Paper SAS1991-2018

Causal Mediation Analysis with the CAUSALMED Procedure

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

Abstract

Important policy and health care decisions often depend on understanding the direct and indirect (mediated) effects of a treatment on an outcome. For example, does a youth program directly reduce juvenile delinquent behavior, or does it indirectly reduce delinquent behavior by changing the moral and social values of teenagers? Or, for example, is a particular gene directly responsible for causing lung cancer, or does it have an indirect (mediated) effect through its influence on smoking behavior? Causal mediation analysis deals with the mechanisms of causal treatment effects, and it estimates direct and indirect effects. A treatment variable is assumed to have causal effects on an outcome variable through two pathways: a direct pathway and a mediated (indirect) pathway through a mediator variable. This paper introduces the CAUSALMED procedure, new in SAS/STAT® 14.3, for estimating various causal mediation effects from observational data in a counterfactual framework. The paper also defines these causal mediation and related effects in terms of counterfactual outcomes and describes the assumptions that are required for unbiased estimation. Examples illustrate the ideas behind causal mediation analysis and the applications of the CAUSALMED procedure.

Introduction

In causal mediation analysis, a treatment variable, T (or an exposure, A, in the field of epidemiology), relates to an outcome variable, Y, by a specific causal mechanism that is represented by the following causal diagram:

$$\begin{tikzcd} \boxed{T} \arrow[rr] \arrow[dr] & & \boxed{Y} \\ & \boxed{M} \arrow[ur] & \end{tikzcd}$$

As depicted in the diagram, the total causal treatment effect of T consists of the following two parts:

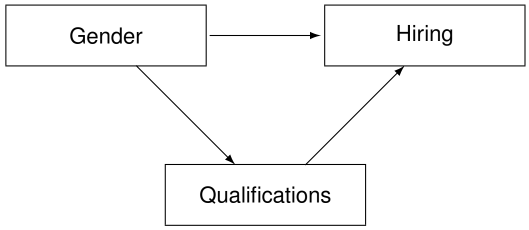
- a direct effect $T \rightarrow Y$
- an indirect or mediated effect $T \rightarrow M \rightarrow Y$, where M is called a mediator

Causal mediation analysis quantifies and estimates the total, direct, and indirect (or mediated) effects. It enables causal interpretations of these effects under the assumptions of the counterfactual framework (Robins and Greenland 1992; Pearl 2001).

This paper introduces the CAUSALMED procedure for causal mediation analysis and is organized as follows. First, the section "Motivation for Causal Mediation Analysis" uses two examples from the literature to illustrate the benefits of causal mediation analysis. Next, the section "Estimating Causal Mediation Effects in Observational Studies" describes the confounding issue of causal effect estimation in observational studies. It extends the estimation problem from that of the total effects to that of the mediation effects. The section "Key Questions in Causal Mediation Analysis" outlines the main research questions that can be addressed by causal mediation analysis, followed by the section "Features of the CAUSALMED Procedure." Then, Example 1 analyzes a simulated observational data set and demonstrates the basic features and output of PROC CAUSALMED. The section "Theory, Assumptions, and Estimation" describes the theoretical background and some technical details of PROC CAUSALMED. More analysis examples are presented next. Example 2 illustrates the mediation effect decompositions and their interpretations. The evaluation of conditional causal mediation effects and controlled direct effects is then demonstrated in Examples 3 and 4, respectively. A summary concludes the paper.

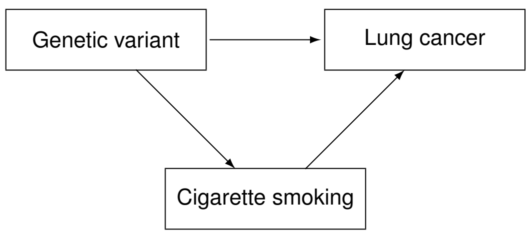
Motivation for Causal Mediation Analysis

Differentiating between the direct and indirect (mediated) components of a treatment effect is not only scientifically important, but it might also shed light on policy decisions. To demonstrate this point, Pearl (2001) gives an example of a hiring decision process, which can be represented by the following causal diagram:



Pearl (2001) argues that evidence about gender discrimination in the hiring process should be established on the *direct* effect of gender on hiring only. Neither the total effect of gender nor the mediation effect via qualifications forms the basis of gender discrimination. If there is indeed a gender bias in hiring but most of the portion of the total effect is due to the mediation, then policy makers should perhaps focus more on gender equality in education, which would potentially equalize the qualifications between the genders.

In the field of cancer research, VanderWeele (2015) provides an example about the direct and indirect effects of genetic variants in a particular chromosome on lung cancer. The genetic effects are represented by the following causal diagram:



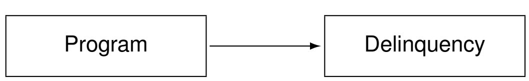
Is the effect on lung cancer largely due to the mediated effect via cigarette smoking? If so, intervention to reduce cigarette smoking would be an effective means of lowering the risk of lung cancer. If the mediation effect is negligible, then either lung cancer is a genetic predisposition or some unspecified mediators are the culprits. In the former case, there might be no cure for the root cause; in the latter case, researchers should continue to search for important mediators in which they could potentially intervene to reduce the risk of lung cancer.

Estimating Causal Mediation Effects in Observational Studies

The section lays out the context for all the data examples that are demonstrated in this paper. It begins with an example that demonstrates the confounding issue of estimating total causal treatment effects in observational studies. The estimation problem is then extended to that of causal mediation analysis.

Estimating Total Causal Treatment Effects

A youth program is designed to provide teenagers with healthful activities to help steer them away from delinquency. Teenagers who are randomly selected to participate in the program form the *treatment* group, and those who are not selected form the *control* group. The following idealized causal diagram represents the design of this randomized experiment:



Because of the random assignment, all factors other than the youth program itself are supposed to be balanced between the treatment and control groups. As a result, the observed difference in delinquency between the treatment and control groups has a valid causal interpretation; this total program effect is not confounded by other factors.

