Paper SAS374-2017

Estimating Causal Effects from Observational Data with the CAUSALTRT Procedure

Michael Lamm and Yiu-Fai Yung, SAS Institute Inc.

ABSTRACT

Randomized control trials have long been considered the gold standard for establishing causal treatment effects. Can causal effects be reasonably estimated from observational data too? In observational studies, you observe treatment T and outcome Y without controlling confounding variables that might explain the observed associations between T and Y. Estimating the causal effect of the treatment T therefore requires adjustments that remove the effects of the confounding variables. The new CAUSALTRT (causal-treat) procedure in SAS/STAT® 14.2 enables you to estimate the causal effect of a treatment decision by modeling either the treatment assignment T or the outcome Y, or both. Specifically, modeling the treatment leads to inverse probability weighting methods, and modeling the outcome leads to regression methods. Combined modeling of the treatment and outcome leads to doubly robust methods that can provide unbiased estimates for the treatment effect even if one of the models is misspecified. This paper reviews the statistical methods that are implemented in the CAUSALTRT procedure and includes examples of how you can use this procedure to estimate causal effects from observational data. This paper also illustrates some other important features of the CAUSALTRT procedure, including bootstrap resampling, covariate balance diagnostics, and statistical graphics.

INTRODUCTION

Researchers, clinicians, and policy makers in many disciplines are often interested in estimating the causal effect that a treatment decision (such as taking a drug versus an alternative therapy), exposure to a condition (such as smoking cessation versus control), or intervention (such as attending a private versus public school) has on an outcome. In an experiment or randomized control trial (where each subject is randomly assigned to a treatment condition), you can safely assume that there are no confounding variables that are associated with both the treatment assignment T and the outcome Y. The causal effect of T on Y can therefore be directly estimated from experimental data without a need to perform any additional adjustments to remove the effects of confounding variables.

In many settings it is either infeasible or immoral to randomly assign subjects to treatment conditions, so the only available data are from nonrandomized trials or observational studies. To estimate causal treatment effects in these settings, the CAUSALTRT (causal-treat) procedure, introduced in SAS/STAT 14.2, implements estimation methods that are designed primarily for use with observational data. Specifically, the procedure estimates the average causal effect of a binary treatment, T, on a continuous or discrete outcome, Y, in the presence of confounding variables.

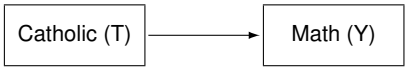
The remainder of the paper is organized as follows. The next section, Example 1, illustrates the difficulties of estimating causal treatment effects in observational studies. It explains how statistical techniques for analyzing experimental data can fail for observational data and thus motivates the need for the statistical techniques implemented by the CAUSALTRT procedure. The main features of the CAUSALTRT procedure are described after Example 1, followed by a section that compares the CAUSALTRT procedure to the PSMATCH procedure. The theoretical foundations of causal treatment effect analysis, on which the CAUSALTRT procedure is based, are then presented. The theoretical section contains technical details and can be skipped in a first r eading. Next, Example 2 extends the analysis of Example 1 and illustrates the application of the CAUSALTRT procedure. It shows how you can specify the syntax of the CAUSALTRT procedure and interpret the analysis results. An additional example, Example 3, provides a further illustration of the diagnostic features of the CAUSALTRT procedure. The final section provides a summary of the paper.

EXAMPLE 1: DOES ATTENDING CATHOLIC SCHOOLS HAVE CAUSAL EFFECTS ON ACADEMIC ACHIEVEMENTS?

Coleman, Hoffer, and Kilgore (1982) studied whether Catholic schools in the United States were more effective than public schools in educating high school students (see also Murnane and Willett 2011 for a discussion of the

analysis). One of the many interesting hypotheses is that attending Catholic schools can enhance high school students' academic performance (for example, their scores in mathematics).

Suppose you could perform an experiment that randomly assigns high school students to Catholic or public schools and then observes their math scores at the end of their high school education. Presumably, the causal effect of attending Catholic schools (T with values Yes or No) on math scores (Y) could be inferred directly from the experimental results. The following causal diagram illustrates such an idealized situation.



The arrow in this diagram represents the direct influence of attending Catholic schools (T) on math scores (Y). Is this direct influence a causal effect, and can you estimate it? The answers are yes if this idealized causal diagram is as simple as it is shown here—that is, there are no missing variables and arrows that if included would open new pathways and complicate the relationship between T and Y. Such an idealized causal diagram might be plausible for experimental data where either you control for all the confounding factors or you use randomization to assign subjects to the conditions of T so that all other factors would average out.

Murnane and Willett (2011) analyzed a data set of 592 students attending Catholic high schools and 5,079 students attending public schools. The students' math scores were measured to indicate their academic achievements (see Example 2 for more details about this data set). For the moment, assume that the data satisfy all the requirements for the preceding idealized causal diagram. Now you want to compute the causal effect of T on Y using these data. One way to do that is to compute the difference in average math scores between Catholic and public school students, as shown in the following statements:

```
proc ttest data=school;
   class Catholic;
   var Math;
run;
```

Figure 1 shows that the average math score for students who attend public schools is 3.8949 points lower than the average score for students who attend Catholic schools. Thus, you could infer that attending a Catholic high school has an average causal treatment effect of 3.8949 points on math scores. This effect is significant at the 0.05 α-level by various *t*-test criteria (not shown) and has a causal interpretation under the current assumptions about the data and the causal diagram.

Figure 1 t Test of Means

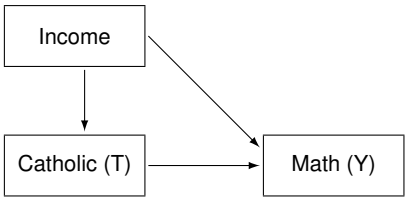
The TTEST Procedure

Variable: Math

Catholic	N	Mean	Std Dev	Std Err	Minimum	Maximum
No	5079	50.6447	9.5343	0.1338	29.8800	71.3700
Yes	592	54.5395	8.4632	0.3478	32.9200	71.0800
Diff (1-2)		-3.8949	9.4283	0.4095		

Unfortunately, the data were not collected from an experiment. They were collected in an observational study where students "selected" their own schools (that is, their selections determined their treatment conditions). As a result, the preceding causal diagram is not an accurate representation of the data, and the estimation of the causal effect by using a *t* test is problematic. How did the observational nature of the study make the estimation of the causal effect by such a simple *t* test problematic?

Consider the following alternative causal diagram for this observational study:



In this causal diagram, the variable **Income** (household income) represents a pretreatment characteristic that affects both the math scores and the decision of whether to attend a Catholic school. Students in families that have higher household incomes are more likely to be able to afford tuition at a Catholic school and also have more resources to invest in the students' academic developments. In this causal diagram, **Income** serves as a pretreatment confounder—it might inflate the association between T and Y so that the observed difference in math scores between Catholic and public school students is larger than the actual causal effect of interest. Consequently, the observed difference in average math scores between Catholic and public school students is a biased causal effect estimate and the corresponding *t* test is not valid.

You can obtain unbiased estimation of causal effects in such a confounded situation by using the statistical methods implemented by the CAUSALTRT procedure. Example 2 continues the analysis of this data set and illustrates one of the causal inference techniques that PROC CAUSALTRT implements.

