

Building a Propensity Score Model with SAS/STAT® Software: Planning and Practice

Michael Lamm, Clay Thompson, and Yiu-Fai Yung, SAS Institute Inc., Cary, NC

ABSTRACT

Propensity score matching is an intuitive approach that is often used in estimating causal effects from observational data. However, all claims about valid causal effect estimation require careful consideration, and thus many challenging questions can arise when you use propensity score matching in practice. How to select a propensity score model is one of the most difficult questions that you are likely to encounter when you do matching. The propensity score model should consider both the tenability of the assumption of no unmeasured confounding and the covariate balance in the matched data. This paper discusses how you can use the PSMATCH procedure in conjunction with other procedures in SAS/STAT® software to tackle some of these practical challenges. In particular, the paper demonstrates how you can use causal graphs to investigate questions related to ignorability and how you can incorporate propensity scores that are computed using approaches other than logistic regression. The paper also illustrates features of PROC PSMATCH that you can use to try to improve covariate balance and control properties of the final matched data set.

INTRODUCTION

Estimating a treatment's causal effect from observational data is a difficult but at times necessary task. Observational data are likely to contain confounding variables that introduce noncausal sources of association between the treatment and the outcome. These noncausal sources of association can bias estimates of a treatment's causal effect. Unlike the effect of sampling variability, the bias that confounding variables introduce cannot be corrected for by increasing the size of the observational data set. The primary challenge that you must address in estimating causal effects from observational data is how to account for the bias that confounding variables can create.

For binary treatments, matching is an intuitive approach that you can use to adjust for confounding variables. Matching methods aim to create from the input data set a new data set in which the two treatment conditions have comparable distributions of the confounders. You typically assess the comparability of the treatment conditions by examining measures of covariate balance. If you achieve sufficient balance, then you can estimate the treatment's causal effect from the matched data by using any number of familiar techniques.

Although matching is an intuitive approach to adjusting for confounders, its implementation requires that you address many difficult questions, starting with how to create the matched data set. A popular way to create a matched data set is to formulate the matching problem around a model for the probability of receiving treatment, given a set of pretreatment variables. This probability is referred to as the propensity score, and as discussed in the section “[Propensity Score Matching](#),” it has theoretical and computational properties that make it an appealing basis for matching. However, difficult questions remain, such as how to model the propensity scores and what constraints to use in the matching problem to help create a well-balanced data set. The examples in this paper illustrate tools in SAS/STAT software that you can use to address the challenges that arise when you perform an analysis based on propensity score matching.

The paper is organized as follows. First, the section “[BACKGROUND](#)” reviews the definition of causal effects in a potential outcome framework, important assumptions that you must consider in evaluating a causal effect estimate, and the role of the propensity score in a matching-based analysis. [Example 1](#) illustrates options in the PSMATCH procedure that you can use to modify the formulation of the matching problem. In particular, the example demonstrates the use of calipers, the use of support regions, and how you can provide precomputed propensity score values to PROC PSMATCH by using the PSDATA statement. [Example 2](#) illustrates the importance of carefully considering the assumptions that determine when an effect estimate has a valid causal interpretation. In particular, the example demonstrates how you can use the new CAUSALGRAPH procedure to analyze graphical causal models and identify sources of association that can inform your selection of a propensity score model.

BACKGROUND

This section reviews how causal effects are defined in a potential outcome framework, the assumptions that support the estimation of a causal effect from observational data, and the role of the propensity score in an analysis based on matching. Throughout this paper, it is assumed that you are interested in estimating the causal effect of a binary treatment decision T on an outcome Y , where $T = 1$ designates the active treatment condition and $T = 0$ designates the control condition.

Defining Causal Effects with Potential Outcomes

This paper primarily considers the definition of causal effects within the Neyman-Rubin potential outcome framework (Rubin 1980, 1990; Neyman, Dabrowska, and Speed 1990). In this framework, potential outcomes are defined for each subject at each possible level of the treatment T . For a binary treatment, each subject has two potential outcomes, denoted by $Y(1)$ and $Y(0)$. The potential outcome $Y(t)$ is interpreted as the outcome that would occur for the treatment assignment t that might be different from the observed treatment T . Potential outcomes are also sometimes referred to as counterfactual outcomes.

In the potential outcome framework, it is natural to define the individual-level treatment effect as the difference $Y(1) - Y(0)$. The individual-level treatment effect is then used to define the average treatment effect (ATE) and the average treatment effect for the treated (ATT). Formally, these two quantities are defined as

$$\begin{aligned} \text{ATE} &= E [Y(1) - Y(0)] \\ \text{ATT} &= E [Y(1) - Y(0) \mid T = 1] \end{aligned}$$

These definitions of a causal effect are appropriate when the potential outcomes capture all the relevant variability in the treatment, when the treatment's effect is only through the individual-level difference in potential outcomes, and when individuals' treatment decisions and outcomes are independent. These conditions are more formally stated in terms of the stable unit treatment value assumption (SUTVA) (Rubin 1980). The potential outcomes are connected to the observed outcome through the consistency assumption that $Y = Y(t)$ if $T = t$.

The estimation of causal effects is simplified when you are working with data from fully randomized experiments. The following intuitive description explains why this is the case. If an external mechanism is used to randomly assign subjects to treatment conditions, then it is expected that at the start of the study, the two treatment conditions should be comparable in all respects except for the treatment. Therefore, any difference that you observe in the outcomes at the end of the study can be attributed to the treatment with a causal interpretation. More formally, when complete randomization is used, the independence assumption $Y(t) \perp\!\!\!\perp T$, for $t = 0, 1$, is plausible. When this assumption holds, you can then easily estimate the potential outcome means from the observed data, because

$$E [Y \mid T = t] = E [Y(t) \mid T = t] = E [Y(t)]$$

In an observational study, or a study with imperfect randomization, estimating causal effects is more complicated because of the presence of confounding variables that are associated with both the treatment and the outcome. For example, a patient's age or health status is likely to affect the treatment decision made by his or her medical providers, and these same factors are also likely to affect many possible outcomes of interest. In observed medical records, a treatment's effectiveness would therefore be confounded by these and other, similar factors.

For observational data, the independence assumption is not likely to be plausible. A different assumption is required in order to ensure that there is sufficient information about the confounders to support the unbiased estimation of a treatment's causal effect. This assumption often takes the form of either weak or strong ignorability. Weak ignorability assumes that there exists a collection of variables X such that

$$Y(t) \perp\!\!\!\perp T \mid X \quad \text{for } t = 0, 1$$

The assumption of weak ignorability is implied by the stricter assumption of strong ignorability,

$$(Y(0), Y(1)) \perp\!\!\!\perp T \mid X$$

Often an assumption of ignorability is referred to as the assumption of no unmeasured confounding. A final assumption that is made when you analyze observational data is the positivity assumption, which requires there to be no values of

X for which the conditional probability of receiving treatment is either zero or one. Under these assumptions, it is possible to obtain unbiased estimates of a treatment’s causal effect by using methods that properly adjust for the confounders X . For more information about the definition of causal effects in a potential outcome framework, see Imbens and Rubin (2015), Hernán and Robins (2019), and references therein.

Graphical Causal Models

Assumptions about the process that generates a data set are key to evaluating whether a set of variables X satisfy the ignorability conditions necessary for the valid estimation of a causal effect. Graphical causal models provide a useful framework for representing assumptions about a data generating process. This section provides a brief overview of how you can represent these assumptions in a directed acyclic graph (DAG). For more information about causal graphs, see Spirtes, Glymour, and Scheines (2001), Pearl (2009a, b), Elwert (2013), and references therein.

A causal DAG consists of a set of nodes that represent variables in the causal model and a set of directed edges between variables. An edge from a variable T into a variable Y represents a possible direct causal effect of T on Y . The absence of an arrow from T to Y therefore represents the belief that T has no direct causal effect on Y . For observational data, the DAG is likely to include nodes that represent variables that influence the data generating process but are unmeasured. There is likely to be uncertainty about how variables, measured or unmeasured, might affect the data generating process. The specification of a causal model therefore depends in part on expert opinions, knowledge of the measurement process, and past experience.

After they are specified, the paths in a DAG can be analyzed to identify and classify sources of association in a causal model. Three of the basic structures in a DAG that correspond to the sources of association are described in Elwert (2013) and summarized in Table 1.

Table 1 Structures in a Directed Acyclic Graph

Path	Association between T and Y along the Path	Role of X in the Path	Effect of Conditioning on X
$T \rightarrow X_1 \rightarrow Y$	Causal	Mediator	Blocks the causal path
$T \leftarrow X_2 \rightarrow Y$	Noncausal	Confounding common cause	Blocks the noncausal path
$T \rightarrow X_3 \leftarrow Y$	None	Collider	Opens the noncausal path

Intuitively, adjusting for a set of variables leads to a valid causal effect estimate if the adjustment blocks all the noncausal paths between treatment T and outcome Y while leaving all causal paths unblocked. For example, in a simple causal model that consists only of the paths shown in Table 1, an adjustment set that contains X_1 would not be valid, because adjusting for X_1 blocks a causal path between T and Y . Similarly, an adjustment set that contains X_3 would not be valid, because adjusting for X_3 opens a noncausal path. The only valid adjustment set would consist of the variable X_2 , because adjusting for X_2 blocks the one noncausal path in the model. The theory of graphical causal models formally develops these concepts to create criteria for the construction of valid adjustment sets (Pearl 2009a; Shpitser, VanderWeele, and Robins 2010; Van der Zander, Liśkiewicz, and Textor 2014).

The analysis of adjustment sets through graphical causal models can be related to the potential outcome framework described in the section “Defining Causal Effects with Potential Outcomes.” In particular, Shpitser, VanderWeele, and Robins (2010) prove that a set of variables X satisfy the assumption of weak ignorability if and only if the variables X satisfy a certain graphical adjustment criterion. Example 2 demonstrates the importance of carefully considering the data generating process when you estimate causal effects and illustrates how this can be done by analyzing graphical causal models with the new CAUSALGRAPH procedure.

