Paper SAS4322-2020

Causal Effect Estimands: Interpretation, Identification, and Computation

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

Abstract

In modern statistics and data science, there is growing attention on estimating causal effects by using data from nonrandomized or imperfectly randomized studies. This task arises in applications such as post-approval analysis of medical treatments, evaluation of public policies, and assessment of marketing campaign efficacy. One challenge of these applications is the variety of causal effects that you can estimate. For example, you might need to determine whether to estimate the average treatment effect (ATE), the average treatment effect for the treated (ATT), a mediated effect, or other conditional effects. Identifying the causal effect most relevant to your application can have important implications for determining what approach to causal inference is most appropriate. This paper provides an overview of different types of causal estimands, a comparison of how the different estimands are interpreted, and guidance on how identifying an appropriate estimand can help you determine an appropriate causal analysis workflow. The CAUSALGRAPH, CAUSALMED, CAUSALTRT, and PSMATCH procedures in SAS/STAT® software are used to demonstrate the workflow. The paper also includes a review of the assumptions that are required for identifying and estimating causal effects.

Introduction

When you are evaluating public policy or business decisions, a common challenge is to estimate the causal effect of a treatment or intervention. Although data from randomized experiments are widely recognized to be the gold standard for causal effect estimation, there are many research questions that cannot be subjected to experimentation because of practical or ethical constraints. These constraints have spurred the development of a rich and growing body of statistical theory related to estimating causal effects when you are using observational data or data from imperfectly randomized experiments.

One foundational element of causal analysis is the formal definition of different types of causal effects. In terms of statistical inference, these different types of (population) causal effects are the estimands that you want to estimate from data. The formal statement of these estimands can help you refine a research question and improve the clarity of your analysis. In fact, your decision about which estimand to consider has important implications in every step of a causal analysis workflow. First, you must choose an estimand that is appropriate for the research question that sparks the analysis. When the estimand has been determined, the process of selecting and implementing an estimation strategy requires you to reason about the properties of the process that generated the data. Finally, a well-defined estimand is essential to the clear dissemination of the results of your analysis.

This paper provides an overview of different causal estimands and an introduction to the frameworks in which they are defined. In particular, it discusses the average treatment effect (ATE), the average treatment effect for the treated (ATT), and direct and indirect (or mediated) effects. Alongside formal definitions and interpretation, emphasis is placed on causal identification assumptions for each estimand and computational methods that can be used to compute an effect estimate.

The paper is organized as follows. The next section provides an overview of each of these causal estimands and offers some insight into the types of research questions for which each estimand is often useful. The estimands are formally defined in the section "Defining Causal Effects in a Counterfactual Framework." Next, the section "Identification Conditions and Causal Graphs" summarizes the assumptions that are required when you attempt to use observational data to compute an estimate that has a valid causal interpretation. After this formal groundwork has been established, the section "Causal Analysis in SAS/STAT Software" provides a brief overview of the procedures in SAS/STAT 15.1 that are designed to support causal analysis. The remainder of the paper consists of examples that demonstrate these procedures. The paper analyzes a simulated data set and considers questions related to estimating the effect of attending a pre-kindergarten program on students' reading proficiency in eighth grade. Questions that could be investigated by using the different causal estimands are discussed, and examples are presented that demonstrate

the use of appropriate estimation strategies. Although the examples are designed to be realistic, they are intended primarily to illustrate the differences between causal estimands and are not meant to be a definitive discussion on this topic.

Overview of Causal Effect Estimands

Suppose you are interested in studying the effect of enrollment in a pre-kindergarten (pre-K) program on subsequent student performance in eighth grade. There could be a variety of different research questions that might spur your interest in studying this topic. This section discusses a few of these questions.

Perhaps the most common research question is what would be the expected difference in the percentage of students who are reading proficiently in the eighth grade if all students do or do not enroll in a pre-K program. Note that this measure of the causal effect is "universal" for the population of students who are represented by the data. That is, if the sample of students for whom you have data is reasonably representative of the target population, then you can make valid inferences from the result of the sampled students to the entire population.

Suppose instead that data you will use to assess the effect of pre-K enrollment were collected as part of a proof-of-concept program for expanding access to pre-K programs. If your goal is to evaluate the effectiveness of the *current* program, then it is reasonable to focus your analysis only on those students who were actually enrolled in the pre-K program. Alternatively, if you are interested in the possible benefits associated with expanding pre-K access to the currently unserved population, you might want to estimate the average "cost" for those students who did *not* attend a pre-K program. In other words, the group of subjects who receive treatment and the group of subjects who do not are two examples of subgroups that you might be interested in studying as part of your research question.

Now consider the case that your data set contains measurements for an intermediate variable, such as reading proficiency in fourth grade. In this case, your research question might consider the extent to which pre-K enrollment boosts eighth-grade reading proficiency by raising achievement in fourth grade. This is a causal question that involves the decomposition of the treatment effect into the direct effect that the treatment has on the outcome and the indirect or mediated effect that it can have through intermediate (mediator) variables. It can be important to quantify the direct and indirect effects of treatment when you are considering alternative interventions that might affect the outcome. For example, if your estimates suggest that pre-K enrollment improves eighth-grade reading proficiency primarily by raising achievement in fourth grade, then a remedial program that directly boosts performance in fourth grade might be an effective tool to boost eighth-grade reading proficiency, even for those students who were not able to attend pre-K. Alternatively, if this particular mediated effect is estimated to be trivial, policy makers might want to focus on expanding access to pre-K programs or perhaps look for other mediators that can lead to a successful remedial intervention.

In each of these situations, your research question is a *causal* question about the effect of pre-K enrollment on eighth-grade reading proficiency. That is, you are not interested in computing a measure of association (for example, the correlation between pre-kindergarten attendance and eighth-grade reading proficiency). Rather, you would like to know the extent to which enrollment in a pre-kindergarten program *causes* success in eighth-grade reading proficiency. However, each of these questions corresponds to a different notion of a causal effect. The variety of possible notions for a causal effect fuels the need for a formal framework that defines different causal estimands. The following section defines several types of causal effect estimands and discusses the different types of causal questions that each estimand represents.

Defining Causal Effects in a Counterfactual Framework

This paper uses the Neyman-Rubin potential outcomes framework (Rubin 1980, 1990) as the basis for defining causal effects. In this framework, multiple potential outcomes are defined for each subject. Intuitively, the potential outcomes can be thought of as describing an idealized situation where it is possible to observe the outcome for every subject under each possible treatment scenario. That is, each potential outcome corresponds to an outcome that would be observed in a treatment scenario that might or might not be contrary to the observed facts.

For simplicity, consider a binary treatment variable T and a continuous or discrete outcome variable Y. In the pre-K example that this paper uses, the treatment variable is binary and indicates whether or not a student was enrolled in a pre-K program; the outcome variable is also binary and indicates whether a student is reading at grade level at the end of eighth grade. Two potential outcomes are required in order to define the total causal effect of T on Y: the potential outcome in the situation where the student is enrolled in a pre-K program, and the potential outcome in the situation where the student is not enrolled in a pre-K program. In this paper, these two quantities are written as Y_1 and Y_0, respectively.

The total causal effect of treatment for a student is the difference between these two potential outcomes, $Y_1 - Y_0$. Typically, it is not possible to observe both both Y_1 and Y_0 for a single individual. For this reason, the individual treatment effect is seldom the target of an analysis but is instead used as the basis for defining different population-level treatment effects.

The broadest population-level effect is the average treatment effect (ATE). The ATE is defined as the expected value of the individual difference in potential outcomes. Formally, the ATE is defined as

$$\text{ATE} = E[Y_1 - Y_0] = E[Y_1] - E[Y_0]$$

As its name suggests, the ATE tells you the average effect of treatment in the population. In terms of the motivating questions discussed for the pre-K example, the ATE is most relevant when you want to compute an effect estimate that is "universal" in the sense that it summarizes the effect for all members of your target population.

Average treatment effects for subpopulations are defined by conditioning on the variables that identify the subpopulation. Two common subpopulations to consider are those individuals who are in the treatment condition and those individuals who are not in the treatment condition. These subpopulations are the basis for the average treatment effect for the treated (ATT) and the average treatment effect for the untreated (ATU). They are formally defined as

$$\text{ATT} = E[Y_1 \mid T = 1] - E[Y_0 \mid T = 1] \quad \text{and} \quad \text{ATU} = E[Y_1 \mid T = 0] - E[Y_0 \mid T = 0]$$

In terms of the pre-K example, the ATT is the estimand that is most relevant when you want to evaluate the effect of a proof-of-concept pilot study. The ATU is the estimand that is most relevant when you want to estimate the "cost" incurred by students who are not enrolled in a pre-K program. Note that when you switch the treatment condition that is designated as $T = 0$, the ATT and ATU estimates reverse up to a sign change. Hence, discussions about estimation methods typically consider only the ATT because you can use the same approaches to estimate the ATU. For this reason, the remainder of this paper focuses primarily on the ATT.