It is important to notice that the "improved" causal diagram shown in this example is still a simplified picture for illustration. Indeed, there could be other confounding pretreatment characteristics in the picture. One critical assumption of the causal inference methods that are implemented by the CAUSALTRT procedure is that all important confounding pretreatment variables must be included in the analysis (that is, "no unmeasured confounding" is assumed). If this assumption and all other necessary assumptions are satisfied, PROC CAUSALTRT can yield unbiased estimation of causal effects. Example 2 assumes that all these confounders have been included and illustrates the estimation of causal effects.

MAIN FEATURES OF THE CAUSALTRT PROCEDURE

The CAUSALTRT procedure estimates two types of causal effects for a binary treatment T on an outcome Y, where the outcome can be continuous or binary. Depending on the subset of the study population of interest, the following two types of causal treatment effects can be estimated by PROC CAUSALTRT:

- Average treatment effect (ATE)—also known as average causal effect (ACE)—is the causal effect of the treatment T within the entire study population.
- Average treatment effect for the treated (ATT)—sometimes also abbreviated as ATET—is the causal effect of the treatment T within the subset of the study population that is in the treatment condition.

To illustrate these two definitions of causal effects, consider the data and causal question from Example 1. If you want to study the effect of school type (Catholic versus public) on the math performance for the entire population of students irrespective of their actual school enrollment, then the ATE is the causal effect that you want to estimate. This is also the default causal effect that PROC CAUSALTRT estimates. If you want to study the effect of school type on the math performance only for those students who are in the treatment condition (that is, enrolled in Catholic schools), then the ATT is the causal effect that you want to estimate. You can use the ATT option in PROC CAUSALTRT to estimate this effect.

To estimate either type of treatment effect with a valid causal interpretation from observational data, adjustments are made to remove the confounding effects of pretreatment characteristics by modeling the treatment T or the outcome Y, or both. According to which models are specified, PROC CAUSALTRT implements the following three broad categories of statistical methods for causal effect estimation:

- Modeling of the treatment T leads to inverse probability weighting (IPW) estimation methods. The CAUSALTRT procedure implements three different IPW estimation methods for the ATE and one IPW estimation method for the ATT. The predicted probability of being in the treatment condition is called the propensity score, and the model for the treatment T is therefore also called the propensity score model.
- Modeling of the outcome Y leads to the regression adjustment estimation method. Both the ATE and ATT can be estimated by regression adjustment.
- Modeling of both the treatment T and outcome Y leads to doubly robust estimation methods, which provide unbiased estimates for the treatment effect even if one of the models is misspecified. The CAUSALTRT procedure implements two doubly robust estimation methods for the ATE.

For all the estimation methods, standard errors and confidence limits can be computed either by using analytic formulas that are based on the asymptotics or by using bootstrap resampling methods. PROC CAUSALTRT uses multithreaded computation for model estimation and for bootstrapping.

Based on the balancing score property of the propensity score (Rosenbaum and Rubin 1983), diagnostics for assessing the propensity score model can be obtained by investigating the balance that a model produces. The CAUSALTRT procedure supports the following diagnostics for assessing the balance that a propensity score model produces:

- weighted and unweighted standardized mean differences (between treatment and control conditions) and variance ratios (treatment to control) for the covariates that are used to fit the propensity score model
- plots of the propensity scores and weights by treatment conditions
- weighted and unweighted kernel density plots for the continuous covariates in the propensity score model

COMPARISON TO THE PSMATCH PROCEDURE

The PSMATCH procedure, new in SAS/STAT 14.2, also provides a variety of tools for propensity score analysis. The PSMATCH procedure either computes propensity scores or reads previously computed propensity scores, and it provides the following methods for using the scores to allow for valid estimation of treatment effect in a subsequent outcome analysis: inverse probability of treatment weighting, stratification, and matching.

Both the CAUSALTRT procedure and the PSMATCH procedure provide tools for addressing the issue of confounding variables in observational data. In particular, both procedures support the computation of inverse probability of treatment weights from the propensity scores and both provide diagnostics for assessing covariate balance after weighting. The PSMATCH procedure also provides diagnostics for assessing covariate balance after matching or stratifying the data based on the propensity scores. PROC PSMATCH supports matching methods and stratification of the data by the propensity scores; PROC CAUSALTRT does not support matching or stratification. Further differences between the PSMATCH and CAUSALTRT procedures are that the PSMATCH procedure does not involve the outcome variable Y and therefore does not perform any outcome analysis. Because the PSMATCH procedure does not involve the outcome variable Y, it can be used to explore how well you can balance the covariates between the treatment conditions by using different propensity score models. After adequate balance is achieved across the confounding variables, the PSMATCH procedure can create an output data set for a subsequent outcome analysis that mimics the analysis you would perform with data from a randomized study. For example, if you used matching with the PSMATCH procedure, a simple univariate test or analysis might be sufficient to estimate the treatment effect.

CAUSAL EFFECTS: DEFINITION, IDENTIFICATION, AND ESTIMATION

This section describes the theoretical foundations on which the CAUSALTRT procedure is based. It contains technical details that some readers might want to skip in their first reading.

The causal effects that PROC CAUSALTRT estimates are defined using the Neyman-Rubin potential outcome framework (Rubin 1980, 1990). Potential outcomes describe an idealized source of data, where it is possible to observe a subject's response for all possible treatment assignments. Suppose T represents a binary treatment designation, where $T = 0$ corresponds to the control condition and $T = 1$ corresponds to the treatment condition.

For each subject, there are two potential outcomes: $Y(1)$ and $Y(0)$. The subject-level treatment effect is then defined by the difference in potential outcomes, $Y(1) - Y(0)$.

To ensure that the potential outcomes and subject-level treatment effect are well defined, Rubin (1980) stated the stable unit treatment value assumption (SUTVA). See also Imbens and Rubin (2015). Two components of the SUTVA are the following:

- No interference. A subject's potential outcomes are not affected by the treatment assignments of the other subjects.
- No hidden variations of treatments. Subjects must receive the same form of treatment at each treatment level.

In order to attempt to estimate the causal effect of a treatment from the data, the following consistency assumption that relates an observed outcome to the potential outcomes is needed:

$$Y = TY(1) + (1 - T)Y(0)$$

This equation states that the observed response Y is equal to the potential outcome with a treatment level that matches the actual assigned treatment level. Because each subject can take part in only one treatment condition, at least half of the potential outcomes would not be observed in the data collection process. Therefore, the unit-level causal effect, as defined by $Y(1) - Y(0)$, is seldom the primary interest in research. Instead, average treatment effects in the population are the more common estimands.

The CAUSALTRT procedure estimates two types of treatment effects:

- The average treatment effect (ATE) for the entire population is given by

  $$\text{ATE} = \mu_1 - \mu_0 = E[Y(1)] - E[Y(0)]$$

  where $\mu_1 = E[Y(1)]$ and $\mu_0 = E[Y(0)]$ are potential outcome means for the treatment and control conditions, respectively.

- The average treatment effect for the treated (ATT, also called ATET) is the average causal effect among only those individuals that receive treatment; it is given by

  $$\text{ATT} = \mu_{1|T=1} - \mu_{0|T=1} = E[Y(1)|T = 1] - E[Y(0)|T = 1]$$

  where $\mu_{1|T=1} = E[Y(1)|T = 1]$ and $\mu_{0|T=1} = E[Y(0)|T = 1]$ are potential outcome means for the treatment and control conditions, respectively, conditional on receiving treatment.