However, if assignments to the treatment and control groups are not random and the teenagers are instead allowed to participate in the youth program voluntarily, the study would become an observational study. The preceding causal diagram is not accurate unless the confounding causes or factors are added (for example, confounding factors in observational studies are often added to the diagram as common causes to the treatment (T) and outcome (Y) variables). Confounding factors or variables induce extraneous associations between T and Y, so the observed mean difference between the treatment and control groups would be a biased estimate of the causal effect.

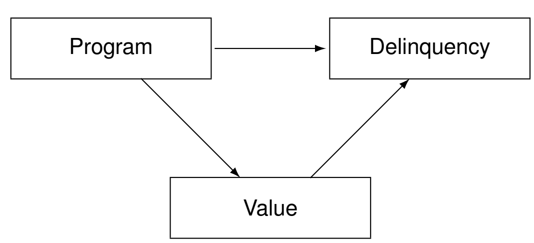
For example, if the youth program is costly or free transportation to the program activities is not provided, teenagers from lower socioeconomic classes might not be able to participate. Consequently, the treatment group would be overrepresented by teenagers with higher socioeconomic status (SES), and the control group would be overrepresented by teenagers with lower SES. The observed relationship between T and Y is thus confounded by SES, making it difficult to determine whether the delinquent behavior is due to the true program effect, the different SES makeups of the treatment and control groups, or both. In this case, you must somehow adjust for SES (which is called a confounding variable or a confounder) when you estimate the causal effect of the youth program.

Generally, in observational studies where random assignments of participants are not made, there could be a lot of confounders that cloud the observed relationships between the treatment and outcome variables. In the current example, in addition to SES, pretreatment characteristics such as teenagers' personality (extroversion/introversion), their degree of parental support, and their academic performance are also potential confounders that you might need to adjust for when you assess the causal effect of the youth program.

The CAUSALTRT and PSMATCH procedures in SAS/STAT software (SAS Institute Inc. 2017) provide statistical adjustments for confounding pretreatment characteristics so that the total causal treatment effect can be estimated without bias. For more information about these procedures as well as application examples, see Lamm and Yung (2017) and Yuan, Yung, and Stokes (2017). This paper extends beyond the estimation of total causal treatment effects and discusses the decomposition of causal effects through causal mediation analysis by using the CAUSALMED procedure.

Mechanism of the Causal Process

In causal mediation analysis, not only is the total causal effect of interest, but the causal mechanism is important as well. The following causal diagram extends the preceding youth program example by including a mediator variable, Value (for teenagers' moral and social values), in the causal process:



This diagram has three variables:

- The outcome variable **Delinquency**—Delinquent behavior, which counts the number of instances of delinquent behavior within two years after participating the youth program.
- The treatment variable **Program**—Youth program, which is a binary variable that indicates whether a teenager participated in the program.
- The mediator variable **Value**—Moral and social values, which reflects the teenagers' moral and social values after participating in the youth program. A psychological instrument is used to measure these values.

This causal diagram theorizes that part of the **Program** effect on **Delinquency** is due to the mediated process of **Value**. That is, the youth program might elevate the teenagers' moral and social values, affecting their observed

Value level, which in turn reduces subsequent delinquent behavior. In addition, the direct pathway from **Program** to **Delinquency** represents the "residual" or direct treatment effect that is not due to the mediation process.

Causal mediation analysis enables you to understand the mechanism of the causal process. If the youth program does reduce subsequent delinquent behavior, is it largely due to the program's intermediate effect in changing the teenagers' moral and social values? If so, then the program administrator (or policy maker) might want to make certain that future youth programs retain or even increase the educational elements for moral and social values. If not, then the program administrator might want to investigate what other mediating factors might have been important in the process and incorporate these factors into the design of future youth programs.

As in the case of estimating causal treatment effects (see Example 1), causal mediation analysis from observational studies is also subject to confounding. Even if treatment assignments have been randomized, causal interpretations and estimation of mediation and related effects might still be subject to confounding because the mediator levels are usually not randomized. Thus, you need to take confounding covariates into account when you estimate causal mediation effects. Example 1 demonstrates a basic analysis by using the CAUSALMED procedure.

Key Questions in Causal Mediation Analysis

To summarize, causal mediation analysis helps researchers find answers to the following key questions:

- Is part of the total causal treatment effect due to the specified mediation process?
- Is there an interaction effect between the treatment and the mediator on the outcome?
- What percentage of the total causal effect is due to the mediation, the interaction, or both?
- Can you intervene in the causal mediation process to achieve desirable outcomes?

To answer these questions, causal mediation analysis must deal with the following main issues:

- Definition—How can various causal mediation effects be defined appropriately in a general modeling situation?
- Interpretation—How do you properly interpret the various causal mediation effects and decompositions?
- Identification—Under what conditions can various causal mediation effects be estimated without bias?
- Estimation—How can various causal mediation effects be estimated?

The counterfactual framework (Robins and Greenland 1992; Pearl 2001) has been proposed to deal with the definition, interpretation, and identification issues of causal mediation analysis. Based on the counterfactual framework, the CAUSALMED procedure implements the regression adjustment method to estimate causal mediation effects (Valeri and VanderWeele 2013; VanderWeele 2014). For more information about the technical details of the counterfactual framework and the estimation methods of PROC CAUSALMED, see the section "Theory, Assumptions, and Estimation."