Propensity Score Matching

Matching attempts to adjust for confounding variables by creating a new data set in which the two treatment conditions are comparable in terms of the relevant confounders. The new matched data set is created by identifying subjects in the initial data set that have similar values of the confounders. If you can achieve comparable treatment conditions, you can estimate the treatment effect by using the matched data set.

The propensity score is the probability of receiving treatment conditional on a set of covariates X and is denoted by

$$e(x) = \Pr(T = 1 \mid X = x)$$

Matching subjects according to the similarity of their propensity score values is computationally appealing, because it uses a single scalar value as the primary comparison instead of making comparisons based directly on the whole vector of covariates X . The use of the propensity score in matching problems also has a justification based on the potential outcome framework. As shown in Rosenbaum and Rubin (1983), the propensity score is a balancing score. The balancing score property means that the components of X and the treatment T are conditionally independent, conditioned on the propensity score. Moreover, if the strong ignorability and positivity assumptions hold, the potential outcomes ($Y(0), Y(1)$) are also independent of the treatment, conditioned on the propensity score value. These properties imply that it is possible to estimate the treatment effect within the levels of the propensity scores, thus making the propensity score a natural basis for a matching problem.

In observational studies, the propensity score values are typically unknown and must be estimated. Assessing the balance of covariates after matching therefore provides important diagnostics for evaluating the quality of a matching and the propensity score model. An analysis that uses propensity score matching often compares the balance in different matched data sets because of uncertainty in how to best formulate either the matching problem or the propensity score model. The formulation of a matching problem might vary in terms of how it is solved or whether constraints are imposed on the problem. For information about different approaches to formulating a matching problem, see Stuart (2010), Austin (2014), Guo and Fraser (2015), Imbens and Rubin (2015), and references therein. To mitigate concerns about how the selection of a matched data set for the outcome analysis might unduly influence the causal effect estimation, you should exclude the outcome from all considerations when you fit the propensity score model, perform the matching, and assess the balance in the matched data set.

The PSMATCH procedure in SAS/STAT software provides a variety of tools for performing propensity score analysis (Yuan, Yung, and Stokes 2017). It enables you to create matched data sets and request numerical and graphical assessments of covariate balance. The examples in the remainder of the paper illustrate how you can use the PSMATCH procedure in conjunction with other procedures in SAS/STAT software to perform an analysis with propensity score matching. [Example 1](#) demonstrates how you can use the PSMATCH procedure to produce and compare matched data sets on the basis of different propensity score models or different matching formulations. [Example 2](#) demonstrates how analyzing graphical causal models with the new CAUSALGRAPH procedure can help inform the selection of a propensity score model and the importance of considering the data generating process in matching.

Example 1: Comparing Different Matches

This example illustrates an analysis in which propensity score matching is used to adjust for confounding before estimating the ATT. It demonstrates how you can compare covariate balance in different matched data sets before selecting the final analysis data set. The matched data sets are produced by using different propensity score models and by imposing different constraints on the matching problem. The data for this example are a subset of data from the NHANES I Epidemiologic Follow-Up Study (NHEFS) that is used in Hernán and Robins (2019). For the study, medical and behavioral information was collected during an initial physical examination and again at follow-up interviews approximately one decade later. The causal question of interest is the effect that quitting smoking has on an individual's change in weight. The initial data set **SmokingWeight** is also analyzed in Examples 1 and 2 in the documentation of PROC CAUSALTRT (SAS Institute Inc. 2018). The current example is not intended to provide a definitive analysis of this data set and causal question.

The variables in the data are as follows:

- **Activity**: level of daily activity, with values 0, 1, and 2
- **Age**: age in 1971
- **BaseWeight**: weight in kilograms in 1971
- **Change**: difference in weight (in kilograms) at the initial and follow-up interviews
- **Education**: level of education, with values 0, 1, 2, 3, and 4
- **Exercise**: amount of regular recreational exercise, with values 0, 1, and 2
- **PerDay**: number of cigarettes smoked per day in 1971
- **Quit**: 1 if an individual quit smoking between the initial and follow-up interviews; 0 otherwise

- **Race:** 0 for white; 1 otherwise
- **Sex:** 0 for male; 1 for female
- **Weight:** weight in kilograms at the follow-up interview
- **YearsSmoke:** number of years that an individual has smoked

Change is the outcome of interest for this analysis, and **Quit** is the treatment variable.

For the matching process, a new data set, **SmokingNoResp**, is created from **SmokingWeight** by dropping the variables **Change** and **Weight**. Excluding these variables ensures that the process of selecting the analysis data set is not influenced by knowledge of the outcome variable. Before creating the new data set, a variable **ID** is added to the **SmokingWeight** data set so the outcome can be easily added to the final analysis data set. To simplify this example, it is also assumed that the missingness in the data is ignorable and that all observations with missing values are removed from the input data set. These tasks are accomplished using the following DATA steps:

```
data SmokingWeight;
  set SmokingWeight;
  if nmiss(of _numeric_) > 0 then delete;
  id = _N_;
run;

data SmokingNoResp;
  set SmokingWeight;
  drop Change Weight;
run;
```

The following statements invoke the PSMATCH procedure to create a matched data set and assess the covariate balance after matching:

```
ods graphics on;
proc psmatch data=SmokingNoResp;
  class Sex Race Education Exercise Activity Quit ;
  psmodel Quit(Treated='1') = Sex Age Education Exercise Activity YearsSmoke PerDay;
  match method=varratio(kmin=1 kmax=4) distance=lps caliper=0.5;
  assess lps var=(age YearsSmoke BaseWeight PerDay) /plots=(CDFPlot BoxPlot StdDiff);
  output out(obs=all)= SmokeMatched1 weight=matchattwgt matchid=MID;
run;
```

The propensity score model for this matching is specified in the PSMODEL statement and is fit by logistic regression. You can use the TREATED= option to specify what level of the treatment variable is modeled. The MATCH statement requests the use of matching and specifies the matching problem to solve. The METHOD=VARRATIO option requests the use of an optimal variable ratio matching. A variable ratio matching allows the ratio of treated to control units in each matched set to vary. You can use the KMIN= and KMAX= options to specify a range of how many subjects in the control condition are matched to a subject in the treatment condition. The matching is optimal in the sense that it minimizes the total absolute difference of the matching distance across all feasible matches. The DISTANCE=LPS option specifies that the logit of the propensity scores is used as the matching distance. The CALIPER=0.5 option specifies the caliper constraint for the matching. The caliper constraint requires that matched subjects have an absolute difference of the matching distance less than the caliper value. By default, the caliper value is determined by multiplying the value specified in the CALIPER= option by an estimate of the standard deviation of the matching distance that is pooled across the treated and control groups. A caliper value provides one tool for enforcing similarity between subjects in the matched data set. For more information about using a caliper in matching, see Austin (2011).

You can use the ASSESS statement in PROC PSMATCH to request numerical and graphical summaries of covariate balance. Balance diagnostics are produced for any continuous or binary variable listed in the VAR= option, and the LPS option requests comparisons for the logit of the propensity scores. The output data set **SmokeMatched1** is created by the OUTPUT statement. The WEIGHT= keyword names a variable that contains the ATT weights used in the balance assessments. You should also incorporate these weights into any outcome analysis. The ATT weights after matching without replacement are one for each matched treated unit, and for a matched control unit the weight is equal to the number of treated units in its matched set divided by the number of control units in its matched set. For unmatched subjects, the ATT weights are set to missing. The WEIGHT= keyword, which is new in SAS/STAT 15.1, replaces four keywords that are used in previous releases.

Figure 1 displays the “Data Information” table. This table provides information about the input data set, any output data that are created, and some properties of the data that are matched. In particular, this table shows that by default the “Extended Treated Group” support region was used for this matching problem. Observations whose propensity score values fall outside the support region are not matched. The extended treated group support region is determined by the range of propensity score values of subjects in the treatment condition. By definition, this support region contains all treated subjects, but it might include only a subset of the subjects in the control condition. In this example, eight subjects from the control condition have propensity score values outside the support region and are not considered for matching.

Figure 1 Data Information
The PSMATCH Procedure

Data Information	
Data Set	WORK.SMOKINGNORESP
Output Data Set	WORK.SMOKEMATCHED1
Treatment Variable	Quit
Treated Group	1
All Obs (Treated)	403
All Obs (Control)	1163
Support Region	Extended Treated Group
Lower PS Support	0.086181
Upper PS Support	0.736673
Support Region Obs (Treated)	403
Support Region Obs (Control)	1155

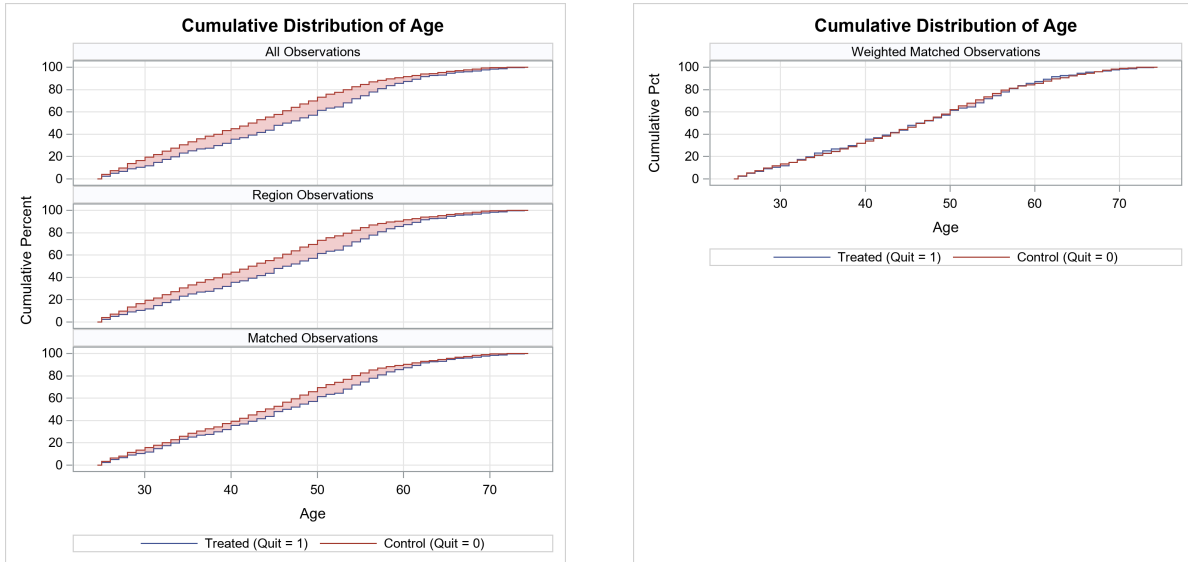
The “Matching Information” table (Figure 2) provides a summary of the matching distance used, the method used to create the matched data set, the caliper value, and a summary of the matched sets.