The assumptions that are made to construct the potential outcome framework and to define the ATE and ATT include an implicit but important condition that is pointed out in Rubin (2005): the potential outcomes are not affected by the treatment assignment mechanism that the researchers use to learn about them. Whether in an observational study or a randomized experiment, potential outcomes remain unique to the subjects given the well-defined causal problem and SUTVA. Although the definitions of causal treatment effects within the potential outcome framework do not depend on the source of the data, the statistical methodology for estimating the effects does depend on the treatment assignment mechanism.

When randomization is used to assign subjects to treatment conditions, you can safely assume that the potential outcomes and treatment T are independent. So

$$E[Y(j)] = E[Y(j)|T = 0] = E[Y(j)|T = 1], \quad j = 0, 1$$

In this case, there is no difference between the ATE and ATT, and the causal treatment effect can be identified from the observed data by

$$E[Y|T = 1] - E[Y|T = 0]$$

In observational studies, subjects "select" the treatment conditions based on their pretreatment characteristics (covariates), $\mathbf{X}$, which could also be associated with the outcome variable. The association of treatment and outcome is confounded by the covariates. In this case, identification of causal effects requires that the strong ignorability

assumption (also called the conditional exchangeability assumption) be satisfied; that is, causal effects can be identified only when the estimation of the causal effects can successfully take into account all the confounding that is caused by the covariates $\mathbf{X}$.

The CAUSALTRT procedure implements several statistical techniques that enable you to estimate treatment effects when you specify confounding covariates, $\mathbf{X}$. PROC CAUSALTRT uses the $\mathbf{X}$ covariates to fit generalized linear models for the treatment T or the outcome Y, or both. Predicted values from these models are then incorporated into the estimation of the causal effects. The next section describes the estimation of the ATE by using the doubly robust augmented inverse probability weighting estimation method.

AUGMENTED INVERSE PROBABILITY WEIGHTING

Estimating the ATE by augmented inverse probability weighting (AIPW) combines modeling of the treatment variable T and of the outcome variable Y. The AIPW estimation method is doubly robust because it produces unbiased estimates for the potential outcome means and the ATE even if one of the models is misspecified.

The inverse probability weighting component of the AIPW estimation method is based on modeling the treatment assignment T. For a set of observed pretreatment covariates $\mathbf{x}_{\mathrm{ps}}$, denote the conditional probability of receiving treatment by

$$e(\mathbf{x}_{\mathrm{ps}}) = \Pr(T = 1 \mid \mathbf{x}_{\mathrm{ps}})$$

The inverse probability weight for an observation is equal to

$$\frac{1}{\Pr\left(T = t \mid \mathbf{x}_{\mathrm{ps}}\right)} = \frac{t}{e(\mathbf{x}_{\mathrm{ps}})} + \frac{1-t}{1-e(\mathbf{x}_{\mathrm{ps}})}$$

The conditional probability $e(\mathbf{x}_{\mathrm{ps}})$ is also called the propensity score, and the model that is used to predict $e(\mathbf{x}_{\mathrm{ps}})$ is called the propensity score model.

When the estimation of a causal effect involves inverse probability weighting, in addition to the SUTVA, the positivity assumption (that $0 < e(\mathbf{x}_{\mathrm{ps}}) < 1$), is required to ensure that each observation has a nonzero probability of being in each treatment condition. If the positivity, SUTVA, and strong ignorability assumptions are all satisfied, then it follows that the ATE can be identified from the data by

$$E\left[\frac{TY}{e(\mathbf{x}_{\mathrm{ps}})}\right] - E\left[\frac{(1-T)Y}{1-e(\mathbf{x}_{\mathrm{ps}})}\right]$$

This expression provides the basis for the inverse probability weighting approach for estimating causal treatment effects. The CAUSALTRT procedure implements three different estimation methods that involve only inverse probability weights and the propensity score model. These three methods are not described in detail in this paper. For more information about these methods, see Chapter 33, "The CAUSALTRT Procedure" (*SAS/STAT 14.2 User's Guide*).

To predict the propensity scores, the CAUSALTRT procedure fits a logistic regression model for the treatment T that uses the covariates $\mathbf{x}_{\mathrm{ps}}$. The parameter estimates for the propensity score model are denoted by $\hat{\boldsymbol{\beta}}_{\mathrm{ps}}$, and the predicted values for the propensity score are given by

$$\hat{e} = \hat{e}(\mathbf{x}_{\mathrm{ps}}) = \frac{\exp\left(\mathbf{x}'_{\mathrm{ps}}\hat{\boldsymbol{\beta}}_{\mathrm{ps}}\right)}{1 + \exp\left(\mathbf{x}'_{\mathrm{ps}}\hat{\boldsymbol{\beta}}_{\mathrm{ps}}\right)}$$

The AIPW estimation method builds on the inverse probability weighting approaches by incorporating a model for the outcome variable Y into the estimation of the ATE. For each treatment condition, the CAUSALTRT procedure fits a separate generalized linear model for the outcome variable by using a set of pretreatment covariates, which are denoted by $\mathbf{x}_{\mathrm{reg}}$. From the outcome models, potential outcomes that correspond to each treatment condition are estimated for all the subjects. The predicted potential outcomes are given by

$$\hat{y}_0 = g^{-1}(\mathbf{x}'_{\mathrm{reg}}\hat{\boldsymbol{\beta}}_c)$$

$$\hat{y}_1 = g^{-1}(\mathbf{x}'_{\mathrm{reg}}\hat{\boldsymbol{\beta}}_t)$$

where g is the link function being used, $\hat{\boldsymbol{\beta}}_{\mathrm{c}}$ are the parameter estimates for the control outcome model, and $\hat{\boldsymbol{\beta}}_{\mathrm{t}}$ are the parameter estimates for the treatment outcome model.

To estimate the potential outcome means and ATE by using the AIPW estimation method, suppose you have observations $(y_i, t_i, \mathbf{x}_i)$, for $i = 1, \ldots, n$, where the vector $\mathbf{x}_i$ includes all the covariates in the propensity score model and the outcome model. The AIPW estimates for the potential outcome means are given by

$$\hat{\mu}_0^{\text{aipw}} = n^{-1} \sum_{i=1}^{n} \frac{(1-t_i)y_i}{1-\hat{e}_i} + \hat{y}_{i0}\left(\frac{t_i - \hat{e}_i}{1-\hat{e}_i}\right)$$

$$\hat{\mu}_1^{\text{aipw}} = n^{-1} \sum_{i=1}^{n} \frac{t_i y_i}{\hat{e}_i} - \hat{y}_{i1}\left(\frac{t_i - \hat{e}_i}{\hat{e}_i}\right)$$

The estimates $\left(\hat{\mu}_0^{\text{aipw}}, \hat{\mu}_1^{\text{aipw}}\right)$ solve the estimating equations

$$\mathbf{S}_{\text{aipw}}(\boldsymbol{\mu}) = \sum_{i=1}^{n} \mathbf{S}_{\text{aipw},i}(\boldsymbol{\mu}) = \mathbf{0}$$

for (μ_0, μ_1), where

$$\mathbf{S}_{\text{aipw},i}(\boldsymbol{\mu}) = \begin{bmatrix} \frac{(1-t_i)y_i}{1-\hat{e}_i} + \hat{y}_{i0}\left(\frac{t_i-\hat{e}_i}{1-\hat{e}_i}\right) - \mu_0 \\ \frac{t_i y_i}{\hat{e}_i} - \hat{y}_{i1}\left(\frac{t_i-\hat{e}_i}{\hat{e}_i}\right) - \mu_1 \end{bmatrix} = \begin{bmatrix} \hat{y}_{i0} - \left(\frac{1-t_i}{1-\hat{e}_i}\right)(\hat{y}_{i0} - y_i) - \mu_0 \\ \hat{y}_{i1} - \left(\frac{t_i}{\hat{e}_i}\right)(\hat{y}_{i1} - y_i) - \mu_1 \end{bmatrix}$$