In order to define mediated effects in the potential outcomes framework, additional notation is required. Assume that the total causal effect consists of two components or pathways:

- a direct effect of T on Y

- a mediated effect in which T affects a mediating variable M, which in turn affects Y

The mediating variable M provides an explanation of how the treatment variable T influences the outcome variable Y. The direct pathway can be interpreted either as a literal direct effect of T on Y or as a summary of all the causal influence of T on Y that does not act through M.

Causal mediation effects can also be defined through a counterfactual framework (Pearl 2001; Robins and Greenland 1992; VanderWeele and Vansteelandt 2009, 2010; Valeri and VanderWeele 2013). Define the counterfactual mediator variables M_1 and M_0 in the same way as the potential outcomes that are defined previously. Now define the additional variables Y_{tm} to be the counterfactual outcomes when an intervention sets the treatment variable to $T = t$ and the mediator variable to $M = m$. When both the treatment and mediator variables are binary, there are four such variables: Y_{00}, Y_{01}, Y_{10}, and Y_{11}. In the pre-K example, the binary mediator variable is an indicator for whether a student is reading at grade level after completing fourth grade.

As with the definition of the ATE, the treatment effect for an individual is the difference between these two potential outcomes—that is, $Y_{1M_1} - Y_{0M_0}$. The only difference between this definition and the definition of the ATE is the additional notation that indicates the presence of the mediator variable. The potential values of the mediator variable (M_1 or M_0) follow "naturally" from the specific treatment conditions in the causal process. Hence, $Y_{1M_1} - Y_{0M_0} = Y_1 - Y_0$ is implicitly assumed.

These new double-indexed potential outcomes facilitate the definition of some mediated effects. The natural direct effect (NDE) is defined as

$$\text{NDE} = Y_{1M_0} - Y_{0M_0}$$

The NDE describes the effect of treatment when a hypothetical intervention sets the mediator variable to the value that would have been observed under the control condition—that is, $M = M_0$. Thus, this is a concept about the treatment effect without mediation.

The natural indirect effect (NIE) is then defined as

$$\text{NIE} = Y_{1M_1} - Y_{1M_0}$$

The NIE is the difference between counterfactual outcomes at the two mediator levels, M_1 and M_0, when an intervention sets the treatment variable to $T = 1$. Thus, this is a concept about the mediation effect when the treatment is applied.

Notice that the "natural" designation in the definitions of the NDE and NIE refers to the fact that the potential mediator value M_1 or M_0 follows naturally from the respective treatment condition rather than being a particular fixed value. A corollary is that M_1 and M_0 could be realized at different values for different individuals. For the purpose of this paper, however, the "natural" designation is dropped for brevity.

As intuition suggests, the NDE and NIE sum up to the total effect $Y_{1M_1} - Y_{0M_0}$, and this forms the basis of effect decomposition and the associated mediation measure. For example, when the NIE and NDE have the same sign, the "percentage mediated" can be defined as

$$\text{Percentage mediated} = \text{NIE}/(\text{NIE} + \text{NDE}) \times 100\%$$

If the NIE and NDE have different signs, you can still interpret the NDE and NIE effects, which have different directions, but you cannot meaningfully interpret the percentage mediated.

It is important to note that assumptions are required to ensure the validity of this approach to defining potential outcomes and treatment effects. For total effects, these assumptions typically take the form of the *stable unit treatment value assumption* (SUTVA). The component of SUTVA that ensures well-defined potential outcomes is the assumption that there are no hidden levels of treatment. This would not be the case, for example, if there were different types of pre-kindergarten programs and there were substantive differences between these programs in terms of their effects on the outcome variable. The second component of SUTVA is that there is no interference between subjects. This means that the treatment and potential outcomes for individuals do not depend on one another. In the context of the pre-K example, this condition could be violated if positions in a public program were to be allocated with a fixed number of students from each school district. In this case, the enrollment status of one student necessarily depends on the enrollment status of all other students in the district. For more information, see Imbens and Rubin (2015).

Identification Conditions and Causal Graphs

Potential outcomes provide a framework for the clear definition of causal effects. But for the definitions to be useful, the quantities must be connected to actual observed data. In general, you cannot observe a subject's outcome under multiple treatment scenarios. Instead, for each subject, you typically observe a single outcome Y under *one* treatment assignment. To connect the observed outcome to the potential outcomes, you need to make the *consistency* assumption:

$$Y = Y_t \quad \text{if} \quad T = t$$

To put this into words, the consistency assumption states that the observed outcome equals the potential outcome corresponding to the actual treatment level t that is received by the subject. For a binary treatment, this still leaves the potential outcome Y_{1-t} unobserved for each subject. Therefore, the consistency assumption itself is not sufficient for determining whether it is possible to obtain an unbiased estimate for the causal estimands that are defined in the section "Defining Causal Effects in a Counterfactual Framework."

Reasoning about whether an estimation method supports the valid estimation of a causal effect requires careful consideration of properties of the estimation method and properties of the data generating process. To illustrate, first consider an idealized situation where you can conduct a randomized controlled trial to study the effect of a pre-K program on students' reading proficiency upon their completing eighth grade. In the design stage of your trial, you need to clearly define the student population of interest and what type of pre-K program you want to study. The answers to these questions help ensure that the study captures a "well-defined intervention" (Hernán and Taubman 2008) and that the SUTVA conditions are plausible. Next, you would recruit students, randomly assign each student to either attend (treated or exposed group) or not attend (control group) a pre-K program, and track their academic performance.

In this design with complete external randomization, it would be plausible to assume that $Y_t \perp\!\!\!\perp T$. Following the consistency assumption and the independence assumption,

$$E[Y \mid T = t] = E[Y_t \mid T = t] = E[Y_t]$$

These equalities imply that for this study design, the ATE, ATT, and ATU are all identical and can be directly identified from the observed data. For example, for a binary treatment, this means that $E[Y_t]$ is identified by estimating $E[Y \mid T = t]$ from the sample means of treatment groups for $t = 1$ and $t = 0$.

Unfortunately, as noted in the introduction to this paper, there are many important causal questions that cannot be investigated by conducting a fully randomized experiment. That is likely to be the case for the pre-K enrollment example. Families might not be willing to make their enrollment decisions on the basis of external randomization. More important, there might be ethical issues associated with limiting pre-K access in the control group. Given these limitations, it is often necessary to use observational or retrospective data.

Often, background factors contribute to the parents' decision about whether to enroll their children in a pre-K program. For example, household income levels of the treated (enrolled) and untreated groups could be very different. Moreover, some of these background factors might play a causal role in determining the values of both the treatment and outcome variables. For example, wealthier families might be more likely to enroll their children in a pre-K program, and children from wealthier families also have better academic performance for reasons other than pre-K attendance. These factors are confounding factors or confounding covariates that could obscure the causal effects you would like to estimate. In order to estimate an average causal effect, you must account for these confounding covariates.

A key assumption is that there exists a collection of *observed* variables (or covariates) X such that, within each stratum of these variables, it is as if the treatment assignment were randomized. This "as good as randomized" assumption is often referred to as *conditional ignorability* or *no unmeasured confounders*. It is often written as a conditional independence statement for the potential outcomes:

$$Y_t \perp\!\!\!\perp T \mid X$$

The set of variables X is often referred to as an *adjustment set*, because you must adjust for (or stratify by) the values of these variables. However, it is not at all clear from the definition of conditional ignorability how you should choose the set of variables X that satisfy the conditional independence assumption. One intuitive and often-used strategy is to include all *related* background characteristics of subjects in the observational study. Yet this strategy involves some degree of subjectivity about the criterion of all related background characteristics.

An increasingly popular approach is to use a causal diagram analysis. Briefly, a causal diagram is a collection of variables (including unobserved variables and latent constructs) that are assumed to play a role in the process that generated the observed data. A directed edge that links two variables indicates the possible presence of a causal effect of one variable on the other variable. The process of constructing a reasonable causal diagram for representing a data generating process requires theories, substantive knowledge, and expert opinions, and completing it is not an easy task. But once a causal diagram is constructed and the treatment and outcome variables are specified, it is relatively easy to determine exactly which variables must be included in the adjustment set X in order to satisfy the conditional ignorability assumption. For more information about the use of causal diagrams, see Elwert (2013).

Using a causal diagram has two major benefits. First, it forces you to write down exactly your assumptions about the data generating process. The diagram then provides a transparent representation of your assumptions so that it facilitates communication with others who are interested in your analysis. Second, once a causal diagram is agreed upon by stakeholders, the rules of causal diagrams, called the do-calculus (Pearl 2009), provide an algorithmic tool to identify the set X. The "Examples" section demonstrates causal diagram analysis by using the CAUSALGRAPH procedure and describes how the adjustment set of covariates can be used to estimate the target causal estimands.

