the final weight variable (ATE) calculated in the previous data step in addition to strata (`_STSTR`) and clusters (`_PSU`). The results of the model are displayed in the table below, while the code to generate this output is as follows:

```
PROC SURVEYLOGISTIC DATA = BRFSS;
  WEIGHT ATE;
  STRATA _STSTR;
  CLUSTER _PSU;
  CLASS FLUSHOT(REF = "0") MARRIED (REF = "0") BLACK HISPANIC HSGRAD POVERTY
    SMOKER DOCTOR INSURANCE BADHEALTH / PARAM = GLM;
  MODEL FLUSHOT = MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR
    INSURANCE BADHEALTH;
RUN;
```

Odds Ratio	95% Confidence Limits	P-Value
1.177	1.146 1.209	< 0.0001

Table 4. Results for Propensity Score IPTW/ATE Weighting

PROPENSITY SCORE ATT WEIGHITNG

Similar to Propensity Score IPTW/ATE weighting, PROC PSMATCH for ATT weighting first specifies a propensity score model:

```
PROC PSMATCH DATA = BRFSS REGION = ALLOBS;
  CLASS MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR INSURANCE
    BADHEALTH;
  PSMODEL MARRIED(TREATED = "1") = BLACK HISPANIC HSGRAD POVERTY SMOKER
    DOCTOR INSURANCE BADHEALTH _LLCPWT;
  ASSESS PS VAR = (BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR INSURANCE
    BADHEALTH _LLCPWT) / VARINFO PLOTS = ALL WEIGHT = ATTWGT;
  OUTPUT OUT(OBS = ALL) = ATT ATTWGT = _ATT_;
RUN;
```

The code for PROC PSMATCH is almost identical for performing IPTW/ATE weighting or ATT weighting, but there are a few minor differences. In the ASSESS statement, the WEIGHT option is changed to WEIGHT = ATTWGT to produce balance diagnostics for ATT weights instead of IPTW/ATE weights. The OUTPUT statement now produces a data set, called ATT, that includes ATT weights as the variable `_ATT_`. The figure on the following page shows the graphical variable balance assessment from the ATT weighting model.

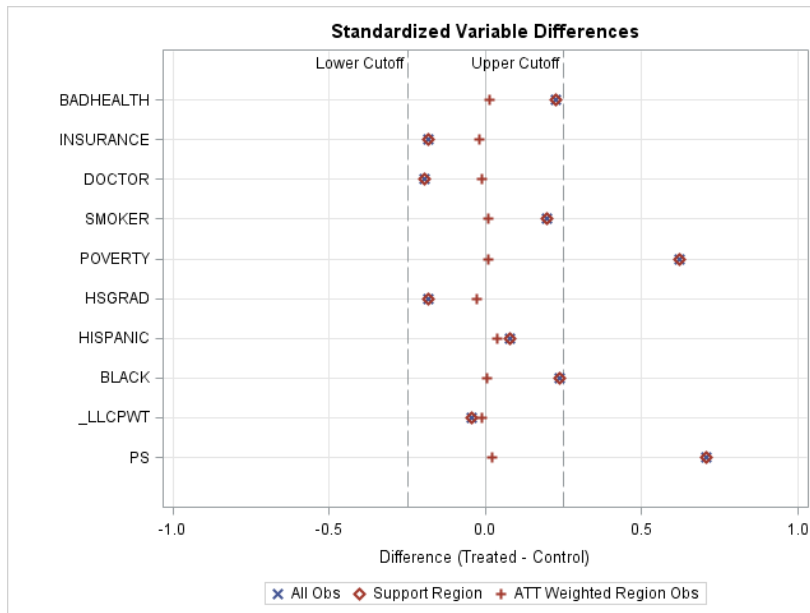


Figure 2. ATT Weighting Variable Balance Assessment

There is good variable balance as all standardized differences are now between -0.25 and 0.25. The DATA step below merges the propensity scores with the BRFSS data set and multiples ATT weights produced with the OUTPUT statement from PROC PSMATCH (_ATT_) and the sampling weights from the BRFSS (_LLCPWT) to create a final weight (ATT). The code for this data step is as follows:

```
DATA BRFSS;
  MERGE BRFSS ATT;
  BY SEQNO;

  ATT = _ATT_*_LLCPWT;
RUN;
```