Figure 2 Matching Information

Matching Information	
Distance Metric	Logit of Propensity Score
Method	Optimal Variable Ratio Matching
Min Control/Treated Ratio	1
Max Control/Treated Ratio	4
Caliper (Logit PS)	0.25032
Matched Sets	403
Matched Obs (Treated)	403
Matched Obs (Control)	1008
Total Absolute Difference	4.78415

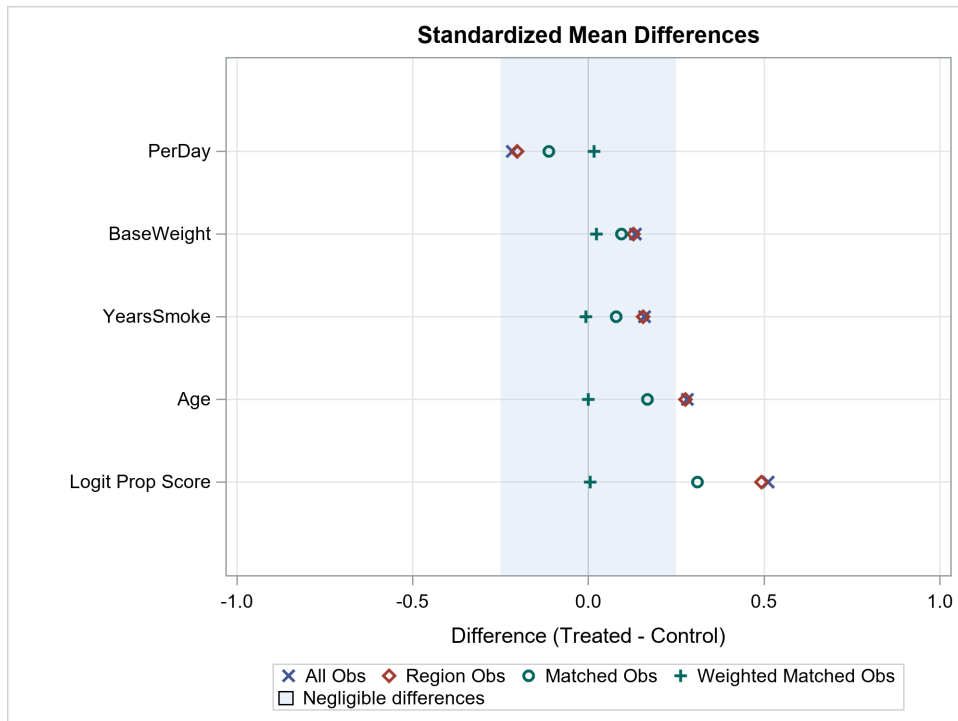
The CDFPlot request in the PLOTS= option produces cumulative distribution function (CDF) plots for continuous variables. The plots compare the CDF of a variable among the different treatment conditions in the entire data set, when restricted to the support region, for the matched data, and for the weighted matched data. Figure 3 displays CDF plots for **Age**. These plots show that in the input data, subjects who quit smoking tended to be older than those who did not quit. The balance for **Age** improves after you incorporate the ATT matching weights.

Figure 3 CDF Plots for Age



The “Standardized Mean Differences” plot (Figure 4) displays the difference in the means of variables between the treatment and control groups. By default, the differences are standardized by dividing by the sample standard deviations of the variables, pooled across the treated and control groups, computed from all the observations. This plot provides a compact assessment of the balance that is achieved across a number of variables. The shaded range of -0.25 to 0.25 is one common recommendation for what constitutes good balance; see Rubin (2001) and Stuart (2010). Other authors (Normand et al. 2001; Mamdani et al. 2005; Austin 2009) recommend a smaller range of -0.1 to 0.1 . You can modify what range is shaded in the “Standardized Mean Differences” plot by using the REF= option in the STDDIFF plot request.

Figure 4 Standardized Mean Differences



The “Standardized Mean Differences” table (Figure 5) displays the values of the standardized mean differences and the treated-to-control variance ratios. The standardized mean differences all show large percentage reductions from the input data to the weighted matched observations. The variance ratios also tend to show improvement with values closer to one after weighting. One exception is a small increase in the variance ratio for **PerDay** after weighting. Although this increase does not disqualify the use of these matched data in the outcome analysis, you might consider examining the balance that is achieved in other matched data sets.

Figure 5 Standardized Mean Differences and Variance Ratios

The PSMATCH Procedure

Standardized Mean Differences (Treated - Control)						
Variable	Observations	Mean Difference	Standard Deviation	Standardized Difference	Percent Reduction	Variance Ratio
Logit Prop Score	All	0.25574	0.50064	0.51083		0.9572
	Region	0.24628		0.49193	3.70	1.0026
	Matched	0.15556		0.31073	39.17	1.2165
	Weighted Matched	0.00276		0.00551	98.92	1.0206
Age	All	3.38522	12.00514	0.28198		1.0731
	Region	3.31742		0.27633	2.00	1.0727
	Matched	2.01199		0.16759	40.57	1.0853
	Weighted Matched	-0.00744		-0.00062	99.78	1.0085
YearsSmoke	All	1.94455	12.23620	0.15892		1.1846
	Region	1.90845		0.15597	1.86	1.1807
	Matched	0.95885		0.07836	50.69	1.1568
	Weighted Matched	-0.08561		-0.00700	95.60	1.0896
BaseWeight	All	2.05205	15.40392	0.13322		1.0606
	Region	1.96795		0.12776	4.10	1.0603
	Matched	1.44707		0.09394	29.48	1.0990
	Weighted Matched	0.34874		0.02264	83.01	1.0967
PerDay	All	-2.58877	11.94772	-0.21667		1.1679
	Region	-2.41953		-0.20251	6.54	1.2013
	Matched	-1.35139		-0.11311	47.80	1.3452
	Weighted Matched	0.19293		0.01615	92.55	1.3787
Standard deviation of All observations used to compute standardized differences						

You can create a new matched data set either by trying a different type of matching or by changing how the propensity score model is fit. One way of changing the propensity score model is to replace the continuous effects in the model with spline effects. The PSMATCH procedure does not support the construction of spline effects, but it does enable you to specify a variable in the input data set that contains precomputed propensity scores or logits of the propensity scores. The following statements invoke the LOGISTIC procedure to fit a propensity score model that includes spline effects:

```
proc logistic data=SmokingNoResp ;
  class Sex Race Education Exercise Activity Quit ;
  Effect splAge = spline(age/basis=bspline);
  Effect splPerDay = spline(PerDay /basis=bspline);
  Effect splYrsSmoke = spline(YearsSmoke/ basis=bspline);
  Effect splBWeight = spline(BaseWeight/basis=bspline);
  model Quit(Event='1') = Sex splAge Education Exercise Activity splYrsSmoke splPerDay ;
  output out=SmokeSPLPred predicted=pScore;
run;
```


The EFFECT statement is used to construct the spline effects that are then specified in the MODEL statement. The OUTPUT statement creates the data set **SmokeSPLPred**, and the PREDICTED= option names the variable that contains the predicted propensity score values. Note that after these statements are executed, the warnings shown in [Figure 6](#) appear in the SAS log.

Figure 6 Quasi-complete Separation Warnings

```
WARNING: There is possibly a quasi-complete separation of data points. The
          maximum likelihood estimate may not exist.
WARNING: The LOGISTIC procedure continues in spite of the above warning.
          Results shown are based on the last maximum likelihood iteration.
          Validity of the model fit is questionable.
```

Quasi-complete separation describes a data configuration for which the maximum likelihood estimates are not unique and might be made infinitely large. The LOGISTIC procedure prints the warnings displayed in [Figure 6](#) when there is an observation that has a very large predicted probability. Based on these warnings, you would expect there to be problems when you are trying to match on the propensity scores that are predicted by this model. Despite these concerns, this example considers matching on the propensity scores that are predicted by this model in order to illustrate features of the PSMATCH procedure.

When matching on precomputed propensity score values, you specify the PSDATA statement instead of the PSMODEL statement. You identify the treatment variable by using the TREATVAR= option in the PSDATA statement, and you can use the PS= option to identify the variable in the input data set that contains the propensity scores. The following statements invoke the PSMATCH procedure to perform variable ratio matching with the propensity scores that PROC LOGISTIC predicts:

```
proc psmatch data=smokeSPLPred;
  class Sex Race Education Exercise Activity Quit ;
  psdata treatvar=Quit(Treated='1') ps=pScore;
  match method=varratio(kmin=1 kmax=4) stat=lps caliper=0.5;
  assess lps var=(age YearsSmoke BaseWeight PerDay) /plots=(CDFPlot BoxPlot StdDiff);
  output out(obs=all)= smokeSPLMatched1 weight=matchattwgt matchid=MID;
run;
```

For this problem, no feasible matching exists, and the error shown in [Figure 7](#) is printed in the SAS log.

Figure 7 No Feasible Matching Error

```
ERROR: A feasible variable ratio matching that has the specified parameters
       does not exist.
```

The large values that are seen in the box-and-whisker plots for the logit of the propensity score values ([Figure 8](#)) are due to the quasi-complete separation of the data for this propensity score model. Matching within the extended treated group support region is not possible for the caliper value that is displayed in the “Matching Information” table ([Figure 9](#)). To illustrate the effects that removing the caliper constraint or changing the support region can have on a matching problem, this example will continue to match on the propensity score values that PROC LOGISTIC predicts despite the quasi-complete separation.