The doubly robust property of the AIPW estimation method follows from the fact that the following equation is equal to the ATE if either the propensity score model or the outcome model is correctly specified:

$$E\left[\frac{TY}{\hat{e}} - \hat{y}_1\left(\frac{T-\hat{e}}{\hat{e}}\right)\right] - E\left[\frac{(1-T)Y}{1-\hat{e}} + \hat{y}_0\left(\frac{T-\hat{e}}{1-\hat{e}}\right)\right]$$
$$= E\left[\hat{y}_1 - \left(\frac{T}{\hat{e}}\right)(\hat{y}_1 - Y)\right] - E\left[\hat{y}_0 - \left(\frac{1-T}{1-\hat{e}}\right)(\hat{y}_0 - Y)\right]$$

Furthermore, estimates for the covariance matrix of $(\hat{\mu}_0^{\text{aipw}}, \hat{\mu}_1^{\text{aipw}})$ can be derived using the theory of M-estimation (Stefanski and Boos 2002). Because the ATE is equal to the difference in the potential outcome means, its variance is computed by applying standard variance rules to the estimate for the covariance matrix for the potential outcomes. This analytic approach of using the estimating equations and the theory of M-estimation to estimate the covariance matrix of the potential outcome means is used for all the estimation methods that PROC CAUSALTRT implements. To provide better theoretical and computational properties when computing the covariance matrix, the CAUSALTRT procedure uses certain asymptotically equivalent expressions for components of the estimates. For descriptions of the modifications, their motivation, and theoretical justification, see Pierce (1982); Robins, Rotnitzky, and Zhao (1995); Lunceford and Davidian (2004); Wooldridge (2010).

Although the AIPW estimates for the potential outcome means and ATE are doubly robust, the estimates for their variances are not. In addition to the analytic variance estimates, PROC CAUSALTRT can compute bootstrap-based standard errors and confidence limits.

EXAMPLE 2: THE AIPW METHOD FOR ESTIMATING THE CAUSAL EFFECT OF SCHOOL TYPE ON MATH PERFORMANCE

This example extends the analysis of Example 1 and illustrates the main features of the CAUSALTRT procedure. Previously, it was argued that the observed difference in average math scores of 3.8949 points between Catholic and public school students was a biased estimate of the average causal treatment effect (ATE) because the pretreatment characteristics were not accounted for. This example demonstrates how to use the CAUSALTRT procedure to perform a doubly robust estimation of the ATE by using the augmented inverse probability weighting (AIPW) method.

To recapitulate, the causal question of interest is the effect that attending a Catholic school has on students' performance in mathematics. The data for this example are a subset of the data from the National Educational Longitudinal Study of 1988 in Murnane and Willett (2011). Demographic information was collected from the students and their families along with standardized test scores in 1988 and subsequent follow-up years. For this analysis, the population is restricted to students whose 1988 family income was at most $75,000. The **School** data set used for this example is created in the following DATA step:

```
data School;
   input
   Income $ FatherEd $ MotherEd $ Math BaseMath  Catholic $;
   datalines;
   Middle HighSchool    HighSchool    49.77 50.27 Yes
   Middle Unknown       Unknown       59.84 51.52 Yes
   Middle NoHighSchool  Postsecondary 50.38 47.56 Yes
   Middle Unknown       Unknown       45.03 46.60 Yes
   High   HighSchool    HighSchool    54.26 60.22 Yes

   ... more lines ...

   Middle HighSchool    HighSchool    41.69 48.29 Yes
   High   College       College       56.60 60.15 Yes
   High   College       SomeSecondary 56.29 59.53 Yes
   High   College       Postsecondary 58.16 57.06 Yes
   High   College       SomeSecondary 63.57 63.51 Yes
;
```

The data set **School** consists of records for 5,671 high school students. The variables in the data set are as follows:

- **BaseMath**: student's score on the mathematics portion of a standardized test in 1988
- **Catholic**: Yes if a student attended a Catholic high school; No otherwise
- **FatherEd**: highest level of education completed by the student's father, with six levels
- **Income**: classification based on total family income, with values of Low, Middle, and High
- **Math**: student's score on the mathematics portion of a standardized test in 1992
- **MotherEd**: highest level of education completed by the student's mother, with six levels

The outcome variable for this example is the student's 1992 score on the mathematics portion of a standardized test, and the treatment variable is the indicator of the type of school the student attends.

As discussed in Example 1, this data set consists of observational data, and the random assignment of students to schools should not be assumed. Instead, pretreatment characteristics such as family income, educational levels of father and mother, and the baseline math performance could have contributed to the observed associations between school type and math performance in 1992. These pretreatment characteristics must also be taken into account in order to properly estimate the causal effect.

The following statements invoke PROC CAUSALTRT to estimate the ATE of attending a Catholic high school on math scores:

```
ods graphics on;
proc causaltrt data=school covdiffps poutcomemod nthreads=2;
   class Income FatherEd MotherEd;
   psmodel Catholic(ref='No') = Income FatherEd MotherEd;
   model Math = BaseMath Income FatherEd MotherEd;
   bootstrap seed=1234 plots=hist(effect);
run;
```

All the estimation methods that PROC CAUSALTRT implements require that you specify the outcome variable in the MODEL statement and the treatment variable in the PSMODEL statement.

In the PSMODEL statement, the treatment variable **Catholic** is specified with the REF='NO' option, which identifies **Catholic**='No' (in other words, public schools) as the reference (control) level for this analysis. The variables that are also specified in the PSMODEL statement (**Income**, **FatherEd**, and **MotherEd**) are pretreatment characteristics that are associated with the choice of Catholic schools. These variables are used to fit the propensity score model and should also be associated with the math performance in 1992.

The MODEL statement specifies the outcome variable **Math** and specifies the variables **BaseMath**, **Income**, **FatherEd**, and **MotherEd** to serve as covariates in the outcome model.

Because the PROC CAUSALTRT statement does not include a METHOD= option to specify an estimation method, the CAUSALTRT procedure chooses a default estimation method based on the model specifications provided in the MODEL and PSMODEL statements. In the current example, the AIPW estimation method is selected because both the outcome and treatment models are specified, in the MODEL and PSMODEL statements, respectively. The AIPW method fits both the treatment and outcome models separately and then combines the results to estimate the treatment's average causal effect.

The CAUSALTRT procedure produces the "Model Information" table (Figure 2), which provides a summary of the treatment variable, outcome variable, and estimation method used. Because the outcome variable is not listed in the CLASS statement and the DIST= option is not specified in the MODEL statement, the CAUSALTRT procedure defaults to using the normal distribution and identity link function to fit the outcome model. The frequency of students in each treatment condition is displayed in the "Treatment Profile" table (Figure 3). A "Response Profile" table would be included in the output if the outcome variable was also binary.