Features of the CAUSALMED Procedure

The CAUSALMED procedure supports a limited set of generalized linear models for describing the relationships among the outcome, treatment, mediator, and confounder variables. With this set of models, you can use PROC CAUSALMED to conduct causal mediation analysis on data of the following types:

- outcome variable Y: binary, continuous, or count
- treatment variable T: binary or continuous
- mediator variable M: binary or continuous
- covariates: categorical or continuous

The main features and output from PROC CAUSALMED include the following:

- estimates of total, controlled direct, natural direct, and natural indirect effects
- effects that are computed on the odds ratio scale and the excess relative risk scale for binary responses
- percentages of the total effect that are attributed to mediation and interaction, and the percentage of the total effect eliminated by controlling the mediator level
- various decompositions of total effects, including several two-, three-, and four-way decompositions
- flexible evaluation of controlled direct effects and conditional mediation effects

Other important features that you can optionally request from the procedure include these:

- analysis with case-control design
- bootstrap estimation of standard errors and confidence intervals
- output of outcome and mediator model estimates

Example 1. Basic Causal Mediation Analysis of the Youth Program Data

This section demonstrates how you can use the CAUSALMED procedure to perform causal mediation analysis of observational data. As described in Example 2, a youth program was designed to foster teenagers' personal growth and thus reduce their delinquent behavior. The analysis also sought to determine whether any of the program effect was due to the mediation of the teenagers' moral and social values during their participation in the process.

A data set for an observational study was simulated. It contains 800 observations with the following variables:

- **Delinquency**, the frequency of delinquent behavior within two years *after* the outset of the youth program; this is the *outcome* variable of interest.
- **Program**, the indicator of youth program participation (Yes or No); this is the *treatment* variable of interest.
- **Value**, the teenager's score on a moral and social values scale one year after the outset of the youth program; this is the *mediator* variable of interest.
- **Delinquency0**, the frequency of delinquent behavior within two years *before* the outset of the youth program
- **Gender**, the gender of the teenager
- **GPA**, the teenager's grade point average in school at the outset of the youth program
- **Introversion**, the indicator of self-described introversion (Yes or No)
- **SES**, the teenager's socioeconomic status (Low, Medium, or High)
- **Support**, the degree of parental support on a four-point scale, which was assessed at the outset of the program
- **Value0**, the teenager's score on a moral and social values scale at the outset of the youth program

The last seven variables are covariates that might have confounded the observed relationships among the outcome, treatment, and mediator variables.

The first 10 observations of the **Youth** data set are shown in Figure 1.

Figure 1 First 10 Observations of the **Youth** Data Set

Obs	Delinquency	Program	Value	Delinquency0	Gender	GPA	Introversion	SES	Support	Value0
1	3	No	20	3	Female	3.21	Yes	Medium	2	18
2	4	No	19	4	Male	3.11	No	High	1	17
3	3	No	19	3	Female	2.97	Yes	Medium	3	19
4	3	No	21	3	Female	3.56	No	Low	3	19
5	4	No	18	4	Female	2.56	No	High	1	17
6	3	Yes	22	3	Female	2.90	No	Medium	3	19
7	4	No	19	4	Male	3.32	Yes	Medium	2	18
8	2	Yes	22	3	Female	3.06	No	Medium	4	20
9	4	No	19	4	Male	3.06	No	Medium	1	16
10	2	Yes	24	3	Female	3.31	Yes	Medium	4	20

You obtain descriptive statistics for the variables by running the following statements:

```
proc freq data=Youth;
   table Program Gender Introversion SES;
run;

proc means data=Youth;
   var Delinquency Delinquency0 Value Value0 GPA Support;
run;
```

The frequency distribution tables (output not shown) indicate that 30% of the teenagers in the study participated in the youth program, 57% were female, and 50% were self-reported introverts. For the SES, 25% were High, 45% were Medium, and 30% were Low.

Figure 2 displays some descriptive statistics of the continuous variables in the analysis.

Figure 2 Descriptive Statistics for Continuous Variables

Variable	N	Mean	Std Dev	Minimum	Maximum
Delinquency	800	3.1487500	0.7761291	1.0000000	5.0000000
Delinquency0	800	3.4000000	0.5726342	2.0000000	5.0000000
Value	800	20.0837500	1.8731315	15.0000000	25.0000000
Value0	800	18.1737500	1.2001506	15.0000000	21.0000000
GPA	800	3.0239858	0.4095261	1.6576112	4.7995488
Support	800	2.4537500	1.0579509	1.0000000	4.0000000

Figure 2 shows that the teenagers' average delinquent behavior decreased after the outset of the program. By comparing the means of **Value** and **Value0**, you see that the teenagers' average moral and social values also improved. But these changes do not distinguish program participants from nonparticipants. So you now examine these statistics separately for the treatment group (program participants) and control group (nonparticipants) by running the following statements:

```
proc means data=Youth;
   class program;
   var Delinquency Delinquency0 Value Value0 GPA Support;
run;
```

For the program participants, Figure 3 shows a 0.7-point reduction in delinquent behavior. It also shows an average of about a 3-point improvement in moral and social values. For the nonparticipants, the average reduction of delinquent behavior is about 0.05, and the average improvement in moral and social values is about 1.5 points. These findings suggest that the teenagers matured naturally, irrespective of participation in the youth program. Hence, those observed mean differences between **Delinquency** and **Delinquency0** and between **Value** and **Value0** for the treatment group reflect not only the causal effects of the youth program but also the natural maturing process. Hence, you must adjust for baseline characteristics such as **Delinquency0** and **Value0** when you estimate the causal mediation effects. These characteristics will be included as covariates in the causal mediation analysis.

Figure 3 Means of the Continuous Variables in the Treatment and Control Groups

Program	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
No	557	Delinquency	557	3.5134650	0.5739379	2.0000000	5.0000000
		Delinquency0	557	3.5655296	0.5607969	2.0000000	5.0000000
		Value	557	19.2477558	1.4628553	15.0000000	24.0000000
		Value0	557	17.7935368	1.1357949	15.0000000	21.0000000
		GPA	557	3.0943297	0.3997079	1.7885487	4.7995488
		Support	557	2.0000000	0.8732640	1.0000000	4.0000000
Yes	243	Delinquency	243	2.3127572	0.4733844	1.0000000	3.0000000
		Delinquency0	243	3.0205761	0.3904708	2.0000000	4.0000000
		Value	243	22.0000000	1.1712817	19.0000000	25.0000000
		Value0	243	19.0452675	0.8344396	17.0000000	21.0000000
		GPA	243	2.8627450	0.3862172	1.6576112	3.8540835
		Support	243	3.4938272	0.6190527	2.0000000	4.0000000

In addition, Figure 3 reveals the confounding issue that is common in observational studies. Generally, confounding covariates are those pretreatment characteristics that are associated with the treatment, mediator, and outcome variables but cannot be controlled or manipulated in an observational study. They cloud the interpretations of the treatment effects by creating extraneous associations among the outcome, treatment, and mediator variables. For example, Figure 3 shows that the mean of **Support** (parental support) of the treatment group is much higher than that of the control group. Instead of a real program effect, such coincidentally better parental support in the treatment group could have explained the larger reduction in delinquent behavior after the youth program. Similarly, **Gender**, **Introversion**, **SES**, and **GPA** are potential confounding variables that you need to take into account when you estimate causal mediation effects.