Figure 8 Box Plots for LPS

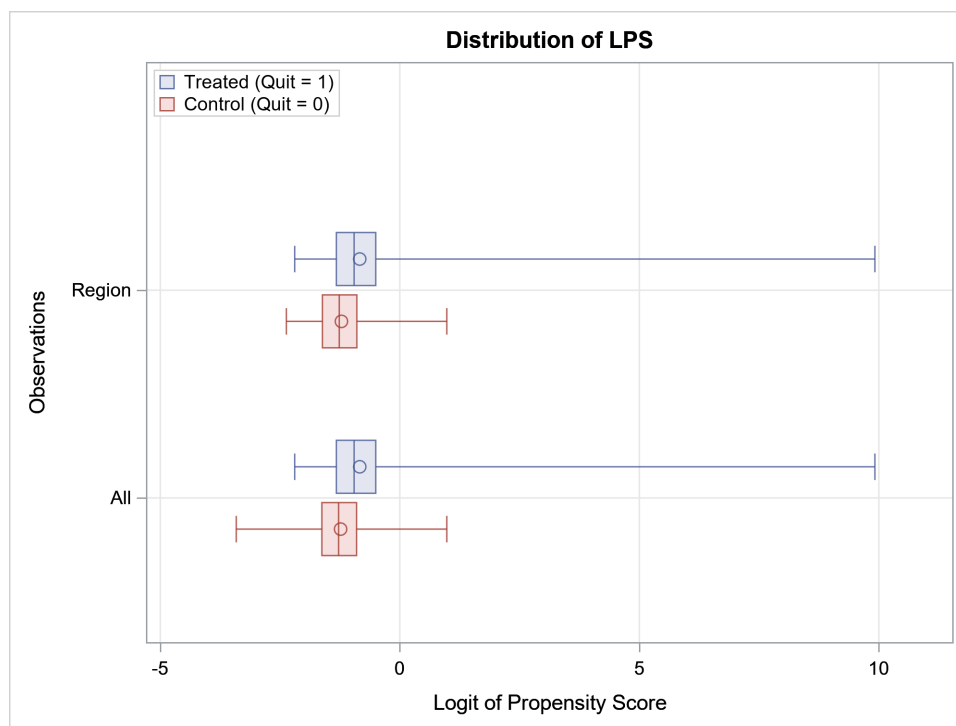


Figure 9 Matching Information

The PSMATCH Procedure

Matching Information	
Distance Metric	Logit of Propensity Score
Method	Optimal Variable Ratio Matching
Min Control/Treated Ratio	1
Max Control/Treated Ratio	4
Caliper (Logit PS)	0.362031

By default, CALIPER=0.25 is used for matching, so to remove the caliper constraint from a matching problem you must specify a missing value in the CALIPER= option. The following statements invoke the PSMATCH procedure to match without a caliper constraint:

```
proc psmatch data=smokeSPLPred;
  class Sex Race Education Exercise Activity Quit ;
  psdata treatvar=Quit(Treated='1') ps=pScore;
  match method=varratio(kmin=1 kmax=4) stat=lps caliper=.;
  assess lps var=(age YearsSmoke BaseWeight PerDay) /plots(nodetails)=(CDFPlot BoxPlot StdDiff);
  output out(obs=all)= smokeSPLMatched2 weight=matchattwtg matchid=MID;
run;
```

In the PLOTS requests, the NODetails option is used to limit the display of the balance assessment plots. When you specify the NODetails option, plots are displayed for only two sets of observations: the set of all observations and the weighted matched observations. The CDF plots for **PerDay** (Figure 10) show improved balance for this matching, and the variance ratio for **PerDay** is closer to one in the “Standardized Mean Differences” table (Figure 11). However, for this matching the percentage reductions in the standardized mean differences are lower than those achieved in the first matching (Figure 5) for other variables likely to be relevant to this causal question. In particular, the percentage reduction for **BaseWeight** is much lower for this matching.

Figure 10 CDF Plots for PerDay

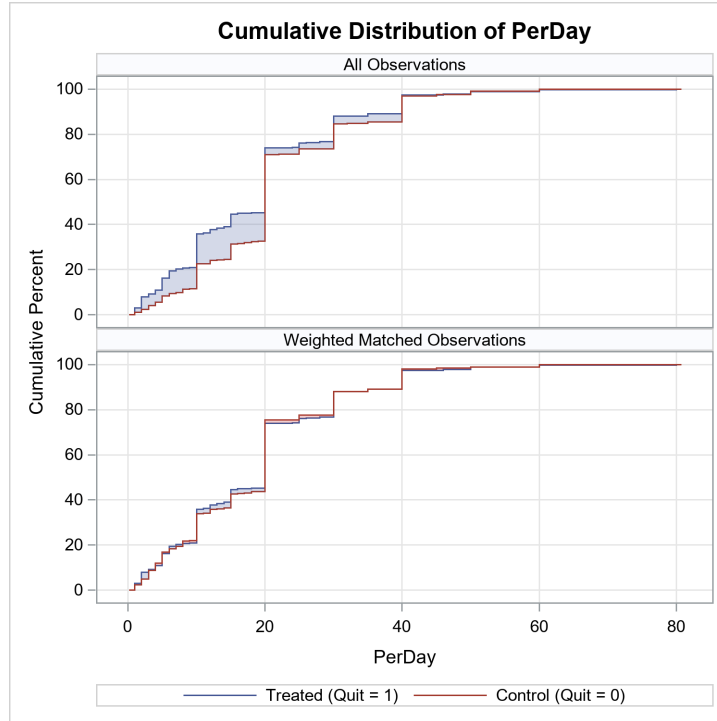


Figure 11 Standardized Mean Differences and Variance Ratios

The PSMATCH Procedure

Standardized Mean Differences (Treated - Control)						
Variable	Observations	Mean Difference	Standard Deviation	Standardized Difference	Percent Reduction	Variance Ratio
Logit Prop Score	All	0.39930	0.72406	0.55147		2.2937
	Region	0.38177		0.52727	4.39	2.4519
	Matched	0.28672		0.39599	28.19	2.8345
	Weighted Matched	0.06345		0.08763	84.11	2.1335
Age	All	3.38522	12.00514	0.28198		1.0731
	Region	3.19025		0.26574	5.76	1.0826
	Matched	2.15485		0.17949	36.35	1.0920
	Weighted Matched	0.16646		0.01387	95.08	1.0168
YearsSmoke	All	1.94455	12.23620	0.15892		1.1846
	Region	1.80142		0.14722	7.36	1.1863
	Matched	1.10071		0.08996	43.40	1.1520
	Weighted Matched	0.26613		0.02175	86.31	1.0356
BaseWeight	All	2.05205	15.40392	0.13322		1.0606
	Region	1.92784		0.12515	6.05	1.0570
	Matched	1.42514		0.09252	30.55	1.0770
	Weighted Matched	1.05227		0.06831	48.72	1.0866
PerDay	All	-2.58877	11.94772	-0.21667		1.1679
	Region	-2.49807		-0.20908	3.50	1.1810
	Matched	-1.65992		-0.13893	35.88	1.2325
	Weighted Matched	0.03205		0.00268	98.76	1.0890

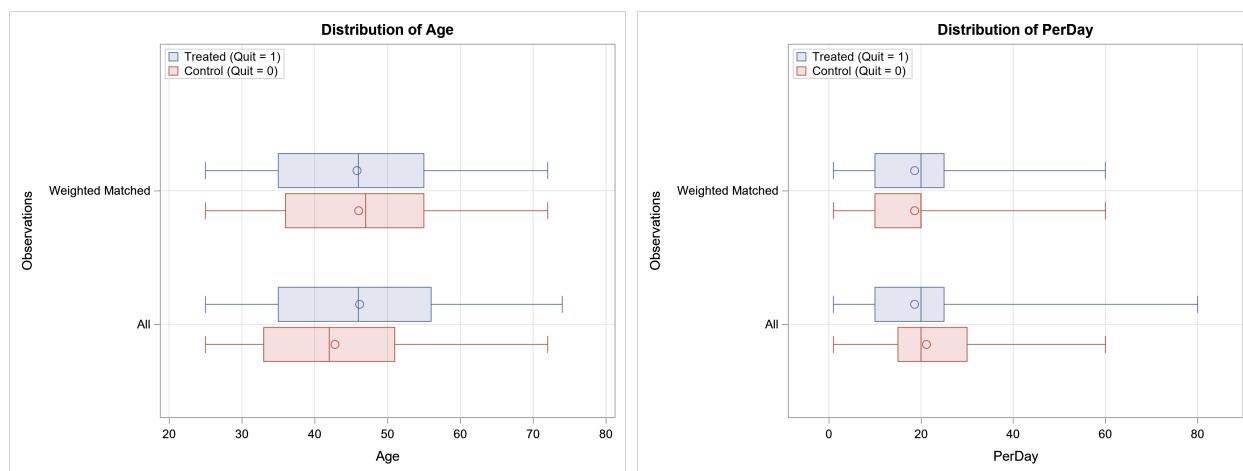
Standard deviation of All observations used to compute standardized differences

Rather than removing the caliper constraint, another way to modify the matching problem is to use the common support region instead of the extended treated group region as the support region. The common support region is the largest interval that contains overlapping propensity score values for subjects from both treatment conditions. The common support region contains subjects with comparable propensity score values and might exclude subjects from both treatment conditions. The following statements use the REGION=CS option in the PROC PSMATCH statement to request matching on the common support:

```
proc psmatch data=smokeSPLPred region=cs;
  class Sex Race Education Exercise Activity Quit ;
  psdata treatvar=Quit(Treated='1') ps=pScore;
  match method=varratio(kmin=1 kmax=4) stat=lps caliper=0.5;
  assess lps var=(age YearsSmoke BaseWeight PerDay) /plots(nodetails)=(CDFPlot BoxPlot StdDiff);
  output out (obs=all)= smokeSPLMatched3 weight=matchattwgt matchid=MID;
run;
```

A concern that arises when you are matching on the common support is that the subjects whose propensity score values are within the support region might not be representative of the target population. When you are estimating the ATT, this problem occurs when the characteristics of the treated subjects differ between the input and matched data sets. Figure 12 shows that the treated subjects in the matched data differ from the treated subjects in the input data set in their range of values for the variables **Age** and **PerDay**. Note that whiskers in the box-and-whisker plots that PROC PSMATCH produces always cover a variable's range and no observations are ever plotted as outliers.