One final assumption is needed in order to estimate a causal effect by using observational data: *positivity*. Intuitively, positivity states that every subject has a nonzero probability of being assigned to each treatment condition. Together with conditional ignorability, the positivity assumption enables you to interpret your observational data as if they were collected from a randomized experiment within each stratum of the adjustment variables X. For more information about positivity, see Hernán and Robins (2019).

Although the identification properties discussed in this section focus on the estimation of total effects, similar conditions exist when you estimate causal mediation effects, including direct and indirect effects. For details, see Example 3.

Causal Analysis in SAS/STAT Software

In SAS/STAT 15.1, there are four procedures that are designed primarily to support causal analysis and causal effect estimation. This section provides a brief overview of these procedures and describes how you can use them to estimate the causal effects that are described in the section "Defining Causal Effects in a Counterfactual Framework."

PROC CAUSALGRAPH

The CAUSALGRAPH procedure provides tools for analyzing graphical causal models. When you submit a directed acyclic graph that represents qualitative causal assumptions about a data generating process, the procedure examines the structure of that graph to find statistical strategies that you can use to obtain an unbiased estimate of a causal effect. PROC CAUSALGRAPH does not perform the estimation. Rather, it tells you which variables to include in an analysis (for example, which variables to use to build a propensity score model). It supports one or more treatment variables and one or more outcome variables. It also includes support for quantities (such as latent constructs) that are hypothesized to play a role in the data generating process but are not observed. For more information about PROC CAUSALGRAPH, see the SAS/STAT documentation (SAS Institute Inc. 2018) and Thompson (2019).

PROC CAUSALMED

The CAUSALMED procedure supports the estimation of various causal mediation effects by using a regression adjustment approach (Valeri and VanderWeele 2013). Specifically, it estimates the total effect, NDE, and NIE, all of which are covered in the earlier sections of this paper. It also estimates many other types of causal mediation effects and decompositions. A limited set of generalized linear models for the mediator or the outcome is supported. The treatment, mediator, outcome, and covariates can be either binary or continuous. For more information about PROC CAUSALMED, see the SAS/STAT documentation (SAS Institute Inc. 2018) and Yung, Lamm, and Zhang (2018). See Example 3 for an illustration of the CAUSALMED procedure.

PROC CAUSALTRT

The CAUSALTRT procedure supports the direct estimation of the ATE and ATT from observational data. The procedure supports estimation methods that adjust for confounding through modeling of the treatment, modeling of the outcome, or both. Modeling of the treatment leads to estimation of the ATE or ATT by inverse probability of treatment weighting (IPTW). Modeling of the outcome leads to estimation of the ATE or ATT by regression adjustment. Combined modeling of the treatment and outcome leads to doubly robust methods of estimating the ATE. For more information about PROC CAUSALTRT and the methods that it supports, see the SAS/STAT documentation (SAS Institute Inc. 2018) and Lamm and Yung (2017). Example 1 shows you how to use PROC CAUSALTRT to estimate the ATE, and Example 2 shows you how to use PROC CAUSALTRT to estimate the ATT.

PROC PSMATCH

The PSMATCH procedure provides tools for performing propensity score analysis. It supports the estimation of total effects for a binary treatment T. PROC PSMATCH does not support the direct estimation of a treatment effect. Instead it takes an input observational data set and, through the use of either stratification, matching, or inverse probability weighting, creates an adjusted output data set that is suitable for subsequent outcome analysis of causal effects. The procedure also supports a number of graphical and numerical diagnostics that you can use to inspect properties of the output data set and the adjustment it represents. For more information about PROC PSMATCH and the methods that it supports, see the SAS/STAT documentation (SAS Institute Inc. 2018), Yuan, Yung, and Stokes (2017), and Lamm, Thompson, and Yung (2019). Example 1 shows you how to use PROC PSMATCH to estimate the ATE, and Example 2 shows you how to use PROC PSMATCH to estimate the ATT.

Examples

This section presents three examples that demonstrate how you can use SAS procedures to estimate the different types of causal estimands that are described in the section "Defining Causal Effects in a Counterfactual Framework." The data that are analyzed in these examples are simulated to provide a simplified but realistic application.

The data generating process that the simulation uses is represented graphically in Figure 1. In this diagram, the effect of pre-kindergarten enrollment (**PreK**) on reading proficiency at the end of eighth grade (**MidProf**) is mediated through reading proficiency at the end of fourth grade (**ElmProf**). Other factors that are assumed to play a role in the data generating process are as follows:

- **ElmSize**: average fourth-grade ("elementary school") class size

- **ESL**: indicator for English as a second language

- **IncRatio**: ratio of household income to the federal poverty line

- **MidSize**: average eighth-grade ("middle school") class size

- **PCGEd**: classification variable for primary caregiver's education

- **Values**: latent construct representing the value that the primary caregiver placed on education

Figure 1 Causal Diagram for the Simulated Data Set
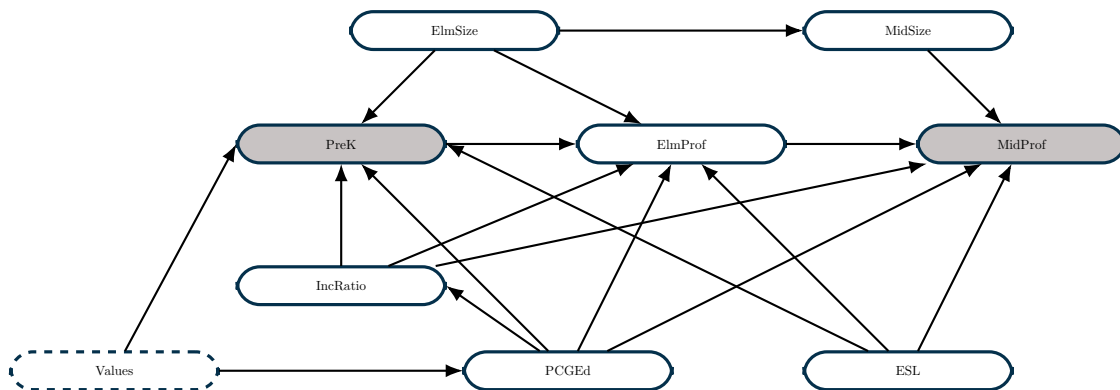
Note that in this data generating process, the variable **Values** is a latent construct and so is not observed. When you analyze observational data, it is important to consider the possible effects that unobserved variables might have in the data generating process. If you fail to properly account for unobserved variables, your analysis might produce biased effect estimates as a result of unobserved confounding.

The simulated data set contains 10,000 observations. Figure 2 displays the first 10 observations in the data set. In the simulation, approximately 41% of students enrolled in a pre-K program.

Figure 2 First 10 Observations in the Simulated Data Set

Obs	ESL	PCGEd	ElmSize	MidSize	IncRatio	PreK	ElmProf	MidProf
1	0	4	19.6	26.3	3.3	0	1	1
2	0	5	19.2	26.1	4.7	1	0	1
3	0	4	18.7	26.1	3.3	1	0	1
4	1	4	17.1	23.4	3.3	0	1	1
5	0	2	17.5	24.5	2.2	0	0	1
6	0	3	17.6	23.8	0.4	0	0	1
7	0	3	19.5	26.0	0.3	1	1	1
8	0	2	19.2	26.4	3.4	1	0	1
9	0	5	19.7	27.0	4.1	1	1	1
10	0	1	20.4	27.3	0.8	0	0	1

The true values of the ATE and ATT for this example are about 0.0369 and 0.0356, respectively. These effects are expressed on the risk difference scale. For example, this means that the difference between the percentage of students who attain eighth-grade reading proficiency in the treatment group (those who were enrolled in pre-K) and the percentage of students in the no-treatment group (those who were not enrolled in pre-K) is 3.69 percentage points. For the ATE (the total effect), this difference consists of two parts: NDE=0.0261 for the direct effect and NIE=0.0108 for the indirect effect. Thus the percentage mediated is about 29%. In the examples that follow, the goal is to use the sample of available data in order to estimate these quantities. Estimation of the ATE and ATT from the simulated data is demonstrated in Example 1 and Example 2, respectively. Estimation of the NDE and NIE from the simulated data is demonstrated in Example 3.