An outcome model is specified using PROC SURVEYLOGISTIC with the final weight (ATT), strata (_STSTR), and clusters (_PSU) included. The table following the code shows the corresponding treatment effect estimate, again showing significantly higher odds of flu shows within the past year for married respondents.

```
PROC SURVEYLOGISTIC DATA = BRFSS;
  WEIGHT ATT;
  STRATA _STSTR;
  CLUSTER _PSU;
  CLASS FLUSHOT(REF = "0") MARRIED(REF = "0") BLACK HISPANIC HSGRAD POVERTY
    SMOKER DOCTOR INSURANCE BADHEALTH / PARAM = GLM;
  MODEL FLUSHOT = MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR
    INSURANCE BADHEALTH;
RUN;
```

Odds Ratio	95% Confidence Limits		P-Value
1.188	1.157	1.220	< 0.0001

Table 5. Results for Propensity Score ATT Weighting

PROPNESITY SCORE STRATIFICATION

PROC PSMATCH is also used to create a propensity score model for stratification. The code is as follows:

```
PROC PSMATCH DATA = BRFSS REGION = ALLOBS;
  CLASS MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR INSURANCE
        BADHEALTH;
  PSMODEL MARRIED(TREATED = "1") = BLACK HISPANIC HSGRAD POVERTY SMOKER
        DOCTOR INSURANCE BADHEALTH _LLCPWT;
  STRATA NSTRATA = 5;
  ASSESS PS VAR = (BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR INSURANCE
        BADHEALTH _LLCPWT) / VARINFO PLOTS = ALL;
  OUTPUT OUT(OBS = ALL) = STRATA STRATA = _STRATA_;
RUN;
```

The STRATA statement is now added in order to indicate propensity score stratification is desired. The NSTRATA = 5 part of the statement requests propensity score quintiles. The table below shows the quintile ranges, which were produced as part of PROC PSMATCH. The subsequent figure shows the variable balance of the propensity score strata.

Strata	Propensity Score Range	Married (Treated)	Not Married (Control)	Total Population Size
1	0.048-0.547	42,599,308	46,178,263	88,777,571
2	0.547-0.641	23,848,303	23,089,374	46,937,677
3	0.641-0.649	41,794,711	21,846,430	41,794,711
4	0.649-0.650	18,344,198	17,788,952	36,133,150
5	0.650-0.661	19,597,412	20,910,615	40,508,027

Table 6. Propensity Score Stratification Table

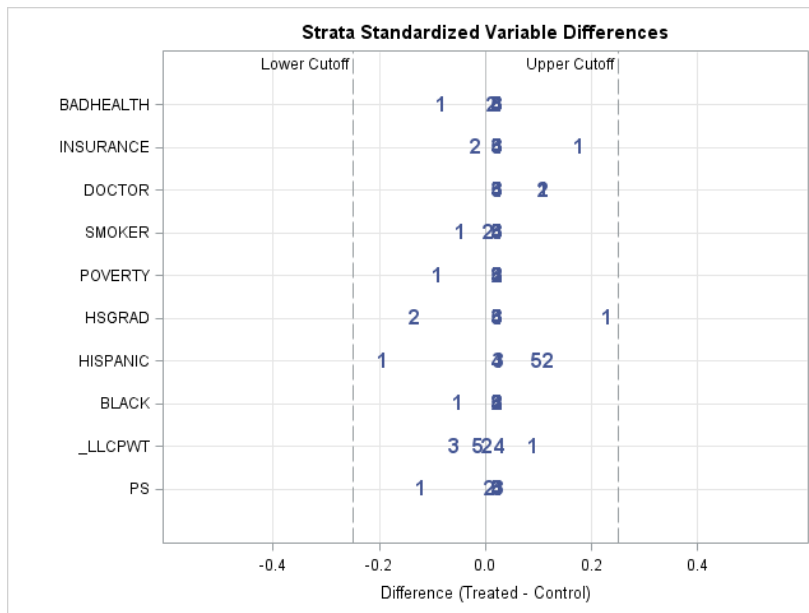


Figure 3. Stratification Variable Balance Assessment

All standardized differences are between -0.25 and 0.25, indicating good variable balance. PROC SURVEYLOGISTIC obtains estimates of treatment effect for each of the five strata. The DOMAIN statement is used to get treatment effect estimates for each strata (_STRATA_). The code is shown on the following page as follows:


```

PROC SURVEYLOGISTIC DATA = BRFSS;
  WEIGHT _LLCPWT;
  STRATA _STSTR;
  CLUSTER _PSU;
  DOMAIN _STRATA_;
  CLASS FLUSHOT(REF = "0") MARRIED(REF = "0") BLACK HISPANIC HSGRAD POVERTY
    SMOKER DOCTOR INSURANCE BADHEALTH / PARAM = GLM;
  MODEL FLUSHOT = MARRIED BLACK HISPANIC HSGRAD POVERTY SMOKER DOCTOR
    INSURANCE BADHEALTH;
RUN;

```

Strata	Odds Ratio	95% Confidence Limits		P-Value
1	1.187	1.134	1.243	< 0.0001
2	1.153	1.087	1.223	< 0.0001
3	1.204	1.131	1.282	< 0.0001
4	1.204	1.123	1.290	< 0.0001
5	1.149	1.076	1.228	< 0.0001

Table 7. Results for Propensity Score Stratification, Estimates by Strata

As the above table shows, the confidence intervals for the five strata overlap. Therefore, the treatment effect is homogenous across the propensity score quintiles (Roberts, 2010).