Figure 2 Model Information

The CAUSALTRT Procedure

Model Information	
Data Set	WORK.SCHOOL
Distribution	Normal
Link Function	Identity
Estimation Method	AIPW
Treatment Variable	Catholic
Outcome Variable	Math
Number of Bootstrap Samples	1000
Bootstrap Seed	1234

Figure 3 Treatment Profile

Treatment Profile		
Ordered Value	**Catholic**	**Total Frequency**
1	Yes	592
2	No	5079

The AIPW estimates for the potential outcome means and ATE are displayed in the "Analysis of Causal Effect" table (Figure 4). After adjusting for the confounding associations that are caused by the pretreatment characteristics, the estimate of the average causal effect of Catholic schooling on math performance is only 1.7001, which is less than half of the unadjusted difference in average math scores of 3.8949. This result supports the belief that the pretreatment confounders have inflated the causal relationship. Although the ATE estimate is significantly different from 0 at the 0.05 α-level, as indicated by the 95% confidence interval (1.2185, 2.1818) or the p-value (< 0.0001), it is relatively small compared to the potential outcome means. Whether this causal effect shows a meaningful educational enhancement of math performance is left to the judgment of the substantive researchers.

Figure 4 Analysis of Causal Effect

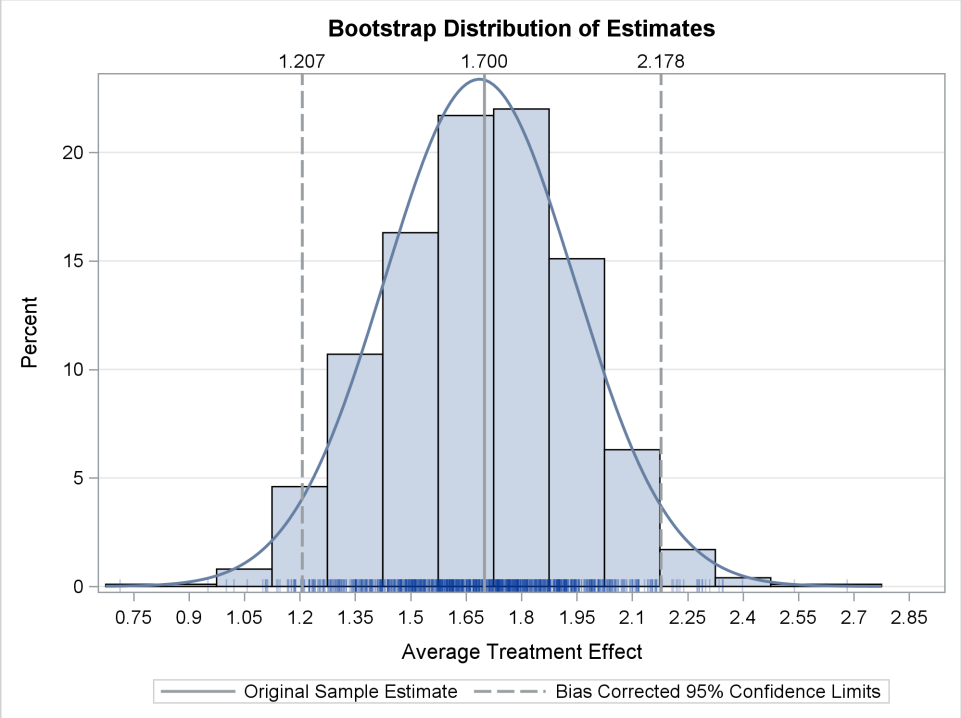
The CAUSALTRT Procedure

Analysis of Causal Effect										
Parameter	**Treatment Level**	**Estimate**	**Robust Std Err**	**Bootstrap Std Err**	**Wald 95% Confidence Limits**		**Bootstrap Bias Corrected 95% Confidence Limits**		**Z**	**Pr > |Z|**
POM	**Yes**	52.5873	0.2560	0.2689	52.0856	53.0890	52.0529	53.1109	205.44	<.0001
POM	**No**	50.8872	0.1280	0.1261	50.6363	51.1382	50.6611	51.1513	397.41	<.0001
ATE		1.7001	0.2457	0.2560	1.2185	2.1816	1.2071	2.1782	6.92	<.0001

The BOOTSTRAP statement requests the computation of bootstrap-based standard errors and confidence limits for the ATE and potential outcome means and adds these estimates to the "Analysis of Causal Effect" table. When you specify the BOOTSTRAP statement, the CAUSALTRT procedure computes bias-corrected confidence intervals from 1,000 bootstrap replications or samples by default. The bootstrap samples are taken from within the treatment conditions, and they include the same number of usable observations in the control and treatment conditions as the number of usable observations that are included in each condition in the input data set. The CAUSALTRT procedure also supports the computation of bootstrap-based confidence limits by using either the percentile method or the normal approximation method. You can control what types of confidence limits are computed and the number of bootstrap samples that are generated by using the BOOTCI and NBOOT= options, respectively, in the BOOTSTRAP statement.

For this example, there is little difference between the analytic and bootstrap-based estimates for the standard errors and confidence limits. You can inspect the distribution of the bootstrap estimates by specifying the PLOTS=HIST option in the BOOTSTRAP statement. When the PLOTS=HIST option is specified, histograms of the bootstrap estimates for the potential outcome means and treatment effect are produced by default and displayed in a panel. The EFFECT suboption specified in this example requests the display of only the bootstrap estimates for the ATE.

Figure 5 Bootstrap Estimates for the ATE



A normal density is overlaid in the histogram of the bootstrap estimates for the ATE in Figure 5. The estimate from each bootstrap sample is marked along the horizontal axis if multiple plots are not grouped together in a panel. When the bias-corrected confidence limits are computed and plots are not grouped together in a panel, the plots also include reference lines for the confidence limits and the original input data set estimate. To replicate the random number stream that was used to generate the bootstrap estimates, you must specify the same value for the SEED= option in the BOOTSTRAP statement and you must ensure that the same number of threads is used for the analytic computations (the number of threads to use can be specified in the NTHREADS= option in the PROC CAUSALTRT statement).

EXAMINING COVARIATE BALANCE AFTER WEIGHTING

One of the important components in AIPW estimation (or other estimation methods that involve inverse probability weighting) is the weighting step, which uses the propensity scores. Because the propensity score used by these methods is also a balancing score (Rosenbaum and Rubin 1983), diagnostics for assessing the propensity score model can be obtained by investigating the balance that a model produces. Essentially, the balancing score property

implies that the weights should create the same type of balance between the pretreatment characteristics of the treatment and control conditions as the type of balance in the data from a randomized experiment.

Therefore, it is important to assess whether balance has been achieved by the weights that are estimated in the propensity score model. A lack of balance would contradict the balancing property of the propensity scores and would call into question the assumption of strong ignorability, which is needed to establish the causal interpretation of the treatment effect.

You can use the COVDIFFPS option in the PROC CAUSALTRT statement to investigate the covariate balance. The COVDIFFPS option computes the weighted and unweighted versions of standardized mean differences (between treatment and control) and the variance ratios (treatment to control) for the covariates that are specified in the PSMODEL statement. These values are displayed in the "Covariate Differences for Propensity Score Model" table in Figure 6.