In a real-world analysis, you should consider whether all necessary identification assumptions are satisfied before attempting to estimate any causal effects. These identification assumptions are described in the section "Identification Conditions and Causal Graphs." For this example, you can assume that both the SUTVA and positivity assumptions hold. For the conditional ignorability assumption, you can use PROC CAUSALGRAPH to analyze the graphical model in Figure 1 and find an adjustment set (the set X that satisfies the conditional ignorability assumption). The following code finds all valid adjustment sets for the causal diagram in Figure 1:

```
proc causalgraph;
   model "ReadingProf"
      ElmProf => MidProf,
      ElmSize => ElmProf MidSize PreK,
      ESL IncRatio => ElmProf MidProf PreK,
      MidSize => MidProf,
      PCGEd => ElmProf IncRatio MidProf PreK,
      PreK => ElmProf,
      Values => PCGEd PreK;
   latent Values;
   identify PreK => MidProf;
run;
```

Because this is a simulated data example with a known data generating process, the causal diagram that represents the data generating process is known exactly and is given in Figure 1. In practice, reasoning about the data generating process and proposing a causal diagram can be a very challenging task that needs to integrate theory, substantive knowledge, and expert opinions. For more information about using expert judgment to form a causal diagram, see Hanea et al. (2018).

The result of the PROC CAUSALGRAPH analysis is shown in Figure 3. There are two possible sets of variables that are sufficient to meet the conditional ignorability assumption. The first set contains the variables **ElmSize**, **ESL**, **IncRatio**, and **PCGEd**. The second set contains these four variables plus the variable **MidSize**.

Figure 3 Valid Adjustment for the Total Effect of **PreK** on **MidProf**

	Size	Minimal	ElmProf	ElmSize	ESL	IncRatio	MidSize	PCGEd
Covariate Adjustment Sets for ReadingProf								
Causal Effect of PreK on MidProf								
1	4	Yes		*	*	*		*
2	5	No		*	*	*	*	*

It is important to note that the variable **ElmProf** is not included in either of the adjustment sets. Including this mediator variable in a predictive model for the outcome variable **MidProf** might be beneficial, but including the mediator variable in the adjustment set can severely bias the estimation of the total causal effect. The reason is that the mediation itself is part of the causal mechanism and therefore should not be included in any adjustment. See Example 1 for a demonstration of the biasing effect of adjusting for the mediator variable **ElmProf** when you are estimating the ATE in the pre-K example.

Example 1: Estimating an Average Treatment Effect

This example estimates the ATE from the simulated data set described in the preceding section. As discussed in the sections "Overview of Causal Effect Estimands" and "Defining Causal Effects in a Counterfactual Framework," the ATE would be a relevant estimand if you are interested in studying what the average effect of enrollment in a pre-K program on eighth-grade reading proficiency would be in the entire population. The causal estimand here is about a "universal" pre-K program for this population.

When you estimate a causal effect, it is important to make sure that the data used for estimation are representative of your target population. Because the ATE is an estimate of the effect of treatment in an entire population, you need to make sure that the entire population is still represented in the data used to obtain an effect estimate. For this reason, if you were to use matching as the adjustment method, a full matching would be appropriate (Stuart 2010). However, for these data, a full matching that results in good balance and a proper adjustment for the confounders is likely to be difficult. In the example data, about 41% of the students are in the treatment condition. Because the treated and untreated groups are of somewhat similar size, a full matching that you perform using PROC PSMATCH is likely to produce many matched groups that contain equal numbers of treated and untreated students. This means that there will be only a limited adjustment for the confounders. For this reason, a full matching is not demonstrated in this example.

Instead of using matching as an adjustment method, you could consider using inverse probability of treatment weighting (IPTW). The following statements use PROC PSMATCH to compute the ATE inverse probability of treatment weights and to create a graphical assessment of the balance that these weights produce:

```
proc psmatch data=ReadingObs;
   class ESL PCGEd PreK;
   psmodel PreK(treated="1") = ElmSize ESL IncRatio PCGEd;
   psweight weight=atewgt;
   assess ps var=(IncRatio ElmSize MidSize) /
          plots(nodetails)=(stddiff box(display=(PS IncRatio)) wgtcloud);
   output out=ATEWgtData weight=_atewgt_;
run;
```

In this example, PROC PSMATCH fits the propensity score model internally by using logistic regression. The propensity score model is specified in the PSMODEL statement. Note that this PSMODEL statement uses the four variables **ElmSize**, **ESL**, **IncRatio**, and **PCGEd** that were output by PROC CAUSALGRAPH as a valid adjustment set. PROC PSMATCH can also use precomputed propensity scores that are specified in a PSDATA statement. For an example of using PROC PSMATCH with precomputed propensity scores, see Example 98.8 in the SAS/STAT 15.1 documentation for PROC PSMATCH (SAS Institute Inc. 2018) or Lamm, Thompson, and Yung (2019). The WEIGHT=ATEWGT option in the PSWEIGHT statement indicates that the ATE weights should be used in the balance diagnostics and as the weights that are added to the output data set that is specified in the OUTPUT statement.[1]

---

[1] The WEIGHT= option and the PSWEIGHT statement are new in the 15.1 release of SAS/STAT software. In previous releases, the same functionality exists but uses different syntax in the ASSESS and OUTPUT statements.

The ASSESS statement produces a number of graphical and numerical assessments of the balance that is achieved using the ATE weights. These diagnostics can help you assess the appropriateness of the propensity score model and the adjustment for the confounders. Figure 4 shows the standardized mean differences plot for the continuous and binary variables that are specified in the ASSESS statement. This plot provides a compact visual summary of the balance achieved in the mean differences across multiple variables. For this example, good balance is seen for all the variables.

Figure 4 Standardized Mean Difference Plot for IPTW-ATE Analysis
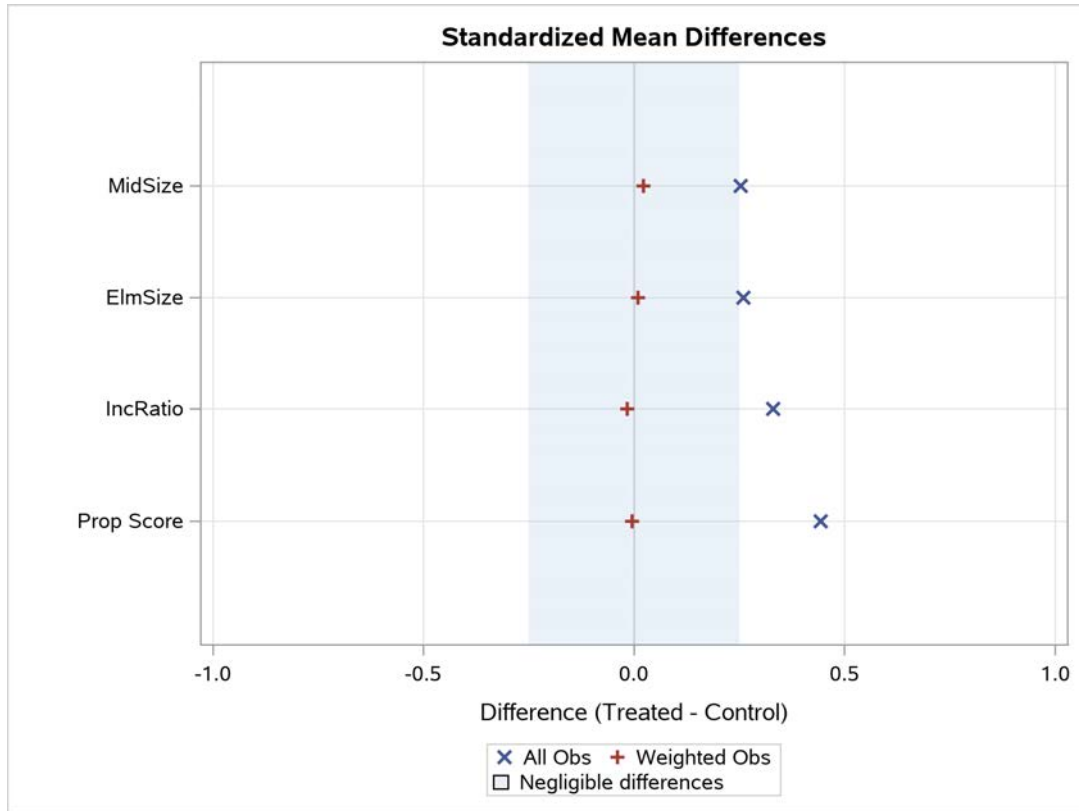
To further inspect the distribution of the confounders after adjusting for inverse probability of treatment weights, you can examine the box plots shown in Figure 5.