Figure 12 Box Plots for Age and PerDay



Printing the observations that correspond to the unmatched treated subjects (Figure 13) confirms that this matching has excluded the oldest treated subjects and the treated subject who smoked the most cigarettes per day in the baseline year. These subjects were not matched because their propensity score values were outside the common support region. Based on this information, you might consider whether it is appropriate to redefine the target population for this analysis and remove from the input data set subjects who do not fit the revised conditions. In some instances, redefining the target population and removing subjects from consideration might be necessary in order to obtain a reliable estimate for a well-defined causal effect. For this example, the target population is not revised, and this matched data set is therefore not appropriate for use in the outcome analysis.

Figure 13 Unmatched Treated Subjects

Obs	id	Age	PerDay	Quit	YearsSmoke	matchattwgt
1037	1155	72	5	1	58	.
1149	1283	72	4	1	60	.
1300	1454	74	5	1	44	.
1504	1677	74	10	1	53	.
1522	1698	56	80	1	38	.
1544	1722	72	4	1	7	.

Using the propensity scores that PROC LOGISTIC predicts has not led to an improvement in the balance achieved in the first matched data. Given the quasi-complete separation of the data for the propensity score model that was fit using PROC LOGISTIC, an additional propensity score model will be considered before proceeding to the outcome analysis.

You can use the ADAPTIVEREG procedure in SAS/STAT software to fit multivariate adaptive regression splines as defined by Friedman (1991). PROC ADAPTIVEREG fits nonparametric models that can capture nonlinear trends and interactions between effects. This procedure also incorporates a model selection process by first creating an overfitted model that is then pruned by using a backward selection technique. The following statements invoke the ADAPTIVEREG procedure to predict the propensity score values and then match on these values by using the PSMATCH procedure:

```
proc adaptivereg data=smokingNoResp;
  class Sex Race Education Exercise Activity Quit;
  model Quit(Event='1') = Sex Age Education Exercise Activity YearsSmoke PerDay /dist=binary;
  output out=adpRegPred pred(ilink)=pScore;
run;

proc psmatch data=adpRegPred ;
  class Sex Race Education Exercise Activity Quit ;
  psdata treatvar=Quit(Treated='1') ps=pScore;
  match method=varratio(kmin=1 kmax=4) stat=lps caliper=0.5;
  assess lps var=(age YearsSmoke BaseWeight PerDay) /plots=(CDFPlot BoxPlot StdDiff);
  output out (obs=all)=SmokeADPred weight=matchattwgt matchID=mid;
run;
```

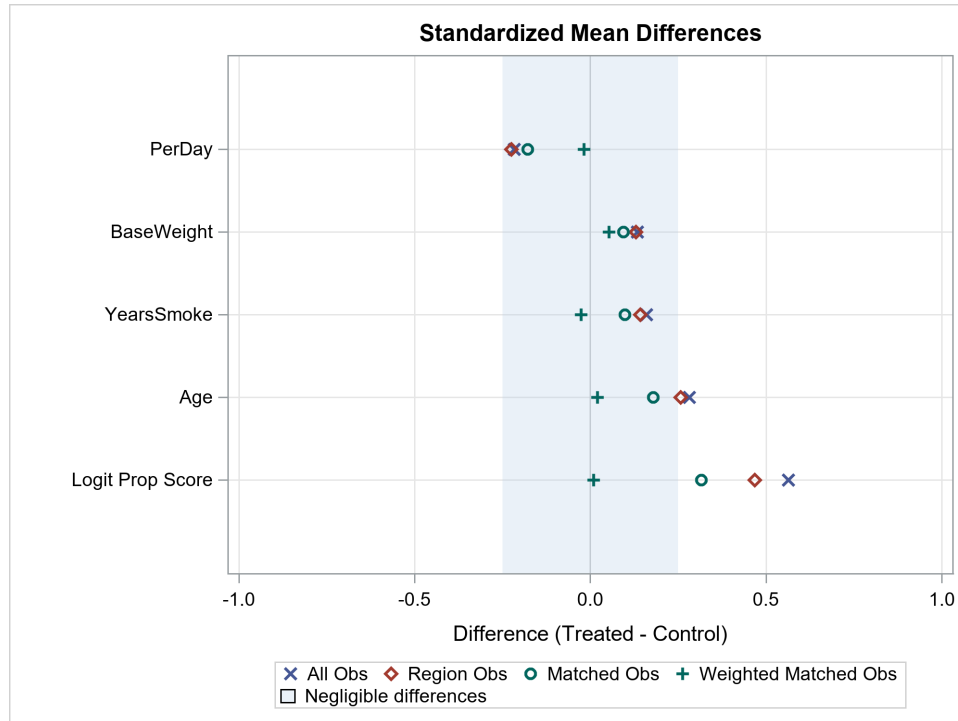
The regression spline model that PROC ADAPTIVEREG selects is displayed in Figure 14. The data set that is created by matching on the predicted propensity score values shows good balance for the continuous variables analyzed in the “Standardized Mean Differences” plot (Figure 15).

Figure 14 PROC ADAPTIVEREG Model Selection Summary

The ADAPTIVEREG Procedure

Regression Spline Model after Backward Selection				
Name	Coefficient	Parent	Variable	Knot Levels
Basis0	-1.3481		Intercept	
Basis2	0.05489	Basis0	PerDay	20.0000
Basis4	-0.07481	Basis0	Age	43.0000
Basis5	0.5633	Basis0	Sex	0
Basis7	-0.1425	Basis0	YearsSmoke	23.0000
Basis8	0.04826	Basis0	YearsSmoke	23.0000
Basis11	-0.1699	Basis4	Education	0
Basis13	-0.00876	Basis7	Age	57.0000
Basis19	0.007521	Basis7	Age	36.0000

Figure 15 Standardized Mean Differences after Matching on Propensity Scores from PROC ADAPTIVEREG



It is also important to investigate the balance for categorical variables. To do so, you can use the FREQ procedure to request contingency tables of the treatment and categorical variables of interest. When you request these tables, it is important to specify the matching weights in the WEIGHT statement. The following statements invoke the FREQ procedure to request contingency tables for the first matched data set, **SmokeMatched1**, and the data set **SmokeADPred**, which is obtained by matching on the propensity score values that PROC ADAPTIVEREG predicts:

```
proc freq data=SmokeADPred;
  table Quit*Exercise Quit*Activity / nocol nopercnt;
  weight matchattwgt;
run;

proc freq data=SmokeMatched1;
  table Quit*Exercise Quit*Activity / nocol nopercnt;
  weight matchattwgt;
run;
```

If you compare the weighted frequencies for **Exercise** and **Activity** between treatment conditions, you see better balance in the **SmokeMatched1** data set (Figure 16) than in the **SmokeADPred** data set (Figure 17). For this reason, **SmokeMatched1** is selected as the analysis data set.

Figure 16 Contingency Tables after Matching on Propensity Scores Predicted by PROC PSMATCH

Frequency Row Pct	Table of Quit by Exercise					Frequency Row Pct	Table of Quit by Activity				
	Quit	Exercise			Total		Quit	Activity			Total
0	62.75	176.583	163.667		403	0	168.333	189.917	44.75	403	
	15.57	43.82	40.61				41.77	47.13	11.10		
1	63	176	164		403	1	170	188	45	403	
	15.63	43.67	40.69				42.18	46.65	11.17		
Total	125.75	352.583	327.667		806	Total	338.333	377.917	89.75	806	

Figure 17 Contingency Tables after Matching on Propensity Scores Predicted by PROC ADAPTIVEREG

Frequency Row Pct	Table of Quit by Exercise					Frequency Row Pct	Table of Quit by Activity				
	Quit	Exercise			Total		Quit	Activity			Total
0	84.5833	162.25	156.167		403	0	175.583	190.667	36.75	403	
	20.99	40.26	38.75				43.57	47.31	9.12		
1	63	176	164		403	1	170	188	45	403	
	15.63	43.67	40.69				42.18	46.65	11.17		
Total	147.583	338.25	320.167		806	Total	345.583	378.667	81.75	806	

Because the confounders are well balanced in the **SmokeMatched1** data set, the outcome analysis can proceed using familiar estimation methods. After the data set **SmokeAnalysis** is created by merging **SmokeMatched1** and **SmokingWeight** by ID, the following statements invoke the TTEST procedure to perform the outcome analysis:

```
data SmokeAnalysis;
  merge SmokeMatched1 SmokingWeight;
  by id;
run;

proc ttest data=SmokeAnalysis;
  class Quit;
  var Change;
  weight matchattwgt;
run;
```

Figure 18 shows an ATT estimate of a 3.2792 kg increase in weight. This estimate is larger than the difference of 2.54 kg observed in the input data set **SmokingWeight** without adjusting for the confounders. Note that if you perform the outcome analysis with the data set **SmokeMatched3** that did not match the oldest treated subjects, the weighted difference in **Change** between the treatment conditions would increase to 3.5408 kg. You might expect this increase in the estimate if you believe that older individuals are less likely to increase in weight over the 10-year period regardless of whether they quit smoking. The difference in the estimates that is caused by excluding the oldest individuals demonstrates the importance of carefully defining the target population for an analysis.