Next, the estimates are combined (DuGoff, 2014). The code and treatment effect estimate for the combined estimate is below:

```

PROC SURVEYLOGISTIC;
  WEIGHT _LLCPWT;
  STRATA _STSTR;
  CLUSTER _PSU;
  CLASS FLUSHOT(REF = "0") MARRIED(REF = "0") _STRATA_ BLACK HISPANIC HSGRAD
    POVERTY SMOKER DOCTOR INSURANCE BADHEALTH / PARAM = GLM;
  MODEL FLUSHOT = MARRIED _STRATA_ BLACK HISPANIC HSGRAD POVERTY SMOKER
    DOCTOR INSURANCE BADHEALTH;
RUN;

```

Odds Ratio	95% Confidence Limits		P-Value
1.165	1.134	1.196	< 0.0001

Table 8. Combined Results for Propensity Score Stratification

SUMMARY OF RESULTS

The summary of all the estimates of the illustrative example are as follows:

Analysis	Odds Ratio	95% Confidence Limits		P-Value
Unadjusted	1.379	1.347	1.413	< 0.0001
Regression Adjustment	1.179	1.149	1.211	< 0.0001
IPTW/ATE Weighting	1.179	1.146	1.209	< 0.0001
ATT Weighting	1.188	1.157	1.220	< 0.0001
Propensity Score Stratification	1.165	1.134	1.196	< 0.0001

Table 9. Summary of Results from Illustrative Example

As shown in the above table, all treatment effect estimates were close to one another regardless of the propensity score method used. All estimates illustrate significantly higher odds of married respondents receiving a flu shot as compared to unmarried respondents.

CONCLUSION

Failure to account for the design elements when working with complex survey data sets and propensity score methods together may lead to biased estimates and may only reflect the sample size instead of generalizing the results to the target population.

Applying propensity scores to complex survey data is still a methodological work in progress. As Austin (2018) highlights, further methodological work is needed in order to overcome some of the limitations of the current approaches.

REFERENCES

- Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, 46: 399-424.
- Austin, P. C., Jembere, N., Chiu, M. (2018). Propensity score matching and complex survey. *Statistical Methods in Medical Research*, 27(4): 1240-1257.
- DuGoff, E. H., Schuler, M., Stuart, E. A. (2014). Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys. *Health Services Research*, 49(1) 284-303.
- Lochner, K. A. & Wynne, M. (2011). Flu shots and the characteristics of unvaccinated elderly Medicare beneficiaries. *Medicare & Medicaid Research Review*, 1(4): E1-E11.
- Roberts, A. L, Gilman, S. E., Fitzmaurice, G., Decker, M. R., Koenen, K. C. (2010). Witness of Intimate Partner Violence in Childhood and Perpetration of Intimate Partner Violence in Adulthood. *Epidemiology*, 21(6): 809-818.
- US CDC (2017). Annual Survey Data. Available online: https://www.cdc.gov/brfss/annual_data/annual_2016.html.
- West, B. T., Sakshaug, J. W., Aurelien, G. A. S. (2016). How Big of a Problem is Analytic Error in Secondary Analyses of Survey Data? *PLOS ONE*. DOI: 10.1371/journal.pone.0158120.
- Yuan, Y., Yung, Y. F., Stokes, M. (2017). Propensity Score Methods for Causal Inference with the PSMATCH Procedure. Paper SAS332-2017. *SAS Global Forum 2017*; Orlando, Florida.
- Zanutto, E. L. (2006). A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data. *Journal of Data Science*, 4: 67-91.

RECOMMENDED READING

- *SAS/STAT® Procedures Guide*
- *Propensity Score Methods for Causal Inference with the PSMATCH Procedure*: <https://support.sas.com/resources/papers/proceedings17/SAS0332-2017.pdf>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Patrick Karabon, MS
Oakland University William Beaumont School of Medicine (OUWB)
pkarabon@oakland.edu
<https://www.oakland.edu/medicine/>