Figure 6 Covariate Differences

The CAUSALTRT Procedure

Covariate Differences for Propensity Score Model

		Standardized Difference		Variance Ratio	
Parameter		**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
Income	**High**	0.3945	0.0083	1.0255	1.0031
Income	**Low**	-0.3829	0.0095	0.5371	1.0107
Income	**Middle**				
FatherEd	**College**	0.1447	0.0096	1.3284	1.0210
FatherEd	**HighScho**	0.0219	0.0017	1.0173	1.0014
FatherEd	**NoHighSc**	-0.2910	0.0039	0.4817	1.0075
FatherEd	**Postseco**	0.1555	0.0153	1.5045	1.0455
FatherEd	**SomeSeco**	0.0463	-0.0035	1.0713	0.9945
FatherEd	**Unknown**				
MotherEd	**College**	0.0994	0.0258	1.2482	1.0622
MotherEd	**HighScho**	0.0262	-0.0081	1.0138	0.9955
MotherEd	**NoHighSc**	-0.3398	0.0133	0.3802	1.0271
MotherEd	**Postseco**	0.1825	0.0074	1.7304	1.0257
MotherEd	**SomeSeco**	0.0532	0.0133	1.0779	1.0195
MotherEd	**Unknown**				

If the weights that are estimated by the propensity score model improve the balance, then the weighted versions should have standardized mean differences that are closer to 0 and variance ratios that are closer to 1 than the unweighted versions have. For this example, all the propensity score model effects show improved balance after weighting—all the weighted standardized mean differences are less than 0.03 in magnitude, and each weighted variance ratio is closer to 1 than the corresponding unweighted variance ratio.

Notice that rows with blank values are for reference levels of the categorical covariates. The blanks indicate that the values were not computed because they can be expressed in terms of the remaining levels. In addition to the displayed columns, the table includes columns for the weighted and unweighted mean and variance for the covariates within each treatment condition; these columns are not displayed but are accessible if you save the table as an output data set by specifying the ODS OUTPUT statement. Alternatively, you can display these columns by modifying the corresponding template. For more information about working with templates and the Output Delivery System, see Chapter 20, "Using the Output Delivery System" (*SAS/STAT User's Guide*).

Because covariate balance is such an important diagnostic in causal analysis, PROC CAUSALTRT provides many graphical plots to help you assess the balance. Example 3 illustrates some of these graphical capabilities.

PARAMETER ESTIMATES IN TREATMENT AND OUTCOME MODELS

Although the parameter estimates in the treatment and outcome models are usually not the main focus of causal effect estimation, you can still display them by specifying the POUTCOMEMOD and PPSMODEL options, respectively, in the PROC CAUSALTRT statement.

This example specifies the POUTCOMEMOD option in the PROC CAUSALTRT statement to request display of the parameter estimates for the outcome model fit within each treatment condition. Two tables are produced. Figure 7 displays the estimates for the control condition—that is, for the public schools. Figure 8 displays the estimates for the treatment condition—that is, for the Catholic schools.

Figure 7 Control Group Estimates

The CAUSALTRT Procedure

Outcome Model Estimates for Control Group

Parameter		Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		10.9061	0.4772	9.9707	11.8414	522.2403	<.0001
BaseMath		0.7774	0.0080	0.7617	0.7932	9327.3034	<.0001
Income	**High**	0.4244	0.1785	0.0745	0.7743	5.6507	0.0174
Income	**Low**	-0.7194	0.1928	-1.0974	-0.3415	13.9185	0.0002
Income	**Middle**	0	.	.	.	.	.
FatherEd	**College**	0.3857	0.3462	-0.2929	1.0642	1.2410	0.2653
FatherEd	**HighScho**	-0.3112	0.3011	-0.9014	0.2790	1.0681	0.3014
FatherEd	**NoHighSc**	-1.1478	0.3338	-1.8020	-0.4936	11.8247	0.0006
FatherEd	**Postseco**	0.5464	0.3965	-0.2308	1.3236	1.8986	0.1682
FatherEd	**SomeSeco**	0.1948	0.3185	-0.4295	0.8191	0.3739	0.5409
FatherEd	**Unknown**	0	.	.	.	.	.
MotherEd	**College**	0.6255	0.3740	-0.1076	1.3586	2.7969	0.0944
MotherEd	**HighScho**	-0.00768	0.3185	-0.6319	0.6166	0.0006	0.9808
MotherEd	**NoHighSc**	-0.5579	0.3590	-1.2615	0.1456	2.4160	0.1201
MotherEd	**Postseco**	0.1472	0.4367	-0.7088	1.0031	0.1135	0.7361
MotherEd	**SomeSeco**	0.4911	0.3358	-0.1671	1.1492	2.1386	0.1436
MotherEd	**Unknown**	0	.	.	.	.	.

Figure 8 Treatment Group Estimates

The CAUSALTRT Procedure

Outcome Model Estimates for Treatment Group							
Parameter		Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		15.6156	1.5447	12.5880	18.6431	102.1926	<.0001
BaseMath		0.7274	0.0250	0.6784	0.7763	849.1167	<.0001
Income	High	-0.3734	0.4885	-1.3309	0.5840	0.5843	0.4446
Income	Low	-1.9732	0.7340	-3.4118	-0.5346	7.2272	0.0072
Income	Middle	0	.	.	.	.	.
FatherEd	College	2.3426	1.0376	0.3090	4.3763	5.0975	0.0240
FatherEd	HighScho	0.6428	0.9968	-1.3108	2.5965	0.4159	0.5190
FatherEd	NoHighSc	0.0547	1.2184	-2.3333	2.4426	0.0020	0.9642
FatherEd	Postseco	1.2191	1.0877	-0.9127	3.3509	1.2562	0.2624
FatherEd	SomeSeco	1.9739	1.0052	0.00371	3.9440	3.8559	0.0496
FatherEd	Unknown	0	.	.	.	.	.
MotherEd	College	-0.1707	1.0823	-2.2920	1.9507	0.0249	0.8747
MotherEd	HighScho	-1.3311	1.0085	-3.3077	0.6454	1.7423	0.1869
MotherEd	NoHighSc	-0.6978	1.3189	-3.2829	1.8873	0.2799	0.5968
MotherEd	Postseco	-0.8308	1.1153	-3.0168	1.3552	0.5548	0.4563
MotherEd	SomeSeco	-0.8607	1.0030	-2.8265	1.1052	0.7363	0.3908
MotherEd	Unknown	0	.	.	.	.	.

EXAMPLE 3: FAMILY AID AND CHILD DEVELOPMENT

This example further demonstrates the graphical and numerical methods for assessing the causal effects that are estimated by the CAUSALTRT procedure. It also discusses some limitations of the analyses. These limitations are related to practical issues that researchers might often encounter in causal analyses. However, the viewpoints expressed in this example are not intended to represent definitive analyses of the data.

The data for this example are a subset of data from the 1997 Child Development Supplement to the Panel Study of Income Dynamics (Hofferth et al. 2001) in Guo and Fraser (2015). Data for the study were collected nationally from families with children whose ages ranged for newborn to 12. The research question of interest was whether receiving a welfare benefit (represented by the **AFDC** variable) would affect children's development. For this example, a child's development was measured by a score on the age-normalized letter-word identification portion of the Woodcock-Johnson Tests for Achievement (represented by the **Lwi** variable).