Figure 18 Outcome Analysis Using **SmokeMatched1**

The TTEST Procedure

Variable: **Change**

Quit	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
0		1.2458	0.7992 1.6925	4.5692	4.3781 4.7779
1		4.5251	3.6684 5.3818	8.7483	8.1831 9.3979
Diff (1-2)	Pooled	-3.2792	-4.1171 -2.4414	6.0627	5.8469 6.2951
Diff (1-2)	Satterthwaite	-3.2792	-4.2447 -2.3138		

Example 2: Causal Diagrams and Model Considerations

This example demonstrates the importance of considering properties of the data generating process when you are selecting a propensity score model. The example uses a simulated data set and considers the causal question of what effect the location of postsurgical physical therapy has on a patient's recovery after coronary artery bypass grafting (CABG) surgery. The example uses a simplified causal model to illustrate important concepts in the analysis of observational data and is not meant to be a definitive discussion of this topic.

The observed variables in the simulated data are as follows:

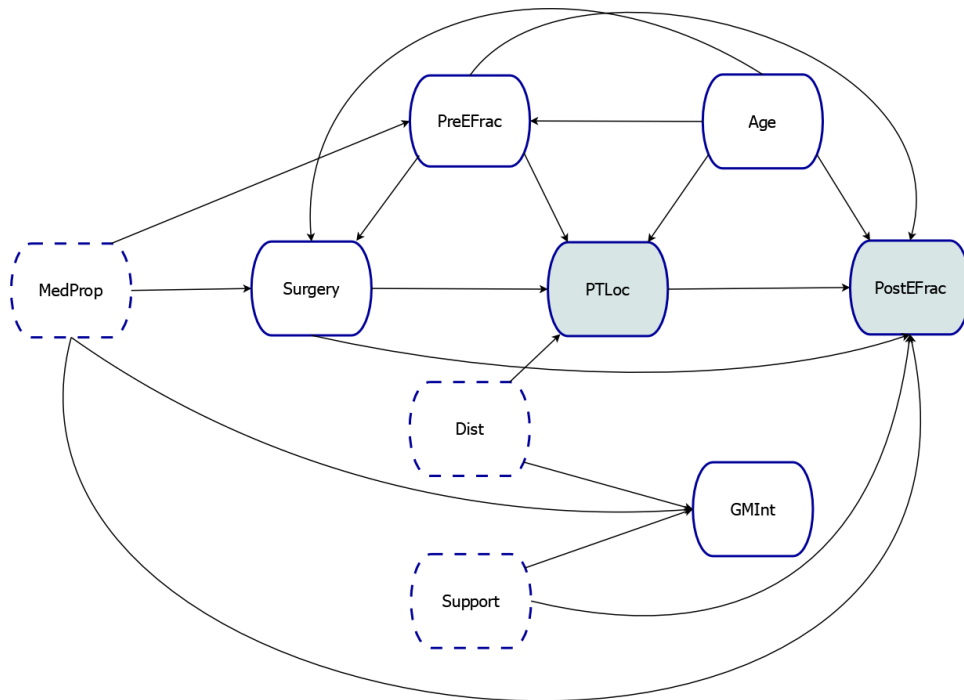
- **Age**: age of the patient
- **GMInt**: indicator of whether the patient expressed interest in attending postsurgical support group meetings at the medical facility
- **PostEFrac**: difference in ejection fraction of the heart before and three months after surgery
- **PreEFrac**: ejection fraction of the heart before surgery
- **PTLoc**: indicator of whether a patient after discharge attends physical therapy at the medical facility or off premises
- **Surgery**: indicator of whether patient received traditional CABG or off-pump CABG

PostEFrac is the outcome variable for the analysis, **PTLoc** is the treatment variable, and the four other observed variables can be considered pretreatment variables. It is assumed that the patients are men for whom no complications arose during surgery. The following three additional variables are used in the data generating process but are excluded from the simulated data:

- **Dist**: distance of the patient's home from the medical facility
- **MedProp**: a latent construct that reflects the patient's receptiveness to medical care
- **Support**: a measure of the regular support that the patient receives from his family and community

The simulation is designed so that the ATT of attending physical therapy off premises, **PTLoc** = 0, is -0.5 . The data generating process for the simulation is represented by the causal diagram shown in Figure 19. The treatment and outcome variables are represented as shaded nodes, and unmeasured variables are represented as nodes with dashed outlines.

Figure 19 Causal Model of Simulated Data



The simulated data set **SimDataN1k** contains 1,000 observations. The unadjusted difference in outcome means that is observed between the treatment conditions in **SimDataN1k** is -0.0890 (Figure 20).

Figure 20 ATT Estimate with No Adjustment, $N=1,000$

The TTEST Procedure

Variable: PostEFrac

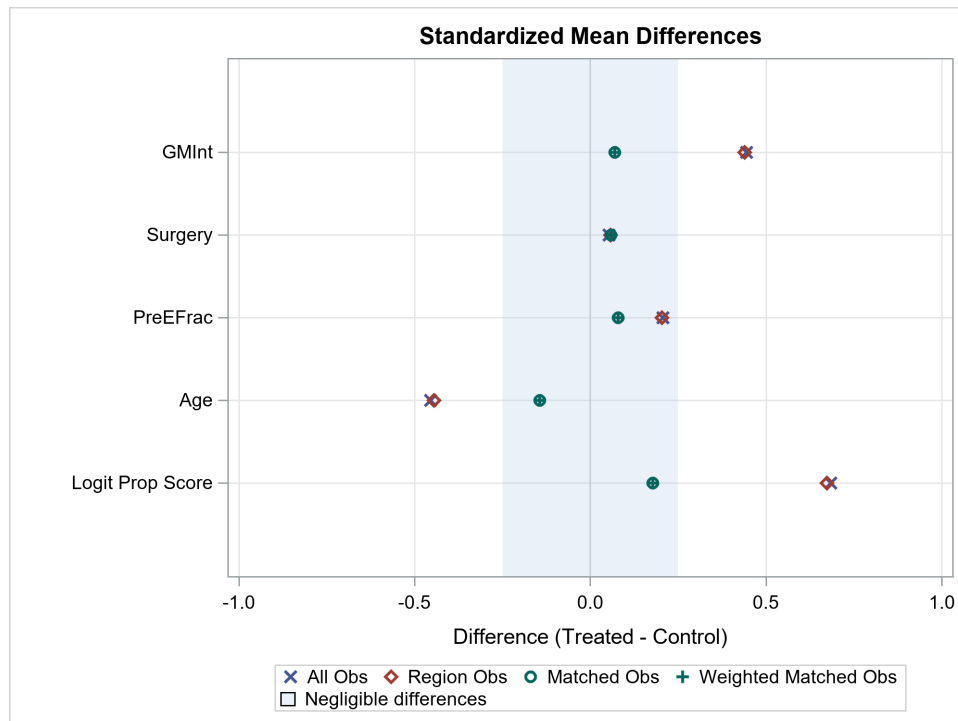
PTLoc	Method	Mean	95%				
			95% CL Mean	Std Dev	CL Std Dev		
0		16.0985	15.7955	16.4015	3.0277	2.8281	3.2578
1		16.1876	15.9487	16.4265	3.0146	2.8549	3.1934
Diff (1-2) Pooled		-0.0890	-0.4740	0.2959	3.0196	2.8928	3.1582
Diff (1-2) Satterthwaite		-0.0890	-0.4744	0.2963			

The following statements use the PSMATCH procedure to perform an optimal one-to-one matching that ignores any consideration of the data generating process and includes all the observed pretreatment variables in the propensity score model:

```
proc psmatch data=SimDataN1k;
  class Surgery PTLoc GMInt;
  psmodel PTLoc(Treated='0') = Surgery Age PreEFrac GMInt;
  match method=optimal stat=lps caliper=0.5;
  assess lps var=(Age Surgery GMInt PreEFrac) / plots=StdDiff;
  output out(obs=match)=Matched1N1k weight=attwgt;
run;
```

The “Standardized Mean Differences” plot (Figure 21) shows that relatively good balance is achieved in the matched data for all four variables and for the logit of the propensity scores.

Figure 21 Balance after Matching on All Variables, $N=1,000$



However, the ATT estimate of -0.0561 that is obtained using the TTEST procedure with the matched data, as shown by the following statements (Figure 22), is quite poor and further from the true value of -0.5 than the unadjusted estimate from the input data set (Figure 20).

```
proc ttest data=Matched1N1k;
  class PTLoc;
  var PostEFrac;
  weight attwgt;
run;
```

Figure 22 ATT Estimate after Matching on All Variables, $N=1,000$

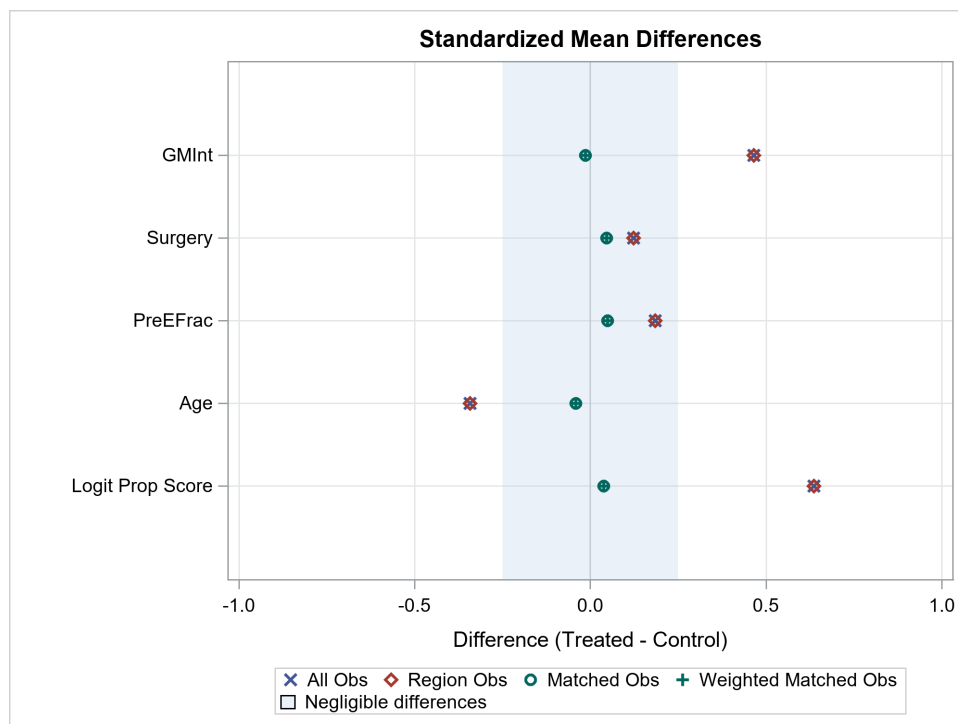
The TTEST Procedure

Variable: PostEFrac

PTLoc	Method	Mean	95%				
			95% CL Mean	Std Dev	CL Std Dev	Std Dev	
0		16.0985	15.7955	16.4015	3.0277	2.8281	3.2578
1		16.1546	15.8604	16.4489	2.9402	2.7464	3.1636
Diff (1-2) Pooled		-0.0561	-0.4778	0.3656	2.9842	2.8424	3.1412
Diff (1-2) Satterthwaite		-0.0561	-0.4778	0.3656			

The poor performance of the estimate after matching on all the pretreatment variables is not due to sampling error. Performing the same analysis with the **SimDataN10K** data set, which contains 10,000 observations, produces the same problems. After you create the data set **Matched1N10k** by matching **SimDataN10K** on all pretreatment variables, good balance is achieved, as shown in Figure 23.