Figure 5 Box Plots for IPTW-ATE Analysis
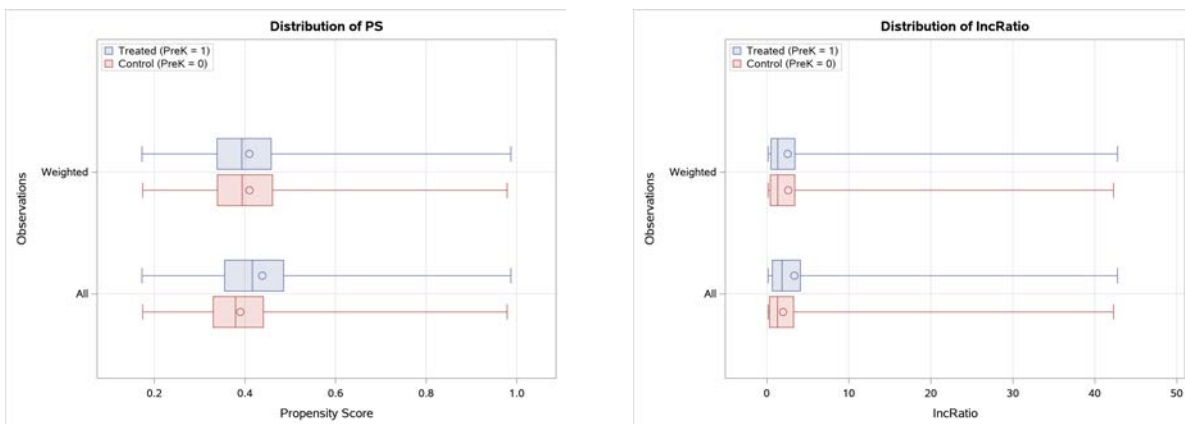
These plots show good balance in the distribution of the propensity scores and the variable **IncRatio** after the procedure incorporates the inverse probability of treatment ATE weights. In particular, there is an improvement in the balance in the left tail of the distribution of the variable **IncRatio** after the weights are incorporated.

One concern in using weighting is that there might be observations that have overly large and influential weights. Figure 6 shows a weight cloud, or jitter plot, of the weights. This figure includes reference lines for identifying possibly extreme weights. By default, these reference lines are placed at 10 times the expected weight for an observation in each treatment condition. There are two observations in the control condition whose weights exceed the recommended threshold. Upon further inspection, these two observations correspond to students in families with very high incomes. For this analysis, these two students and their large inverse probability of treatment weights are not considered problematic and the analysis will proceed.

Figure 6 ATE Inverse Probability of Treatment Weights
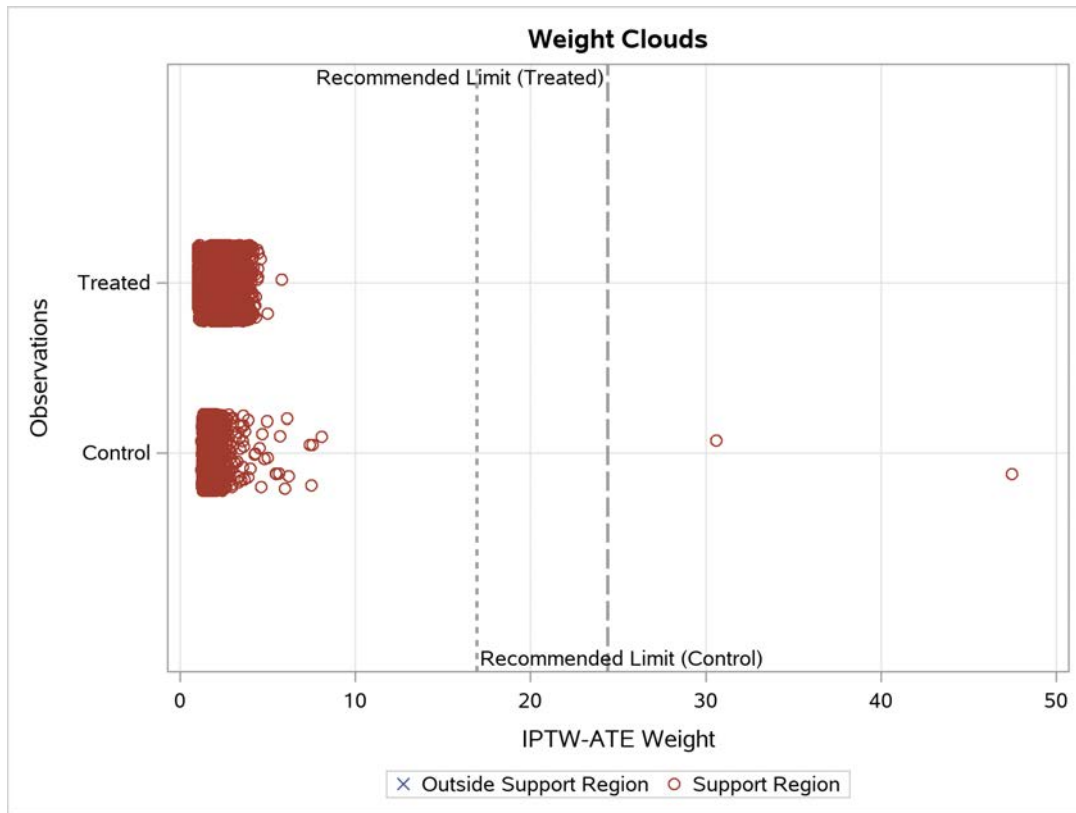
After you have determined that the weights computed by PROC PSMATCH produce sufficient balance, you can perform an outcome analysis by using the output data set **ATEWgtData**. This output data set stores the inverse probability weights in the variable **_atewgt_**. The following code uses this data set in PROC TTEST to compute the ATE:

```
proc ttest data=ATEWgtData;
   class PreK;
   var MidProf;
   weight _atewgt_;
run;
```

The output from PROC TTEST is shown in Figure 7. The estimated ATE of 0.0325 is consistent with the known true ATE of 0.0369.

Figure 7 ATE Estimate Obtained by Inverse Probability of Treatment Weighting

PreK	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.7529	0.7419	0.7640	0.5616	0.5517	0.5719
1		0.7854	0.7728	0.7980	0.6415	0.6279	0.6557
Diff (1-2)	Pooled	-0.0325	-0.0490	-0.0159	0.5957	0.5875	0.6040
Diff (1-2)	Satterthwaite	-0.0325	-0.0492	-0.0157			

So far, this example demonstrates a two-step approach to estimating the ATE. In the first step, you use PROC PSMATCH to examine the balance produced by the inverse probability of treatment weights. In the second step, those weights are incorporated into an outcome analysis that is performed outside PROC PSMATCH. In this approach, the weights that are used in the second step are treated as known values, and there is no adjustment for their estimation from the data.

An alternative approach to the analysis would be to use PROC CAUSALTRT to directly estimate the ATE. PROC CAUSALTRT also supports estimation of the ATE by inverse probability of treatment weighting. Because PROC CAUSALTRT fits the propensity score model and estimates the treatment effect internally, it is able to incorporate the estimation of the weights from the data into the relevant standard error computations.

The CAUSALTRT procedure also supports estimation of the ATE by doubly robust estimation methods that incorporate both modeling of the treatment and modeling of the outcome variable. An estimation method is said to be doubly robust if it produces an unbiased estimate so long as one of the two models that it uses, but not necessarily both, is correctly specified. The following syntax uses PROC CAUSALTRT to estimate the ATE by the doubly robust augmented inverse probability weighting (AIPW) estimation method:

```
proc causaltrt data=ReadingObs;
   class ESL MidProf PCGEd PreK / desc;
   psmodel PreK = ElmSize ESL IncRatio PCGEd;
   model MidProf = ElmSize ESL IncRatio PCGEd;
   bootstrap seed=1976;
run;
```

When you use PROC CAUSALTRT, the outcome variable must always be specified in the MODEL statement and the treatment variable must always be specified in the PSMODEL statement. In these statements, you also provide any additional model specifications that you need for the estimation method. The BOOTSTRAP statement is specified to request bootstrap-based estimation of standard errors and confidence limits for the potential outcome means and treatment effect estimate. For more information about PROC CAUSALTRT and the estimation methods that it supports, see the documentation for PROC CAUSALTRT (SAS Institute Inc. 2018) and Lamm and Yung (2017).

Figure 8 shows the "Analysis of Causal Effect" table that PROC CAUSALTRT produces. The ATE estimate of 0.0327 is close to the true value of 0.0369 and similar to the estimate obtained by using PROC PSMATCH and PROC TTEST. Moreover, the empirical and bootstrap-based confidence limits for the ATE and potential outcome means are similar.

Figure 8 AIPW Estimate for the ATE

							Bootstrap Bias Corrected 95% Confidence Limits			
	Treatment		Robust	Bootstrap	Wald 95%					
Parameter	Level	Estimate	Std Err	Std Err	Confidence Limits		Limits		Z	Pr > |Z|
POM	1	0.7855	0.00670	0.00646	0.7723	0.7986	0.7718	0.7979	117.18	<.0001
POM	0	0.7528	0.00562	0.00574	0.7417	0.7638	0.7419	0.7644	133.90	<.0001
ATE		0.03270	0.00872	0.00873	0.01561	0.04980	0.01487	0.04813	3.75	0.0002

To demonstrate the importance of using a valid adjustment set, the following code uses the AIPW estimation method, with the mediator variable **ElmProf** included in the MODEL statement:

```
proc causaltrt data=ReadingObs;
   class ESL MidProf PCGEd PreK / desc;
   psmodel PreK = ElmSize ESL IncRatio PCGEd;
   model MidProf = ElmSize ESL IncRatio PCGEd ElmProf;
   bootstrap seed=1976;
run;
```

Including the mediator variable in the adjustment set results in a poor ATE estimate of 0.02429, which underestimates the true ATE value of 0.0369 (Figure 9). Although including the mediator variable in the model for the outcome might increase the predictive accuracy of the outcome model, doing so produces a biased estimate of the ATE, as this example demonstrates. The inclusion of the mediator variable is ruinous for the model's causal properties because the mediating pathway is part of the treatment's total effect.