The analysis here is restricted to children whose primary caregiver was under the age of 36. This restriction results in a data set, named **Children**, which contains 1,003 records and is created by the following DATA step:

```
data Children;
   input
   Sex Race Age Ratio PcgEd PcgAFDC AFDC Lwi;
   datalines;
   0 0 4  0.6089 12 0 1 81
   0 0 12 0.4113 9  0 1 93
   0 0 12 4.9965 12 0 0 109
   1 0 6  1.0683 11 1 0 74
   1 0 4  1.0683 11 1 0 79
   0 0 4  3.1081 12 1 0 88

   ... more lines ...

   0 1 4  1.5719 11 1 0 108
   1 1 3  1.1919 12 1 0 108
   1 1 3  0.3129 9  0 1 101
   0 1 5  2.3229 12 0 0 79
;
```

The variables in the data set are as follows:

- **AFDC**: indicator for whether the child received support from a public assistance program
- **Age**: age of the child in 1997
- **Lwi**: child's score on the letter-word identification test
- **PcgAFDC**: indicator for whether the child's primary caregiver received support from a public assistance program during any year when the primary caregiver was between the ages of 6 and 12
- **PcgEd**: number of years of schooling for the child's primary caregiver
- **Race**: indicator for whether the child is African-American
- **Ratio**: ratio of family income to the poverty threshold in 1996
- **Sex**: indicator for whether the child is male

The following statements invoke PROC CAUSALTRT to estimate the ATE of receiving a welfare benefit on letter-word identification test scores and to request diagnostics for assessing the model:

```
proc causaltrt data=Children covdiffps nthreads=2;
   class AFDC PcgAFDC Race Sex;
   psmodel AFDC(ref='0') = Sex Race Age PcgEd PcgAFDC/
           plots=(pscovden weightcloud);
   model Lwi = Sex PcgEd Ratio;
   bootstrap seed=1776;
run;
```

Because models for the outcome (**Lwi**) and treatment (**AFDC**) are specified, the AIPW estimation method is used to estimate the ATE. As compared with the POMs (potential outcome means), the estimate for the ATE of –4.5867 in Figure 9 indicates a relatively small decrease in letter-word identification test scores that can be attributed to receiving some welfare assistance. Because the BOOTSTRAP statement is specified, bootstrap-based standard errors and confidence limits are included in Figure 9. The empirical estimate for the ATE standard error of 1.6437 is much less than the bootstrap-based estimate of 2.2470. Moreover, the Wald confidence limits that are computed with the empirical standard error estimate and the bootstrap-based bias-corrected confidence limits differ in their coverage of 0. Hence, the analytic results show a significant negative causal effect, but the bootstrap results shows a nonsignificant causal effect. This inconsistency suggests that either the models are misspecified or there are some irregularities in the distribution, or a combination of both.

Figure 9 Analysis of Causal Effect

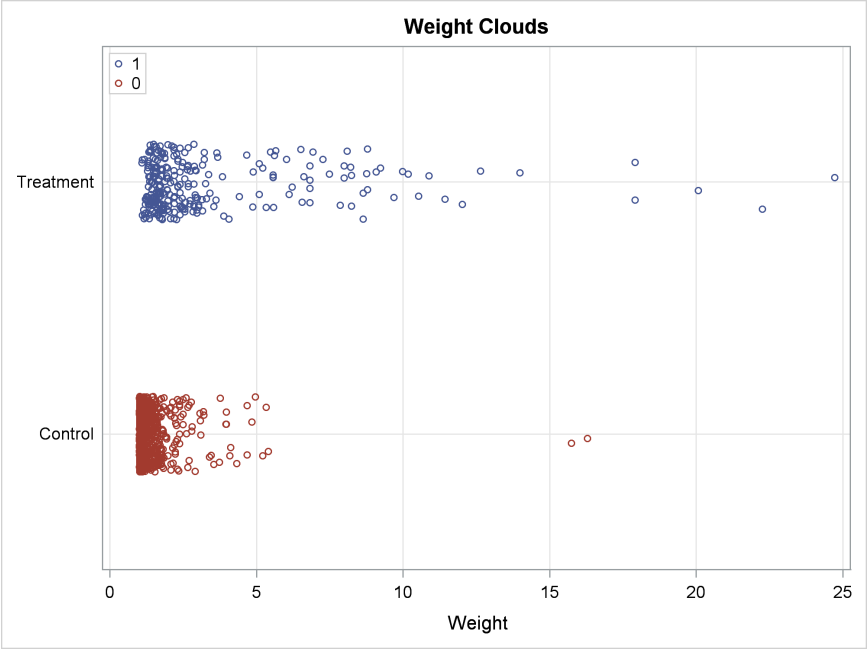
The CAUSALTRT Procedure

Analysis of Causal Effect

Parameter	Treatment Level	Estimate	Robust Std Err	Bootstrap Std Err	Wald 95% Confidence Limits		Bootstrap Bias Corrected 95% Confidence Limits		Z	Pr > |Z|
POM	1	98.5565	1.4458	2.1344	95.7229	101.39	94.5800	103.17	68.17	<.0001
POM	0	103.14	0.8086	0.7919	101.56	104.73	101.74	104.64	127.56	<.0001
ATE		-4.5867	1.6437	2.2470	-7.8082	-1.3652	-8.8993	0.4247	-2.79	0.0053

In order to investigate the weights used for the estimation, the WEIGHTCLOUD suboption in the PLOTS= option produces point clouds that jitter the weights within the treatment and control conditions. The point clouds in Figure 10 show a few observations in each treatment condition that have weights greater than 15. Fortunately, there are no observations with more extreme weights that might have influenced the estimation of the ATE. If there were any observations with a weight greater than 50, PROC CAUSALTRT would add a note to the SAS log. You can change the value that is used to flag the existence of extreme weights by using the WGTFLAG= option in the PSMODEL statement.

Figure 10 Weight Clouds



EXAMINING COVARIATE BALANCE AFTER WEIGHTING

The COVDIFFPS option in the PROC CAUSALTRT statement and the PSCOVDEN suboption in the PLOTS= option in the PSMODEL statement assess the balance induced by the propensity score model for this example. The PSCOVDEN suboption produces density plots for the continuous covariates in the propensity score model. To produce plots for only a subset of the covariates, you can use the EFFECTS suboption. Each plot includes a kernel density estimate of the covariate for the treatment and control condition. Two plots are produced for each covariate: one displays unweighted kernel density estimates, and the other displays kernel density estimates weighted by the inverse probability weights. By default, multiple plots will appear in one output panel, as seen in Figure 12 for this example.