Figure 23 Balance after Matching on All Variables, $N=10,000$



However, once again the ATT estimate of -0.2101 that is obtained from the matched data set **Matched1N10k** (Figure 24) is further from the truth than the unadjusted estimate of -0.3919 from the input data set **SimDataN10K**, without any adjustment for the confounders (Figure 25).

Figure 24 ATT Estimate after Matching on All Variables, $N=10,000$

PTLoc	Method	Mean	95% CL	Mean	Std Dev	95%	
						CL	Std Dev
0		14.2421	14.1416	14.3426	3.0625	2.9930	3.1352
1		14.4522	14.3519	14.5525	3.0569	2.9876	3.1296
Diff (1-2) Pooled		-0.2101	-0.3521	-0.0681	3.0597	3.0103	3.1107
Diff (1-2) Satterthwaite		-0.2101	-0.3521	-0.0681			

Figure 25 ATT Estimate with No Adjustment, $N=10,000$

The TTEST Procedure

Variable: PostEFrac

PTLoc	Method	Mean	95% CL	Mean	Std Dev	95%	
						CL	Std Dev
0		14.2421	14.1416	14.3426	3.0625	2.9930	3.1352
1		14.6340	14.5595	14.7085	3.0463	2.9946	3.0999
Diff (1-2) Pooled		-0.3919	-0.5168	-0.2671	3.0521	3.0104	3.0950
Diff (1-2) Satterthwaite		-0.3919	-0.5170	-0.2669			

The poor performance of the ATT estimate that is obtained after matching on all the pretreatment variables can be explained by examining the causal diagram for this model. The CAUSALGRAPH procedure, new in SAS/STAT 15.1, provides tools for analyzing graphical causal models. The following statements use the CAUSALGRAPH procedure to analyze the causal model shown in Figure 19 and test the adjustment set that contains all observed pretreatment variables:

```
proc causalgraph;
  model "PostSurgery"
    MedProp ==> Surgery GMInt PostEFrac PreEFrac,
    Support ==> GMInt PostEFrac,
    Dist ==> PTLoc GMInt,
    Surgery ==> PTLoc PostEFrac,
    PreEFrac ==> Surgery PTLoc PostEFrac,
    Age ==> PreEFrac PostEFrac PTLoc Surgery,
    PTLoc ==> PostEFrac
  ;
  unmeasured MedProp Support Dist;
  identify PTLoc ==> PostEFrac;
  testid "Adjust All"
    Surgery GMInt PreEFrac Age / paths=(noncausal nonblocked);
run;
```

To specify a causal model by using the CAUSALGRAPH procedure, you can specify directed paths or covariance paths in the MODEL statement. Each model that you specify in a MODEL statement must be given a label specified in a quoted string. Arrows are used to specify the directed edges in the model. For each variable that is listed near the tail of an arrow, there is an edge to, and causal effect on, each variable listed near the head of the arrow. You can specify paths that contain multiple arrows, so it is important to check the location of the commas that delineate the specified paths.

You specify variables that should be treated as unmeasured in the causal model by listing them in the UNMEASURED statement. In the IDENTIFY statement, you specify that the causal effect of **PTLoc** on **PostEFrac** is of interest. To test whether a proposed adjustment set is valid, you can use the TESTID statement. The PATHS= option requests the enumeration of treatment-to-outcome paths in the model. Based on the adjustment set, the paths in the model are classified as either causal or noncausal and either blocked or nonblocked. The NONCAUSAL and NONBLOCKED options in the PLOTS= request specify that only the treatment-to-outcomes paths that are noncausal and nonblocked should be enumerated. You can specify multiple MODEL and TESTID statements. Each adjustment set that you specify in a TESTID statement is tested for each model that you specify in a MODEL statement.

For this model, the “Covariate Adjustment Test” that PROC CAUSALGRAPH produces (Figure 26) shows that the collection of all observed pretreatment variables is not a valid adjustment set. Therefore, as observed in the analysis of the simulated data, the causal effect estimates that you obtain by adjusting for all these variables are not valid. The “Treatment-to-Outcome Paths” table (Figure 27) shows that for the proposed adjustment set, there are two unblocked noncausal paths. Both paths include **GMInt** as a collider, and adjusting for **GMInt** is responsible for opening these noncausal paths.

Figure 26 Covariate Adjustment Test for All Pretreatment Variables

The CAUSALGRAPH Procedure

Covariate Adjustment Test: Adjust All
Causal Effect of PTLoc on PostEFrac

Covariates

Model	Size	Valid	Minimal	Age	GMInt	PreEFrac	Surgery
PostSurgery	4	No	No	*	*	*	*

Figure 27 Unblocked Noncausal Paths for the Test Adjustment Set

The CAUSALGRAPH Procedure

Treatment-to-Outcome Paths
Test Adjust All for Model PostSurgery
Causal Effect of PTLoc on PostEFrac Adjusted for
{Age GMInt PreEFrac Surgery}

Causal Blocked Paths

1*	No	No	PTLoc <== Dist ==> GMInt <== Support ==> PostEFrac
2*	No	No	PTLoc <== Dist ==> GMInt <== MedProp ==> PostEFrac

* indicates a biasing path

As in the following statements, when you specify a causal effect to analyze in the IDENTIFY statement but do not specify any adjustment sets to test in the TESTID statement, the CAUSALGRAPH procedure constructs and displays all the valid adjustment sets:

```
proc causalgraph;
  model "PostSurgery"
    MedProp ==> Surgery GMInt PostEFrac PreEFrac,
    Support ==> GMInt PostEFrac,
    Dist ==> PTLoc GMInt,
    Surgery ==> PTLoc PostEFrac,
    PreEFrac ==> Surgery PTLoc PostEFrac,
    Age ==> PreEFrac PostEFrac PTLoc Surgery,
    PTLoc ==> PostEFrac
  ;
  unmeasured MedProp Support Dist;
  identify PTLoc ==> PostEFrac;
run;
```

The table of covariate adjustment sets (Figure 28) shows that there is one valid adjustment set for this causal model. The adjustment set consists of the variables **Age**, **Surgery**, and **PreEFrac**.

Figure 28 Adjustment Set for the Simulated Model
The CAUSALGRAPH Procedure

Covariate Adjustment Sets for PostSurgery						
Causal Effect of PTLoc on PostEFrac						
Covariates						
	Size	Minimal	Age	GMInt	PreEFrac	Surgery
1	3	Yes	*	*	*	*

The following statements reanalyze the **SimDataN1k** data set with the variable **GMInt** omitted from the propensity score model:

```
proc psmatch data=SimDataN1k;
  class Surgery PTLoc GMInt;
  psmodel PTLoc(Treated='0') = Surgery Age PreEFrac ;
  match method=optimal stat=lps caliper=0.5;
  assess lps var=(Age Surgery GMInt PreEFrac) / plots=StdDiff;
  output out(obs=match)=Matched2N1k weight=attwgt;
run;

proc ttest data=Matched2N1k;
  class PTLoc;
  var PostEFrac;
  weight attwgt;
run;
```

After you perform matching on this propensity score model, **Figure 29** shows good balance in the standardized mean differences for the variables in the adjustment set and no improvement in the balance for **GMInt**. The lack of balance for **GMInt** is acceptable for this example, based on the analysis of the causal model that the CAUSALGRAPH procedure performed. The outcome analysis that is performed with this matched data set provides an ATT estimate of -0.4953 (**Figure 30**), which is close to the true value of -0.5 specified by the simulation's design.