Figure 9 Incorrect ATE Estimate in Adjusting for a Mediator

							Bootstrap Bias Corrected 95% Confidence Limits			
Parameter	Treatment Level	Estimate	Robust Std Err	Bootstrap Std Err	Wald 95% Confidence Limits				Z	Pr > |Z|
POM	1	0.7804	0.00659	0.00653	0.7675	0.7934	0.7671	0.7937	118.36	<.0001
POM	0	0.7562	0.00557	0.00559	0.7452	0.7671	0.7456	0.7674	135.74	<.0001
ATE		0.02429	0.00851	0.00856	0.007615	0.04097	0.006470	0.03973	2.86	0.0043

In causal mediation analysis, some very special types of mediation effects might require you to adjust for the mediator. Even so, the computation of the total effect in causal mediation analysis never adjusts for the mediator variable. Example 3 provides an example that illustrates the use of PROC CAUSALMED to perform causal mediation analysis. These different considerations of how to handle a mediator variable when you are estimating a total effect versus when you are performing causal mediation analysis illustrate the importance of using estimation methods that are appropriate to your target estimand and data generating process.

Finally, the importance of adjusting for the confounding covariates is illustrated by the poor estimate of the ATE obtained for the following PROC TTEST code, which compares the raw difference in the outcome between the treatment conditions (Figure 10):

```
proc ttest data=ReadingObs;
   class  PreK ;
   var MidProf;
run;
```

Figure 10 Poor ATE Estimate without Adjustment

PreK	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.7450	0.7339	0.7561	0.4359	0.4282	0.4439
1		0.7987	0.7864	0.8110	0.4010	0.3925	0.4099
Diff (1-2)	Pooled	-0.0537	-0.0705	-0.0369	0.4220	0.4162	0.4279
Diff (1-2)	Satterthwaite	-0.0537	-0.0702	-0.0371			

Example 2: Estimating an Average Treatment Effect for the Treated

This section considers the estimation of the ATT for the simulated example. When you estimate the ATT, the target population of interest is those students who received the treatment—that is, those students whose families enrolled them in a pre-K program. As discussed in the sections "Overview of Causal Effect Estimands" and "Defining Causal Effects in a Counterfactual Framework," the ATT is interpreted as the average effect of treatment for those who receive the treatment. For this pre-K example, the ATT might be the relevant estimand if the data are being used to evaluate the effectiveness of a pilot program for providing pre-K access.

Matching is a popular method of estimating the ATT, and there are many different approaches you can use to formulate the matching problem. For more 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. The following statements use PROC PSMATCH to perform an optimal one-to-one matching:

```
proc psmatch data=ReadingObs region=treated;
   class ESL PCGEd PreK;
   psmodel PreK(treated="1") = ElmSize ESL IncRatio PCGEd;
   match method = optimal caliper=.;
   assess ps var=(ElmSize IncRatio MidSize) /
         plots(nodetails)=(stddiff box(display=(PS IncRatio)));
   output out=psATTMatchData weight=attWgt;
run;
```

Just like when you are estimating the ATE in Example 1, you use the PSMODEL statement to specify the treatment variable and the variables to be used in the propensity score model. For this example, the four variables that are identified by PROC CAUSALGRAPH are used: **ElmSize**, **ESL**, **IncRatio**, and **PCGEd**. The MATCH statement in PROC PSMATCH requests the use of matching and is used to formulate the matching problem. The CALIPER=. option requests that the matching be performed without a caliper constraint, and the REGION=TREATED option in the PROC PSMATCH statement requests the use of the extended treated group as the support region. When you estimate the ATT, the combination of the extended treated group support region and an optimal matching method ensures that all the treated subjects are included in the matched data set, and therefore the matched data remain representative of the target population. For more information about the different matching methods that PROC PSMATCH supports, see the SAS/STAT documentation (SAS Institute Inc. 2018), Yuan, Yung, and Stokes (2017), and Lamm, Thompson, and Yung (2019).

As in Example 1, the ASSESS statement is used to request diagnostic plots to inspect the balance that is achieved after matching. Note that when the input data set is large, as in this example, performing the matching and producing the graphical diagnostics can be quite resource-intensive. The NODETAILS option and DISPLAY= options in the plot requests are included to control the number of plots produced.

Figure 11 shows the standardized mean differences plot for the data after matching. The plot shows an improvement in the balance for the mean differences for all variables after matching. However, the standardized mean differences for the variable **IncRatio** and the propensity scores are just within the shaded region between –0.25 and 0.25 and would fail to meet a more stringent threshold (for example, a difference between –0.1 and 0.1). As discussed in Example 1, matching is best suited to data for which there is a large control population from which to draw matches for the treated subjects. For this example, where there is not a substantial difference in the size of the treatment and control conditions, the adjustments that are possible using an optimal matching method are limited.

Figure 11 Standardized Mean Difference Plot for Matching ATT Analysis
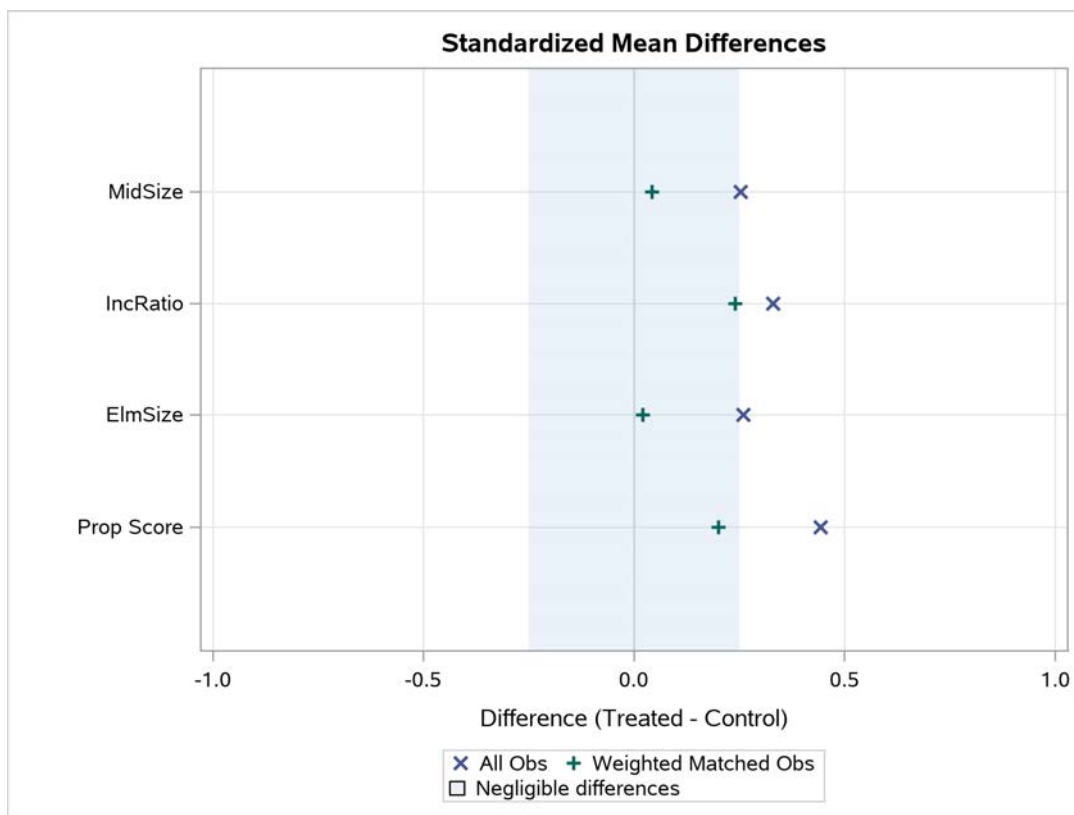
Figure 12 shows the box plots for the propensity score values and the variable **IncRatio** that are requested in the ASSESS statement. These box plots indicate improved balance in the distributions after matching, but with small differences remaining in the mean values, as seen in the standardized mean differences plot.