In sum, to estimate the treatment (youth program) and mediator (moral and social values) effects with valid causal interpretations, you must figure out all those confounding covariates in the causal mediation process.

You can use the following statements to specify all these required elements for analyzing the data in the current example:

```
proc causalmed data=Youth;
   class Program(descending) Gender SES Introversion;
   model    Delinquency = Program Value Program*Value;
   mediator Value       = Program;
   covar    Support GPA Value0 Delinquency0 Gender SES Introversion;
run;
```

The PROC CAUSALMED statement invokes the causal mediation analysis that uses regression adjustment methods (Valeri and VanderWeele 2013; VanderWeele 2014) to estimate causal treatment and mediation effects. The DATA= option in the PROC CAUSALMED statement inputs the **Youth** data set.

The CLASS statement specifies the categorical variables in the analysis. Variables that are not specified in the CLASS statement are treated as continuous variables. The DESCENDING option for the **Program** variable in this statement signifies that the 'Yes' response level represents the treatment condition (instead of using the 'No' response level in accordance with the default alphabetical order rule).

In the MODEL statement, you specify the outcome variable, **Delinquency**, and how it is to be causally modeled by the treatment and mediator variables. You can specify two types of models: a model with only main effects from the treatment and mediator variables or, as in this example, a model that includes both main effects and an interaction effect. Generally, without strong prior knowledge about the absence of the interaction effect, the latter type of model specification is recommended. In the MEDIATOR statement you specify the mediator variable, **Value**, on the left side of the equal sign and the treatment variable, **Program**, on the right side of the equal sign. The treatment and mediator variables that you specify in the MEDIATOR statement must match the variables specified in the MODEL statement. Finally, in the COVAR statement, you specify all the relevant confounding covariates.

Figure 4 displays the modeling information, which includes the outcome, treatment, and mediator variable specifications. Because both the outcome and mediator variables are continuous, linear models (with identity link functions) are used. There are no missing data, so the numbers of observations read and used are both 800. The "Class Level Information" table shows the response levels in the classification variables. Notice that the first (or treatment) level of **Program** is 'Yes' because the DESCENDING option is specified in the CLASS statement.

Figure 4 Modeling Information

Model Information	
Data Set	WORK.YOUTH
Outcome Variable	Delinquency
Treatment Variable	Program
Mediator Variable	Value
Outcome Distribution	Normal
Outcome Link Function	Identity
Mediator Distribution	Normal
Mediator Link Function	Identity

Number of Observations Read	800
Number of Observations Used	800

Class Level Information		
Class	**Levels**	**Values**
Program	2	Yes No
Gender	2	Female Male
SES	3	High Low Medium
Introversion	2	No Yes

Figure 5 displays the main estimation results. The first four rows display causal effect estimates, and the last three rows display the percentages of specific effects in relation to the total effect.

Figure 5 Summary of Estimation of Causal Mediation Effects

Summary of Effects	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.5621	0.0413	-0.6430	-0.4812	-13.62	<.0001
Controlled Direct Effect (CDE)	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
Natural Direct Effect (NDE)	-0.4479	0.0572	-0.5600	-0.3357	-7.83	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	20.3246	6.0692	8.4291	32.2201	3.35	0.0008
Percentage Due to Interaction	7.6512	4.2381	-0.6553	15.9577	1.81	0.0710
Percentage Eliminated	16.9866	4.6894	7.7956	26.1777	3.62	0.0003

The total treatment effect of –0.5621 represents a significant total effect of **Program** on **Delinquency**, assuming that all important confounding covariates have been included in the analysis (that is, the analysis satisfies the "no unmeasured confounding assumption"—to be explained in the section "Theory, Assumptions, and Estimation"). The minus sign indicates that the youth program *reduces* delinquent behavior. An interpretation of this total effect is that on average, a participant in the youth program shows about a half unit of delinquent behavior *less* than if he or she had not participated in the program. Or, on average, a nonparticipant would show a *decrease* in delinquent behavior by about a half unit if he or she had taken part in the program.

The controlled direct effect (CDE) is –0.4666, which is also significantly below 0. This represents the causal effect of the youth program on the delinquent behavior if everyone's moral and social values are fixed at the sample mean level (that is, 20.0838 for the variable **Value**; see Figure 2). Notice that in this example PROC CAUSALMED defines the CDE at the mean mediator level by default. Other controlled values, which might be more meaningful in specific situations, for the mediator level can be requested by using the EVALUATE statement. For more information about applications and interpretations of the CDE, see Example 4.

Next, the natural direct and indirect effects constitute a decomposition of the total effect. Loosely speaking, the natural indirect effect (NIE) refers to the program effect on **Delinquency** that is mediated by affecting the moral and social values (the variable **Value**) of the participants. The natural direct effect (NDE) is the residual, or direct, program effect that is not mediated by **Value**. The decomposition of the total causal effect of –0.5621 in this example is an NIE of

–0.1142 and an NDE of –0.4479. Both NIE and NDE are significantly lower than 0, meaning that both effects lead to reductions in delinquent behavior.