Figure 11 Covariate Differences

The CAUSALTRT Procedure

Covariate Differences for Propensity Score Model

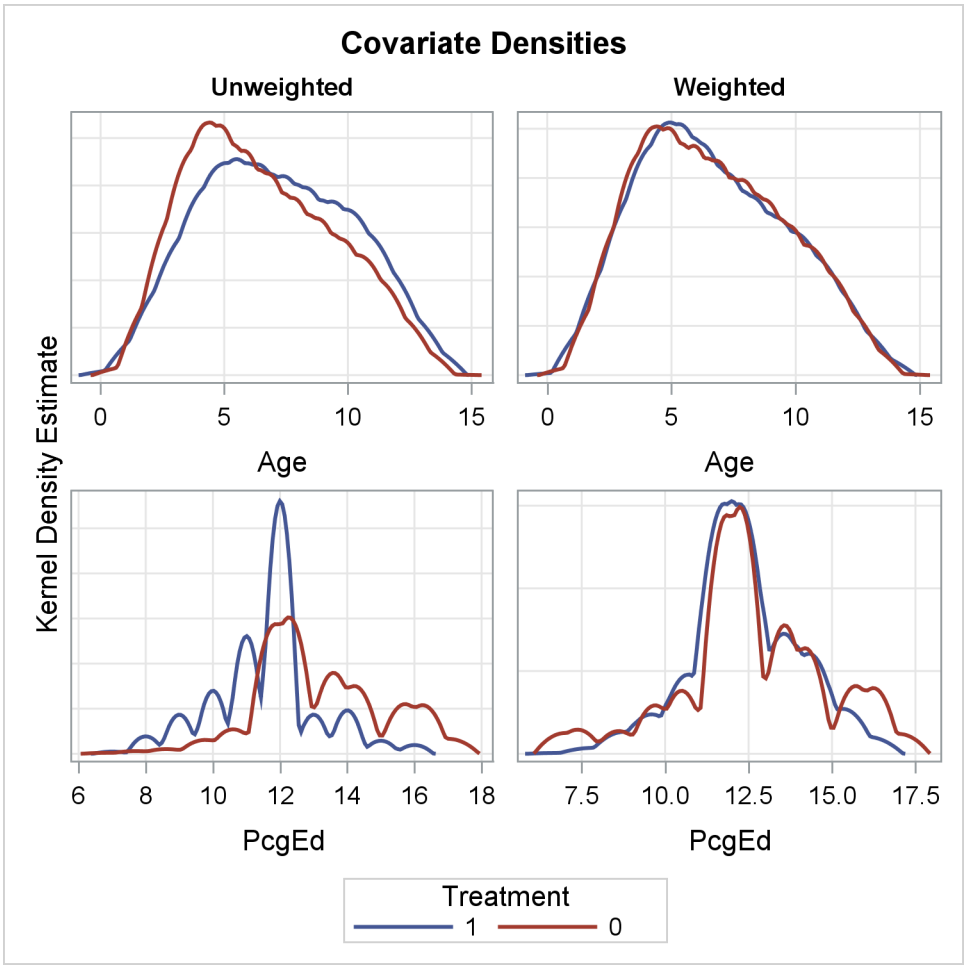
		Standardized Difference		Variance Ratio	
Parameter		**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
Sex	**0**	0.0335	-0.0433	1.0036	0.9936
Sex	**1**				
Race	**0**	-0.9343	-0.0621	0.7404	0.9989
Race	**1**				
Age		0.2196	0.0020	1.0266	0.9650
PcgEd		-0.9067	-0.0974	0.6789	0.5439
PcgAFDC	**0**	-0.6476	-0.0660	1.8658	1.0739
PcgAFDC	**1**				

The densities for the variable **Age** show an improvement in balance after weighting. This improvement is also echoed in the numeric comparison of the standardized mean differences (the weighted version is closer to 0) and variance ratios (the weighted version is closer to 1) for **Age** in Figure 11.

In Figure 12, the densities for the variable **PcgEd** also show an overall improvement in balance after weighting. However, although Figure 11 shows that the corresponding standardized mean difference for **PcgEd** decreases in

magnitude after weighting, the variance ratio decreases from 0.6789 to 0.5439, which is farther away from 1. This lack of balance in variance ratio for **PcgEd** might raise concerns about the appropriateness of the propensity score model.

Figure 12 Kernel Density Estimates



For the categorical variables **Race** and **PcgAFDC**, Figure 11 shows improvement in balance for the weighted version. For the variable **Sex**, both the unweighted and weighted versions show adequate balance.

Because of the lack of balance for the variable **PcgEd** and the nonsignificance of ATE by using the bootstrap-based confidence intervals, the causal treatment effect of **AFDC** on **Lwi** is doubtful.

SOME LIMITATIONS OF THE ANALYSES

It is important to discuss some potential limitations of the causal effect analyses in this example. First, the treatment indicator **AFDC** did not capture the type of welfare program from which the child received a benefit or the duration of the time for which the benefit was received. These limitations might suggest a violation of the "no hidden variations of treatment" component of the stable unit treatment value assumption (SUTVA). Hence, the estimated average treatment effect might not be unambiguously defined. To deal with this problem, it might be useful to study the causal effects separately for subjects within homogeneous welfare programs and within the same benefit durations.

Another concern is that some children's pretreatment characteristics could have disqualified them from certain welfare benefits. For example, children from families with very high socioeconomic status might have a zero probability of receiving welfare aid—a possible violation of the positivity assumption. To deal with this issue, you might want to limit the study of causal effects to the set of children within a certain range of socioeconomic status (low to middle, for example) so that the positivity assumption would be more tenable.

SUMMARY

Estimating the causal effect of a treatment T on an outcome Y from observational data requires the use of a different set of statistical techniques than those that are used for data from randomized control trials. In randomized control trials, you can assume by design that there are no systematic differences between the values of pretreatment variables for subjects in the different treatment conditions. Therefore, the causal effect of T on Y can be directly observed from the data. To estimate the causal effect of T on Y from observational data, additional adjustments must be made to remove the effects of confounding variables that are associated with both T and Y. As described in this paper, the theoretical justification of these adjustments requires an additional set of assumptions to ensure that the causal effect can be properly identified from the data.

The causal inference methods that the CAUSALTRT procedure implements are designed to be used with data from observational studies. These methods adjust for the effects of confounding variables by modeling either the treatment T or the outcome Y, or both. The augmented inverse probability weighting (AIPW) method implemented by PROC CAUSALTRT and illustrated in this paper is a doubly robust estimation method that incorporates modeling of both T and Y. The AIPW estimation method is called doubly robust because it produces unbiased estimates for the potential outcome means and treatment effects even if one of the models is misspecified. In addition to the AIPW estimation method, the CAUSALTRT procedure implements the doubly robust inverse probability weighted regression adjustment (IPWREG) method, along with inverse probability weighting methods (IPW, IPWR, and IPWS) that model only T and regression adjustment (REGADJ) that models only Y.

In addition to the estimation of causal treatment effects, checking the covariate balance (which is induced by the estimated inverse probability weights) is also an important diagnostic step in causal analysis. This paper illustrates some numeric checks that compare standardized mean differences and variance ratios between the unweighted and weighted versions for the samples. Moreover, as illustrated in the last example, graphical plots that are produced by the CAUSALTRT procedure are also useful for assessing the covariate balance.

REFERENCES

Coleman, J. S., Hoffer, T., and Kilgore, S. (1982). *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic Books.

Guo, S., and Fraser, M. W. (2015). *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Hofferth, S., Stafford, F. P., Yeung, W. J., Duncan, G. J., Hill, M. S., Lepkowski, J., and Morgan, J. N. (2001). "Panel Study of Income Dynamics, 1968–1999: Supplemental Files." Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor). http://doi.org/10.3886/ICPSR03202.v3.

Imbens, G. W., and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.

Lunceford, J. K., and Davidian, M. (2004). "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study." *Statistics in Medicine* 23:2937–2960.

Murnane, R. J., and Willett, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. New York: Oxford University Press.

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).

Pierce, D. A. (1982). "The Asymptotic Effect of Substituting Estimators for Parameters in Certain Types of Statistics." *Annals of Statistics* 10:475–478.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data." *Journal of the American Statistical Association* 90:106–121.

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. (2005). “Causal Inference Using Potential Outcomes: Design, Modeling, Decisions.” *Journal of the American Statistical Association* 100:322–331.

Stefanski, L. A., and Boos, D. D. (2002). “The Calculus of M-Estimation.” *American Statistician* 56:29–38.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

ACKNOWLEDGMENTS

The authors are grateful to Anne Baxter for editing this manuscript.

CONTACT INFORMATION

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

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