Figure 12 Box Plots for One-to-One Matching
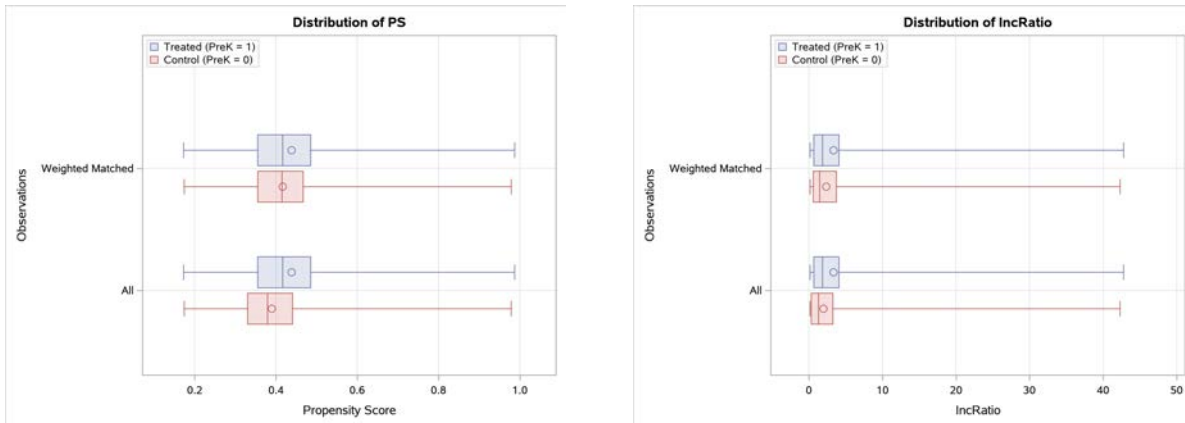
To decide whether to proceed to an outcome analysis, you need to determine whether these matched data show sufficient improvement in balance. This is a discussion that should involve all the different stakeholders in the analysis. If it is decided that the improvement in balance is insufficient, different formulations of the matching problem could be considered. For more information about different approaches to formulating a matching problem by using PROC PSMATCH, see Lamm, Thompson, and Yung (2019). For now, this example will proceed under the assumption that these matched data are acceptable for use in an outcome analysis. The following statements use PROC TTEST to estimate the ATT:

```
proc ttest data=psATTMatchData;
    class PreK;
    var MidProf;
    weight attWgt;
run;
```

Note that by default, the output data set **psATTMatchData** produced by PROC PSMATCH contains all the observations in the input data set, not just the matched observations. The variable **attWgt** contains the matching ATT weights, which are set to missing for each unmatched subject. It is therefore important to specify **attWgt** in the WEIGHT statement to ensure that the unmatched subjects are excluded from the outcome analysis. Figure 13 shows an ATT estimate of 0.0410, compared to the true ATT value of 0.0356.

Figure 13 ATT Estimate Obtained by Matching

PreK	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.7577	0.7446	0.7708	0.4285	0.4195	0.4380
1		0.7987	0.7864	0.8110	0.4010	0.3925	0.4099
Diff (1-2)	Pooled	-0.0410	-0.0590	-0.0230	0.4150	0.4088	0.4215
Diff (1-2)	Satterthwaite	-0.0410	-0.0590	-0.0230			

You can also estimate the ATT by using inverse probability of treatment weighting. As with the estimation of the ATE by inverse probability of treatment weighting, as shown in Example 1, you can perform the analysis in a two-stage approach with PROC PSMATCH or in a single step with PROC CAUSALTRT. The following statements use PROC CAUSALTRT to estimate the ATT by inverse probability of treatment weighting:

```
proc causaltrt data=ReadingObs att;
    class ESL MidProf PCGEd PreK / desc;
    psmodel PreK = ElmSize ESL IncRatio PCGEd;
    model MidProf;
    bootstrap seed=1976;
run;
```

Just as before, the propensity score model that is used to estimate the weights is specified in the PSMODEL statement. The MODEL statement identifies the outcome variable, but no additional model specification is provided because

estimation by inverse probability of treatment weighting does not require modeling of the outcome variable. By default, PROC CAUSALTRT estimates the ATE. To estimate the ATT, specify the ATT option in the PROC CAUSALTRT statement. Figure 14 shows the inverse probability of treatment estimate for the ATT.

Figure 14 ATT Estimate Obtained by Inverse Probability of Treatment Weighting

					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	0.7987	0.00626	0.00611	0.7864	0.8110	0.7862	0.8106	127.51	<.0001
POM	0	0.7644	0.00636	0.00643	0.7519	0.7768	0.7517	0.7774	120.26	<.0001
ATT		0.03430	0.00883	0.00889	0.01700	0.05160	0.01540	0.05075	3.89	0.0001

The ATT estimate of 0.0343 is close to the true value of 0.0356, and the empirical and bootstrap-based confidence limits are similar. For this example, the ATT and ATE values are similar. However, note that because the "Analysis of Causal Effect" table in Figure 14 is for an ATT estimate, the potential outcome mean estimates that it displays are conditioned on subjects being in the treatment condition. This results in larger estimates for the potential outcome mean values than the estimates in Figure 8 that were obtained in estimating the ATE.

Example 3: Estimating Direct and Indirect Effects

In Figure 1, the total causal effect of **PreK** on **MidProf** consists of two components. The first component is the indirect effect that consists of the effect of **PreK** on the mediator variable **ElmProf** and then the effect of **ElmProf** on **MidProf**. In the counterfactual framework, this effect is also called the natural indirect effect (NIE). The second component is the direct effect of **PreK** on **MidProf**. In the current context, this effect is best interpreted as the effect of **PreK** on **MidProf** that is not mediated through **ElmProf**. In the counterfactual framework, this direct effect is also called the natural direct effect (NDE). For brevity, the "natural" designation of these effects is dropped here.

A general goal of a causal mediation analysis is to quantify and estimate the direct and indirect effects. To understand how the NDE and NIE concepts are useful in practice, imagine a hypothetical situation where a sizable amount of the total positive effect of pre-K education (**PreK**) on eighth-grade reading proficiency (**MidProf**) is due to the mediation of fourth-grade reading proficiency (**ElmProf**). If underprivileged children were unable to attend pre-K, then a sizable NIE implies that it is still possible to boost eighth-grade reading proficiency by designing a better elementary education program for reading—that is, through an intervention on the mediator **ElmProf**. In contrast, in an extreme situation where **ElmProf** is not an effective mediator and the total effect is almost the same as the direct effect, it would be futile to try to boost **MidProf** by designing interventions that would enhance fourth-grade reading proficiency (**ElmProf**).

Identification of causal mediation effects such as the NIE and NDE in observational studies requires certain conditions to be satisfied. The discussion in the section "Identification Conditions and Causal Graphs" focuses on the conditions that must be met in order for you to meaningfully estimate a total causal effect. The current section continues to add the identification conditions for the NDE and NIE. Specifically, according to Shpitser and VanderWeele (2011), the set X that satisfies the conditional ignorability assumption must also satisfy

$$Y_{tm} \perp\!\!\!\perp \{T, M\} | X$$
$$M_t \perp\!\!\!\perp T | X$$

Valeri and VanderWeele (2013) outline an apparently different set of identification conditions, which also includes a so-called cross-world independence assumption. But Shpitser and VanderWeele (2011) suggest that the cross-world independence assumption would be satisfied if the other assumptions are met. Although these conditions are opaque as written, you can check them easily by using PROC CAUSALGRAPH. You can use the following code to assess whether the adjustment set X = {**ElmSize**, **ESL**, **IncRatio**, **PCGEd**} (which, as you have already seen, satisfies the conditional ignorability assumption for the total effect) satisfies the first of these new conditions:

```
proc causalgraph;
   model "ReadingProf"
      ElmProf => MidProf,
      ElmSize => ElmProf MidSize PreK,
```

```
        ESL IncRatio => ElmProf MidProf PreK,
        MidSize => MidProf,
        PCGEd => ElmProf IncRatio MidProf PreK,
        PreK => ElmProf,
        Values => PCGEd PreK;
    latent Values;
    identify PreK ElmProf => MidProf;
    testid 'CMedId1' ElmSize ESL IncRatio PCGEd;
run;
```

Note that the IDENTIFY statement replaces the actual treatment variable **PreK** with the combination of the treatment variable and the mediator variable **ElmProf**. Then the TESTID statement is used to test whether the condition holds for the specified adjustment set of covariates.

You can use the following code to test the second condition:

```
proc causalgraph;
    model "ReadingProf"
        ElmProf => MidProf,
        ElmSize => ElmProf MidSize PreK,
        ESL IncRatio => ElmProf MidProf PreK,
        MidSize => MidProf,
        PCGEd => ElmProf IncRatio MidProf PreK,
        PreK => ElmProf,
        Values => PCGEd PreK;
    latent Values;
    identify PreK => ElmProf;
    testid 'CMedId2' ElmSize ESL IncRatio PCGEd;
run;
```

The results of the two PROC CAUSALGRAPH tests are shown in Figure 15 and Figure 16. Because the set X satisfies both conditions (that is, the set is a valid adjustment in both situations), it is possible to estimate the NDE and NIE with valid causal interpretations under the data generating process that is represented by Figure 1. You can use PROC CAUSALMED to estimate the NDE and NIE with a specified adjustment covariate set X.