The row with the label Percentage Mediated displays the percentage of NIE in relation to the total effect. Approximately 20% of the program effect is attributed to the mediation of moral and social values. Therefore, **Value**, as the mediator variable here, accounts for a substantial portion of the youth program effect. The remaining 80% is due to unspecified causal mechanisms of the program.

The row with the label Percentage Due to Interaction displays the percentage of the total program effect due to the interaction between **Program** and **Value**, which is about 8%. Because the 95% confidence interval for this percentage includes zero, the percentage contribution due to the interaction is not statistically significant.

The last row shows the percentage eliminated (PE), which is the percentage of total effect that could have been eliminated if the mediator **Value** were fixed at the sample average level. The percentage eliminated is mathematically defined as (1–CDE/Total Effect)100%. Interpretations of the PE, in relation to the CDE, at various levels of the mediator are discussed in more detail in Example 4.

Theory, Assumptions, and Estimation

This section summarizes the theoretical background and technical details of the CAUSALMED procedure. Readers who would like to first learn the causal mediation concepts from practical examples may skip to Example 2.

Mediation analysis has a relatively long history in psychology, as typified by the work of Baron and Kenny (1986). However, this more traditional approach suffers from several theoretical and practical problems, including the following:

- It lacks a general framework to define causal mediation and related effects, so these causal effects are not well defined when binary mediators or outcomes are considered.
- It excludes the interaction effect between treatment and mediator.
- It does not explain the assumptions and identification conditions for valid causal effect estimation.

The counterfactual framework (Robins and Greenland 1992; Pearl 2001) offers solutions to these problems. Within this framework, direct and indirect effects are well defined in terms of counterfactual outcomes. Using these definitions, VanderWeele and Vansteelandt (2009, 2010) derived analytic results for computing causal mediation effects across a large class of parametric models for various types of treatment and outcome variables. Valeri and VanderWeele (2013) extended these results to binary mediators and count outcomes. This line of development provides the theoretical foundation of the CAUSALMED procedure.

The Counterfactual Framework

In the counterfactual framework for causal treatment effect analysis (without mediation), a counterfactual outcome is the outcome that you would observe under a hypothetical intervention that you can set the treatment T to a particular level, t. As the name implies, counterfactual outcomes, which are also called potential outcomes, are therefore defined for scenarios that might be contrary to the factual outcomes.

In the counterfactual framework for causal mediation analysis, interventions on the mediator level are also used in various hypothetical scenarios to define mediation effects. The following subject-level notation is used for counterfactual outcomes that depend on interventions:

- M_t is the counterfactual outcome of M when an intervention sets the treatment level to $T = t$.
- Y_{tm} is the counterfactual outcome of Y when an intervention sets the treatment level to $T = t$ and the mediator level to $M = m$.
- $Y_{tM_{t*}}$ is the counterfactual outcome of Y when an intervention sets the treatment level to $T = t$ and the mediator to the level M_{t*}, the natural mediator level for a possibly different treatment level t_*.

This notation places no restriction on variable types. The variables Y, T, and M can be continuous or binary.

Suppose for the moment that the treatment is binary, so t is either 0 or 1, denoting the control (no treatment) and treatment conditions, respectively. The total effect (TE) for a subject is defined as the difference between the counterfactual outcomes at the treatment and control levels:

$$\text{TE} = Y_{1M_1} - Y_{0M_0}$$

In this equation, the first subscript in the counterfactual outcomes denotes the level (at either 1 or 0) that is set for the treatment, and the second subscript denotes the mediator value (either M_1 or M_0) that would follow from setting the treatment to a specific level. Notice that M_1 and M_0 are not fixed values, so their realized values could be different for different subjects. Because Y_{iM_i} in the TE definition has consistent treatment levels i, Y_i is an equivalent notation for Y_{iM_i} and therefore TE can also be defined by $Y_1 - Y_0$.

The controlled direct effect (CDE) for a subject is defined as the difference between the counterfactual outcomes at the two treatment levels when an intervention sets the mediator to a particular level, $M = m$. That is,

$$\text{CDE}(m) = Y_{1m} - Y_{0m}$$

The natural direct effect (NDE) for a subject is defined as the difference between the counterfactual outcomes at the two treatment levels when an intervention sets the mediator value to $M = M_0$, which is the natural level of the mediator when there is no treatment. That is,

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

The natural indirect effect (NIE) for a subject is defined as the difference between the counterfactual outcomes at the two mediator levels at M_1 and M_0 when an intervention sets the treatment to $T = 1$. That is,

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

If the treatment variable T is continuous, then the treatment levels must be defined according to the treatment and control levels of interest. For example, if t_1 and t_0 are the treatment and control levels on a continuous scale and they represent the levels of substantive interest, then they should replace the values 1 and 0, respectively, for the treatment and control levels in the definitions. However, this more general notation is not used here because it would unnecessarily complicate the presentation.

So far, TE, CDE, NDE, and NIE have been defined at the subject level. The corresponding effects at the population level are defined by the average or expectation of these effects—that is, $E[\text{TE}]$, $E[\text{CDE}(m)]$, $E[\text{NDE}]$, and $E[\text{NIE}]$, respectively. PROC CAUSALMED estimates these population-level effects, as illustrated in Example 1.

The counterfactual definitions of effects have two important properties. First, they yield the following well-known formula of a two-way decomposition of the total effect (TE):

$$\text{TE} = \text{NDE} + \text{NIE}$$

Or, using the population notation, the two-way decomposition is of the form

$$E[\text{TE}] = E[\text{NDE}] + E[\text{NIE}]$$

The percentage of total effect that is mediated (PM) is then computed as

$$\text{PM} = E[\text{NIE}]/E[\text{TE}] * 100\%$$

Second, and more importantly, because these definitions are independent of the model for the outcome or mediator, you can apply them to a wide range of modeling situations. That is, these definitions of effects and the total effect decomposition formulas are applicable to various variable types and linear or nonlinear models with or without an interaction effect between T and M.