Figure 29 Standardized Mean Differences after Matching on Valid Adjustment Set, $N=1,000$

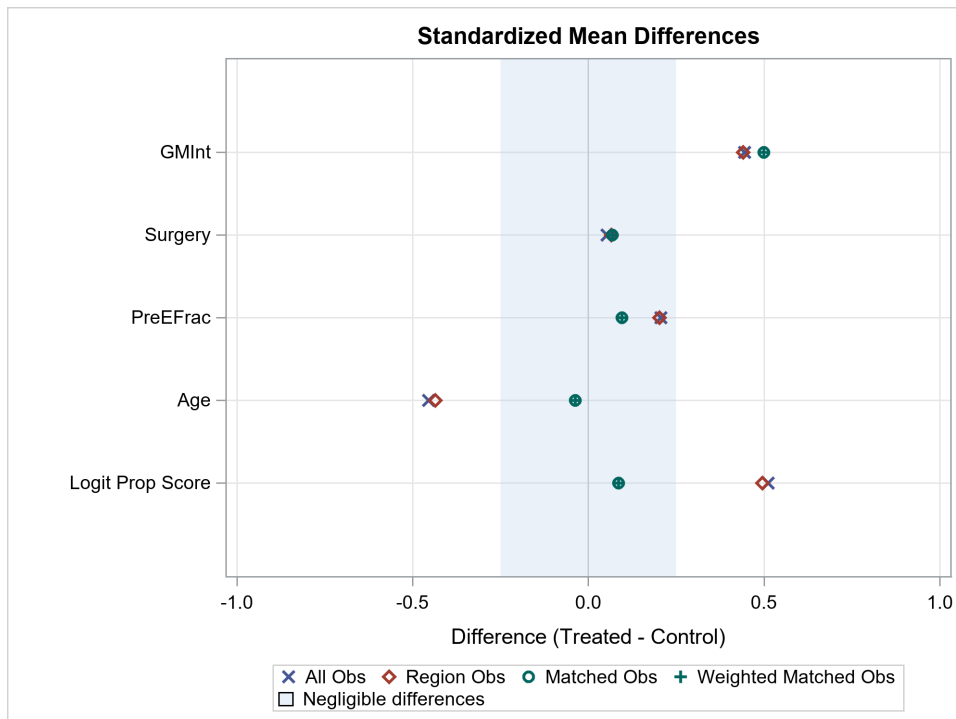


Figure 30 ATT Estimate after Matching on the Valid Adjustment Set, $N=1,000$

PTLoc	Method	Mean	95%				
			95% CL	Mean	Std Dev	CL	Std Dev
0		16.0985	15.7955	16.4015	3.0277	2.8281	3.2578
1		16.5938	16.2912	16.8964	3.0239	2.8245	3.2537
	Diff (1-2) Pooled	-0.4953	-0.9228	-0.0677	3.0258	2.8819	3.1849
	Diff (1-2) Satterthwaite	-0.4953	-0.9228	-0.0677			

Although this example is simplified and uses simulated data, it does demonstrate the importance of considering the causal model for the data generating process. For this example, creating balance across all the observed pretreatment variables is possible, but it does not lead to a valid causal effect estimate because of unmeasured variables in the data generating process. Unmeasured variables can take the form of latent constructs that are difficult to capture in the data, such as **MedProp** in this simulation. The variable **Dist** in this example represents another type of unmeasured variable that is used in the data generating process but is omitted from the data set for privacy considerations or similar factors.

In this example, estimation of the causal effect is possible when you adjust for only a subset of the observed variables. This might not be typical of all problems. If no valid adjustment set exists for a causal model and the data are not yet collected, you might consider how to modify the design of the study to change the data generating process, or you might identify additional information that you can collect to make the estimation of the causal effect feasible. This example is also atypical because it uses simulated data and because there is no doubt about the correctness of the causal model analyzed by the CAUSALGRAPH procedure. In practice, the correct form of the data generating process's causal model is likely to be uncertain. In addition to analyzing adjustment sets, you can also use PROC CAUSALGRAPH to produce information about conditional independence assumptions that are encoded in the causal model. Checking whether the assumptions that are implied by the causal model are observed in the data provides one basis for comparing the plausibility of different causal models. For more information about the CAUSALGRAPH procedure, see Thompson (2019). You should also conduct sensitivity analyses to assess how unmeasured confounding might affect the results of an analysis when you are uncertain about the underlying causal model. For more information about performing sensitivity analyses, see Rosenbaum (2010), Liu, Kuramoto, and Stuart (2013), Imbens and Rubin (2015), Pan and Bai (2016), VanderWeele and Ding (2017), and references therein. An example of one approach to performing a sensitivity analysis after a one-to-one matching is provided in the documentation for the PSMATCH procedure (SAS Institute Inc. 2018).

SUMMARY

In observational studies, confounding variables can introduce noncausal sources of association between the treatment and outcome of interest. The noncausal association that confounding variables introduce can bias many traditional methods of estimating a treatment's causal effect. Determining whether a causal effect can be estimated from an observational data set requires careful consideration of how the data were generated. Assumptions about the data generating process are key to evaluating whether adjusting for a set of variables blocks all the noncausal sources of association between the treatment and outcome of interest. Graphical causal models provide a useful framework for representing assumptions about the data generating process. In SAS/STAT 15.1, you can use the CAUSALGRAPH procedure to analyze graphical causal models to test whether an adjustment set is valid or to construct all valid adjustment sets for a model.

Propensity score matching is an intuitive approach that you can use to mitigate the effects of confounding variables. In a well-matched data set, the treatment conditions are comparable in terms of the relevant confounders, and you can then estimate the treatment effect of interest by using familiar techniques. You can use the PSMATCH procedure in SAS/STAT software to perform propensity score matching and to evaluate the covariate balance achieved in the matched data set. The propensity score values that are used to create a matched data set can be estimated internally by PROC PSMATCH using a logistic regression model, or you can provide precomputed propensity score values by using the PSDATA statement. You can also modify properties of the matched data set by using different methods of matching subjects, by imposing constraints on the matching using caliper values, by requiring the matching to be exact on certain classification variables, or by using different support regions. The flexibility that these methods provide enables you to consider different formulations for a matching problem and to examine the covariate balance achieved in the different matched data sets before you select a data set to use in an outcome analysis.

REFERENCES

- Austin, P. C. (2009). "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." *Statistics in Medicine* 28:3083–3107.
- Austin, P. C. (2011). "Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies." *Pharmaceutical Statistics* 10:150–161.
- Austin, P. C. (2014). "A Comparison of 12 Algorithms for Matching on the Propensity Score." *Statistics in Medicine* 33:1057–1069.
- Elwert, F. (2013). "Graphical Causal Models." In *Handbook of Causal Analysis for Social Research*, edited by S. L. Morgan, 245–273. Dordrecht: Springer.
- Friedman, J. H. (1991). "Multivariate Adaptive Regression Splines." *Annals of Statistics* 19:1–67.
- Guo, S., and Fraser, M. W. (2015). *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: Sage Publications.
- Hernán, M. A., and Robins, J. M. (2019). *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC. Forthcoming.
- Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.
- Liu, W., Kuramoto, S. J., and Stuart, E. A. (2013). "An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-experimental Prevention Research." *Prevention Science* 14:570–580.
- Mamdani, M., Sykora, K., Li, P., Normand, S. L., Streiner, D. L., Austin, P. C., Rochon, P. A., and Anderson, G. M. (2005). "Reader's Guide to Critical Appraisal of Cohort Studies: 2. Assessing Potential for Confounding." *BMJ* 330:960–962.
- Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). "On the Application of Probability Theory to Agricultural Experiments: Essay on Principles, Section 9." *Statistical Science* 5:465–472. Translated and edited by Dabrowska and Speed from the Polish original by Neyman (1923).
- Normand, S.-L. T., Landrum, M. B., Guadagnoli, E., Ayanian, J. Z., Ryan, T. J., Cleary, P. D., and McNeil, B. J. (2001). "Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores." *Journal of Clinical Epidemiology* 54:387–398.
- Pan, W., and Bai, H. (2016). "A Robustness Index of Propensity Score Estimation to Uncontrolled Confounders." In *Statistical Causal Inferences and Their Applications in Public Health Research*, edited by H. He, P. Wu, and D.-G. Chen, 91–100. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-41259-7_4.
- Pearl, J. (2009a). "Causal Inference in Statistics: An Overview." *Statistics Surveys* 3:96–146.
- Pearl, J. (2009b). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.
- Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York: Springer-Verlag.
- Rosenbaum, P. R., and Rubin, D. B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41–55.
- Rubin, D. B. (1980). "Comment on D. Basu, 'Randomization Analysis of Experimental Data: The Fisher Randomization Test'." *Journal of the American Statistical Association* 75:591–593.
- Rubin, D. B. (1990). "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5:472–480.
- Rubin, D. B. (2001). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services and Outcomes Research Methodology* 2:169–188.
- SAS Institute Inc. (2018). *SAS/STAT 15.1 User's Guide*. Cary, NC: SAS Institute Inc. <http://go.documentation.sas.com/?docsetId=statug&docsetTarget=titlepage.htm&docsetVersion=15.1&locale=en>.

- Shpitser, I., VanderWeele, T., and Robins, J. M. (2010). "On the Validity of Covariate Adjustment for Estimating Causal Effects." In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, edited by P. Grünwald and P. Spirtes, 527–536. Corvallis, OR: AUAI Press.
- Spirtes, P., Glymour, C., and Scheines, R. (2001). *Causation, Prediction, and Search*. 2nd ed. Cambridge, MA: MIT Press.
- Stuart, E. A. (2010). "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25:1–21.
- Thompson, W. C. (2019). "Causal Graph Analysis with the CAUSALGRAPH Procedure." In *Proceedings of the SAS Global Forum 2019 Conference*. Cary, NC: SAS Institute Inc.
- Van der Zander, B., Liškiewicz, M., and Textor, J. (2014). "Constructing Separators and Adjustment Sets in Ancestral Graphs." In *Proceedings of the Thirtieth Conference on Causal Inference: Learning and Prediction*, edited by N. L. Zhang and J. Tian, 11–24. Corvallis, OR: AUAI Press.
- VanderWeele, T. J., and Ding, P. (2017). "Sensitivity Analysis in Observational Research: Introducing the E-Value." *Annals of Internal Medicine* 167:268–274. <https://doi.org/10.7326/M16-2607>.
- Yuan, Y., Yung, Y.-F., and Stokes, M. (2017). "Propensity Score Methods for Causal Inference with the PSMATCH Procedure." In *Proceedings of the SAS Global Forum 2017 Conference*. Cary, NC: SAS Institute Inc. <http://support.sas.com/resources/papers/proceedings17/SAS0332-2017.pdf>.

ACKNOWLEDGMENTS

The authors are grateful to Ed Huddleston for his valuable editorial assistance and to Gordon Brown for his helpful suggestions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors:

Michael Lamm	Clay Thompson	Yiu-Fai Yung
SAS Institute Inc.	SAS Institute Inc.	SAS Institute Inc.
SAS Campus Drive	SAS Campus Drive	SAS Campus Drive
Cary, NC 27513	Cary, NC 27513	Cary, NC 27513
Michael.Lamm@sas.com	Clay.Thompson@sas.com	Yiu-Fai.Yung@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.