Figure 15 First Identification Test for Mediated Effects **Figure 16** Second Identification Test for Mediated Effects

Covariate Adjustment Test: CMedId1								Covariate Adjustment Test: CMedId2								
Causal Effect of {ElmProf PreK} on MidProf								Causal Effect of PreK on ElmProf								
Model	Size	Valid	Minimal	ElmSize	ESL	IncRatio	MidSize	PCGEd	Model	Size	Valid	Minimal	ElmSize	ESL	IncRatio	MidProf	MidSize	PCGEd
ReadingProf	4	Yes	Yes	*	*	*		*	ReadingProf	4	Yes	Yes	*	*	*			*

PROC CAUSALMED adopts the regression adjustment approach that is based on Valeri and VanderWeele (2013). There are two complications that arise when you attempt to apply this approach to the current example. The first complication is the scale on which the causal effects are computed. In the regression adjustment approach that PROC CAUSALMED implements, effects on binary outcomes (such as the current example) are computed on the odds ratio and excess relative risk scales. The derivation of the odds ratio and excess relative risk mediation effects assumes that the binary outcomes are rare events (Valeri and VanderWeele 2013). This assumption is not tenable in the current example because the population mean is 0.782 for the treatment and 0.745 for the control. In addition, the causal estimands in the pre-K examples are expressed on the risk difference scale. Note that PROC CAUSALMED does provide risk difference effects on continuous outcomes without assuming rare events. The second complication is that PROC CAUSALMED generally computes conditional effects rather than marginal effects. However, when linear models are applied to the outcome and mediator models, the causal mediation effects that PROC CAUSALMED estimates are also estimates of marginal effects, which is the type of effects that are reported in Example 1 and Example 2.

Therefore, the solution to these complications is to treat the the binary mediator (**ElmProf**) and binary outcome (**MidProf**) variables as though they were continuous variables and apply linear models as an approximation. In this way, estimates of the total causal effect, direct effect, and indirect effect are all interpreted on the risk difference scale. This maintains consistency with the scales of those estimands in Examples 1 and 2 and avoids the necessity of the rare outcome assumption. You can use the following code to conduct a causal mediation analysis of the current data:

```
proc causalmed data=ReadingObs;
   class ESL PCGEd PreK / desc;
   model MidProf = Prek | ElmProf;
   mediator ElmProf = PreK;
   covar ElmSize ESL IncRatio PCGEd;
run;
```

The MODEL statement specifies the outcome variable **MidProf** as a function of the treatment variable **PreK** and the mediator variable **ElmProf**. An interaction effect of the treatment and mediator variables is assumed. The MEDIATOR statement specifies the mediator variable **ElmProf** as a function of the treatment variable **PreK**.

The results of the analysis are shown in Figure 17. The estimated total effect of 0.034 is consistent with the ATE estimate that was obtained earlier and also consistent with the known true value of the ATE (0.0369). The estimate of the NIE is 0.0084 (the true value is 0.0108), and the estimate of the NDE is 0.0256 (the true value is 0.0261). Note that the total effect is the sum of the NIE and the NDE, exactly as expected. An estimated 25% of the total effect is due to mediation (the ratio of the NIE to the total effect). Therefore, the estimates that PROC CAUSALMED obtains using linear model approximations are quite reasonable.

Figure 17 displays other types of causal mediation effects (such as the controlled direct effect) that this paper does not cover. For more information about these advanced concepts of causal mediation effects and their decompositions, see Valeri and VanderWeele (2013). PROC CAUSALMED has a wide range of coverage of these advanced effects and decompositions (Yung, Lamm, and Zhang 2018).

Figure 17 Causal Mediation Analysis

Summary of Effects						
	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	0.03396	0.008735	0.01684	0.05108	3.89	0.0001
Controlled Direct Effect (CDE)	0.02537	0.008546	0.008624	0.04212	2.97	0.0030
Natural Direct Effect (NDE)	0.02555	0.008559	0.008778	0.04233	2.99	0.0028
Natural Indirect Effect (NIE)	0.008407	0.001927	0.004631	0.01218	4.36	<.0001
Percentage Mediated	24.7556	7.6304	9.8003	39.7109	3.24	0.0012
Percentage Due to Interaction	-0.7598	1.4251	-3.5530	2.0333	-0.53	0.5939
Percentage Eliminated	25.2832	7.6608	10.2683	40.2981	3.30	0.0010

Summary

There can be many different substantive questions that spark an interest in studying a treatment's causal effect on an outcome. In the counterfactual framework, causal estimands that are relevant to different motivating causal questions can be defined clearly, and conditions for the identification of these causal estimands can be made explicit. Moreover, this paper demonstrates the importance of using estimation methods that are appropriate given the estimand of interest and properties of the data generating process. The selection of an incorrect estimation method or a modeling choice that is inappropriate for the target estimand—for example, adjusting for a mediator variable in estimating a total effect—can result in biased and incorrect effect estimates.

The use of a formal causal analysis framework to reason about causal estimands is an important part of a rigorous causal analysis workflow. As you begin a causal analysis, the first step is to match your research question to a specific population of interest and then to choose the causal estimand best suited to answer your research question in your chosen population. When you have specified these pieces, you can use a causal diagram to reason about the data generating process that is assumed to give rise to the available data. It is also important to think about the identification assumptions for causal analysis and any possible violations of those assumptions that might be present in your data. The tools for analyzing graphical causal models in PROC CAUSALGRAPH can be beneficial when you are reasoning about a data generating process and the identifiability conditions. Finally, with all these pieces in place, you can use procedures in SAS/STAT, such as PROC PSMATCH, PROC CAUSALTRT, and PROC CAUSALMED, to compute an estimate that corresponds to your estimand of interest.

REFERENCES

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.

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

Hanea, A. M., McBride, M. F., Burgman, M. A., and Wintle, B. C. (2018). "Classical Meets Modern in the IDEA Protocol for Structured Expert Judgement." *Journal of Risk Research* 21:417–433.

Hernán, M. A., and Robins, J. M. (2019). *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC. Forthcoming.

Hernán, M. A., and Taubman, S. L. (2008). "Does Obesity Shorten Life? The Importance of Well-Defined Interventions to Answer Causal Questions." *International Journal of Obesity* 32:Suppl 3:S8–14.

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

Lamm, M., Thompson, W. C., and Yung, Y.-F. (2019). "Building a Propensity Score Model with SAS/STAT Software: Planning and Practice." In *Proceedings of the SAS Global Forum 2019 Conference*. Cary, NC: SAS Institute Inc. https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3056-2019.pdf.

Lamm, M., and Yung, Y.-F. (2017). "Estimating Causal Effects from Observational Data with the CAUSALTRT Procedure." In *Proceedings of the SAS Global Forum 2017 Conference*. Cary, NC: SAS Institute Inc. http://support.sas.com/resources/papers/proceedings17/SAS0374-2017.pdf.

Pearl, J. (2001). "Direct and Indirect Effects." In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, edited by J. Breese and D. Koller, 411–420. San Francisco: Morgan Kaufmann.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.

Robins, J. M., and Greenland, S. (1992). "Identifiability and Exchangeability for Direct and Indirect Effects." *Epidemiology* 3:143–155.

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.

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., and VanderWeele, T. J. (2011). "A Complete Graphical Criterion for the Adjustment Formula in Mediation Analysis." *International Journal of Biostatistics* 7:1–24.

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.

Valeri, L., and VanderWeele, T. J. (2013). "Mediation Analysis Allowing for Exposure-Mediator Interactions and Causal Interpretation: Theoretical Assumptions and Implementation with SAS and SPSS Macros." *Psychological Methods* 18:137–150.

VanderWeele, T. J., and Vansteelandt, S. (2009). "Conceptual Issues Concerning Mediation, Interventions and Compositions." *Statistics and Its Interface* 2:457–468.

VanderWeele, T. J., and Vansteelandt, S. (2010). "Odds Ratios for Mediation Analysis for a Dichotomous Outcome." *American Journal of Epidemiology* 172:1339–1348.

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.

Yung, Y.-F., Lamm, M., and Zhang, W. (2018). "Causal Mediation Analysis with the CAUSALMED Procedure." In *Proceedings of the SAS Global Forum 2018 Conference*. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

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

Clay Thompson
Email: clay.thompson@sas.com

Michael Lamm
Email: michael.lamm@sas.com

Yiu-Fai Yung
Email: yiu-fai.yung@sas.com

SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513