VanderWeele (2014) took a step further and introduced the following four-way decomposition of the total effect:

$$\text{TE} = \text{CDE} + \text{IRF} + \text{IMD} + \text{PIE}$$

The four component effects in this equation are characterized by how they represent the interaction and mediation effects, as follows:

- CDE—controlled direct effect: the component that is not due to interaction or mediation
- IRF—reference interaction: the component that is due to interaction but not mediation
- IMD—mediated interaction: the component that is due to both interaction and mediation
- PIE—pure indirect effect: the component that is due to mediation but not interaction

In VanderWeele (2014), IRF is denoted as INT_{ref} and IMD is denoted as INT_{med}. These four component effects are also defined in terms of counterfactual outcomes (see VanderWeele 2014). For an illustration of the four-way decomposition, see Example 2.

Again, all the preceding effect notations for the four-way decomposition are defined at the subject level. The corresponding population parameters are obtained by taking expectations of the subject-level effects over the entire population. The composite relationship still holds for the population component effects. That is,

$$E[\text{TE}] = E[\text{CDE}] + E[\text{IRF}] + E[\text{IMD}] + E[\text{PIE}]$$

Dividing each of these population component effects by $E[\text{TE}]$ yields the corresponding proportion contributions of these components to the total effect. However, these contributions might not be interpretable when the component effects have mixed signs.

Two important relationships between the four-way decomposition and the preceding two-way decomposition are expressed by the following equations:

$$\begin{aligned} \text{NDE} &= \text{CDE} + \text{IRF} \\ \text{NIE} &= \text{PIE} + \text{IMD} \end{aligned}$$

The first equation expresses the natural direct effect (NDE) as the composite component of the controlled direct effect and reference interaction. The second equation expresses the mediation effect or natural indirect effect (NIE) as the composite component of the pure indirect effect and mediated interaction. Three-way decompositions of the total effect follow from substituting the expression for either the NDE or the NIE into the two-way total effect decomposition.

Another useful composite component is the "portion attributed to interaction" (PAI) between T and M, which is defined as

$$\text{PAI} = \text{IRF} + \text{IMD}$$

At the population level, the percentage of total effect that is due to the interaction is therefore computed as

$$E[\text{PAI}]/E[\text{TE}] * 100\%$$

Finally, the portion eliminated (PE), which is defined as

$$E[\text{PE}] = 1 - E[\text{CDE}]/E[\text{TE}]$$

is the portion of the total effect that is eliminated when an intervention sets the mediator variable to a particular level. Multiplying the right side of the PE formula by 100% yields the definition of "percentage eliminated." For an illustration of the use of the CDE and PE estimates, see Example 4 .

For further discussion of the various two-way and three-way decompositions and their relationships to the four-way decomposition, see VanderWeele (2014). You can use the DECOMP option in the CAUSALMED procedure to obtain several two-way decompositions, several three-way decompositions, and the four-way decomposition. For an illustration and further discussion of the various decompositions, see Example 2.

When the outcome responses are binary, various mediation and related effects are defined on the odds ratio or excess relative risk scale (Valeri and VanderWeele 2013; VanderWeele 2014). Parallel conceptions about effect components and decompositions to those that have been discussed in this section are still applicable. See the PROC CAUSALMED documentation (SAS Institute Inc. 2017) for examples.

Identification of Causal Mediation Effects

This section describes the identification conditions of causal mediation effects and their implications for applying statistical methods that aim to obtain unbiased estimation of the effects.

First, it is useful to distinguish the following three types of confounding covariates:

- a treatment-outcome confounder that confounds the relationship between T and Y
- a mediator-outcome confounder that confounds the relationship between M and Y
- a treatment-mediator confounder that confounds the relationship between T and M

Let C denote the set of all covariates that could be one of these three types of confounders. Thus, controlling for C in regression analysis means that all types of confounding covariates are being controlled for.

According to Valeri and VanderWeele (2013), the following four assumptions are required for the identification of causal mediation effects:

- no unmeasured treatment-outcome confounders given C
- no unmeasured mediator-outcome confounders given (C, T)
- no unmeasured treatment-mediator confounders given C
- no mediator-outcome confounder is affected by T (directly or indirectly) given C

The identification of the controlled direct effect (CDE) assumes the first two conditions, and the identification of the natural direct effect (NDE) and the natural indirect effect (NIE) assumes all four conditions. These four assumptions are collectively called the "no unmeasured confounding assumption." Formal statements about these identification conditions can be found in the appendix of Valeri and VanderWeele (2013) and in VanderWeele (2015).

In practice, these assumptions imply that in order to have valid causal interpretations of the mediation effects, you must be able to measure all relevant confounding covariates C and include them in a causal mediation analysis.

Estimation of Causal Mediation Effects

The CAUSALMED procedure implements regression methods for estimating causal mediation effects that assume the identification conditions of the preceding section along with correct specification of the following two models:

- the outcome model for Y given T, M, and C
- the mediator model for M given T and C

For a class of generalized linear models and under the counterfactual framework, VanderWeele and Vansteelandt (2009, 2010) and Valeri and VanderWeele (2013) derived analytic formulas for computing various causal mediation effects for different variable types. PROC CAUSALMED implements these analytic formulas for estimating causal mediation effects and their standard errors after maximum likelihood estimation of the outcome and mediator models. For the case that has a binary outcome and a continuous mediator, the analytic formulas assume that the outcome Y is a rare event (Valeri and VanderWeele 2013; VanderWeele 2014). If Y is not rare, then the formulas are still valid if Y is modeled by using a log link.

Overall and Conditional Causal Mediation Effects

Because of nonlinearity and inclusion of interaction terms in the outcome models, causal mediation effects are usually different at different levels of covariate values. Let θ represent the vector that collects all parameters in the outcome and mediator models. Under the correct specification of models and the identification assumptions, the causal effects in a mediation analysis are functions of θ conditional on the covariate values. That is, a causal effect, which is denoted by ef, can be expressed as a known function g_{ef} of θ given $C = c$,

$$g_{\text{ef}}(\theta \mid C = c),$$

where c represents some fixed values for covariates C. By default, PROC CAUSALMED computes $g_{\text{ef}}(\theta \mid C = c)$ with $c = c_0$, where c_0 is the sample mean value of C. This default setting provides "overall" measures of various causal mediation effects. It is consistent with the treatment of the SAS® macros that are implemented by Valeri and VanderWeele (2013). If the outcome and mediator models are linear, then the overall effects are also marginal effects. For more information about the computation of c_0, see the PROC CAUSALMED documentation (SAS Institute Inc. 2017).

However, in some applications you might be interested in the *conditional* mediation effects instead of the overall counterparts. Denote a conditional causal mediation effect at the covariate value c_k as

$$g_{\text{ef}}(\theta \mid C = c_k)$$

For example, a particular c_1 value defines a highly motivated subpopulation and a particular c_2 value defines a less motivated subpopulation. In this case, you want to compare $g_{\text{ef}}(\theta \mid C = c_1)$ with $g_{\text{ef}}(\theta \mid C = c_2)$ to see whether the specific mediation effects would be the same for the subpopulations with different motivational levels. You can use the EVALUATE statement in the CAUSALMED procedure to estimate these conditional causal mediation effects. See Example 3 for an illustration.

Example 2. Effect Decompositions

Example 1 shows that 20% of the youth program effect could be attributed to the mediation process in which moral and social values of the teenagers were elevated by the program. This example continues the analysis and illustrates various mediation effect decompositions that are available in the CAUSALMED procedure. The only new option that you need to add in this example is the DECOMP option in the PROC CAUSALMED statement, as shown in the following specification:

```
proc causalmed data=Youth decomp;
   class Program(descending) Gender SES Introversion;
   model    Delinquency = Program Value Program*Value;
   mediator Value       = Program;
   covar    Support GPA Value0 Delinquency0 Gender SES Introversion;
run;
```

Both Figure 6 and Figure 7 show six different decompositions of the total effect: three of them are two-way decompositions, two of them are three-way, and one is four-way.

Figure 6 Various Decompositions of Total Effect

Decompositions of Total Effect

Decomposition	Effect	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
NDE+NIE	Natural Direct	-0.4479	0.0572	-0.5600	-0.3357	-7.83	<.0001
	Natural Indirect	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
CDE+PE	Controlled Direct	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
	Portion Eliminated	-0.0955	0.0243	-0.1432	-0.04782	-3.93	<.0001
TDE+PIE	Total Direct	-0.5096	0.0429	-0.5936	-0.4257	-11.89	<.0001
	Pure Indirect	-0.0525	0.0210	-0.09373	-0.01123	-2.49	0.0127
NDE+PIE+IMD	Natural Direct	-0.4479	0.0572	-0.5600	-0.3357	-7.83	<.0001
	Pure Indirect	-0.0525	0.0210	-0.09373	-0.01123	-2.49	0.0127
	Mediated Interaction	-0.0618	0.0324	-0.1252	0.001660	-1.91	0.0563
CDE+PIE+PAI	Controlled Direct	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
	Pure Indirect	-0.0525	0.0210	-0.09373	-0.01123	-2.49	0.0127
	Portion Due to Interaction	-0.0430	0.0226	-0.08725	0.001231	-1.91	0.0567
Four-Way	Controlled Direct	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
	Reference Interaction	0.0188	0.00992	-0.00068	0.03820	1.89	0.0585
	Mediated Interaction	-0.0618	0.0324	-0.1252	0.001660	-1.91	0.0563
	Pure Indirect	-0.0525	0.0210	-0.09373	-0.01123	-2.49	0.0127
Total	Total Effect	-0.5621	0.0413	-0.6430	-0.4812	-13.62	<.0001

Note: NDE=CDE+IRF, NIE=PIE+IMD, PAI=IRF+IMD, PE=PAI+PIE, TDE=CDE+PAI.

Figure 7 Various Percentage Decompositions of Total Effect

Decomposition	Effect	Percent	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
NDE+NIE	Natural Direct	79.68	6.07	67.78	91.57	13.13	<.0001
	Natural Indirect	20.32	6.07	8.43	32.22	3.35	0.0008
CDE+PE	Controlled Direct	83.01	4.69	73.82	92.20	17.70	<.0001
	Portion Eliminated	16.99	4.69	7.80	26.18	3.62	0.0003
TDE+PIE	Total Direct	90.66	3.68	83.44	97.89	24.61	<.0001
	Pure Indirect	9.34	3.68	2.11	16.56	2.53	0.0113
NDE+PIE+IMD	Natural Direct	79.68	6.07	67.78	91.57	13.13	<.0001
	Pure Indirect	9.34	3.68	2.11	16.56	2.53	0.0113
	Mediated Interaction	10.99	6.08	-0.93	22.90	1.81	0.0707
CDE+PIE+PAI	Controlled Direct	83.01	4.69	73.82	92.20	17.70	<.0001
	Pure Indirect	9.34	3.68	2.11	16.56	2.53	0.0113
	Portion Due to Interaction	7.65	4.24	-0.66	15.96	1.81	0.0710
Four-Way	Controlled Direct	83.01	4.69	73.82	92.20	17.70	<.0001
	Reference Interaction	-3.34	1.86	-6.99	0.31	-1.79	0.0731
	Mediated Interaction	10.99	6.08	-0.93	22.90	1.81	0.0707
	Pure Indirect	9.34	3.68	2.11	16.56	2.53	0.0113

Percentage Decompositions of Total Effect

Note: NDE=CDE+IRF, NIE=PIE+IMD, PAI=IRF+IMD, PE=PAI+PIE, TDE=CDE+PAI.

The ‘NDE+NIE’ Decomposition

The most commonly reported decomposition is the ‘NDE+NIE’ two-way decomposition, which consists of the natural direct and natural indirect effects. The estimation of these effects is shown in Figure 6, and their corresponding percentages are shown in Figure 7. These results were shown previously in the default effect summary table for Example 1 (see Figure 5), and their interpretations are the same here.

The ‘CDE+PE’ Decomposition

Whereas the ‘NDE+NIE’ decomposition is more related to the etiology of the change in delinquent behavior due to the youth program, the ‘CDE+PE’ decomposition is more related to the intervention or policy-making issue regarding the youth program. By definition, the CDE is the effect of the youth program if there is an intervention such that everyone in the population of interest has his or her mediator variable (moral and social values) fixed to a specific level. By default, the CAUSALMED procedure uses the sample mean of the mediator as the intervention level. Figure 6 shows that this effect is –0.4666, which differs from the total effect by 0.0955. This means that the program effect under this intervention on the mediator will be slightly smaller in magnitude than the total effect—it reduces less delinquent behavior in this case. Figure 7 shows this reduction in terms of percentage (portion eliminated, or PE), which is about 17%. For further illustration of the use of the CDE and the PE estimates, see Example 4.

The ‘NDE+PIE+IMD’ Decomposition

The natural indirect effect (NIE) in the ‘NDE+NIE’ decomposition consists of the pure mediation effect (PIE) and the mediated interaction effect (IMD). Splitting the NIE thus forms the ‘NDE+PIE+IMD’ three-way decomposition. Figure 6 shows that the PIE value is –0.0525, and Figure 7 shows that the corresponding percentage due to pure mediation is about 9%. In addition, the IMD value is –0.0618, and the corresponding percentage is about 11%. But because both the 95% confidence intervals for the IMD effect and percentage cover the zero point, the mediated interaction effect is not significantly different from 0.

The ‘TDE+PIE’ Decomposition

Combining the NDE and the IMD effect in the preceding ‘NDE+PIE+IMD’ three-way decomposition yields this ‘TDE+PIE’ two-way decomposition. Figure 6 shows that the total direct effect (TDE) is –0.5096 and the corresponding percentage of the total effect is about 91%, which is quite substantial. This also means that pure mediation accounts for less than 10% of the total effect.

The 'CDE+PIE+PAI' Decomposition

The most interesting effect in this three-way decomposition is perhaps the "portion attributed to interaction" (PAI) component, which pools the mediated interaction (IMD) and reference interaction (IRF) together. Because both the 95% confidence intervals for PAI and its corresponding percentage cover the zero point, the interaction effect between the youth program and the moral and social values is not evident.

The Four-Way Decomposition: 'CDE+IRF+IMD+PIE'

Finally, Figure 6 and Figure 7 show the four-way decomposition (VanderWeele 2014) of the effects and their percentages. All four of these effects have been discussed previously in various decompositions. In fact, you can view all the lower-order decompositions as some particular collapsing of the component effects of the four-way decomposition. The conceptual advantage of the four-way decomposition is that it delineates the contributions due to mediation but not interaction (PIE), interaction but not mediation (IRF), both mediation and interaction (IMD), and neither mediation nor interaction (CDE). In the current case, both IMD and IRF percentages are not statistically significant.

Example 3. Conditional Causal Mediation Effects

By default, PROC CAUSALMED evaluates the causal mediation effects at the "averaged" sample covariate values. These *overall* effects are useful for a general assessment of the mediation effects and their percentages. However, researchers sometimes want to examine the conditional causal mediation effects, given particular levels of covariates of interest. This example illustrates how you can use the EVALUATE statement to study conditional mediation effects.

Using the same data as in Examples 1 and 2, you want to examine whether the mediation patterns are the same for different subpopulations (or subgroups). In the following statements, you specify the same outcome and mediator models as those in Examples 1 and 2, but you add four EVALUATE statements that define specific subpopulations for later comparisons:

```
proc causalmed data=Youth;
   class Program(descending) Gender SES Introversion;
   model    Delinquency = Program Value Program*Value;
   mediator Value       = Program;
   covar    Support GPA Value0 Delinquency0 Gender SES Introversion;
   evaluate 'Delinquency0=2'  Delinquency0=2;
   evaluate 'Delinquency0=5'  Delinquency0=5;
   evaluate 'Delinquency0=2 Support=4 Value0=21' Delinquency0=2 Support=4 Value0=21;
   evaluate 'Delinquency0=5 Support=1 Value0=15' Delinquency0=5 Support=1 Value0=15;
run;
```

The levels for the covariates that are specified in these EVALUATE statements allow for the following two comparisons of causal mediation patterns:

- low versus high initial delinquency subpopulations
- low-risk versus high-risk subpopulations

For the first comparison, you define the low and high initial delinquency subpopulations by using the minimum and maximum sample values of **Delinquency0** (see Figure 2). Hence, the first EVALUATE statement specifies `Delinquency0=2`, and the second EVALUATE statement specifies `Delinquency0=5`. The quoted strings in these statements are used as labels in the output results.

For the second comparison, you define the low-risk and high-risk subpopulations of the teenagers according to the covariate levels summarized in Table 1. The low and high levels for these covariates are defined by the maximum and minimum sample values, which are shown in Figure 2 and specified in the last two EVALUATE statements.

Table 1 Subpopulation Definitions

Covariate	Low-Risk Subpopulation	High-Risk Subpopulation
Delinquency0 (initial frequency of delinquent behavior)	Low	High
Support (level of family support)	High	Low
Value0 (initial moral and social values)	High	Low

Figure 8 and Figure 9 show the effect summary tables for the first comparison. It appears that the total effect, NDE, NIE, percentage mediated, and other statistics in these two figures are quite similar. Therefore, the causal mediation patterns for the low and high initial delinquency subpopulations differ minimally.

Figure 8 Low Initial Delinquency

Summary of Effects: Delinquency0=2

	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.5654	0.0410	-0.6458	-0.4850	-13.78	<.0001
Controlled Direct Effect (CDE)	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
Natural Direct Effect (NDE)	-0.4512	0.0563	-0.5615	-0.3408	-8.01	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	20.2067	5.9857	8.4751	31.9384	3.38	0.0007
Percentage Due to Interaction	8.1867	4.5840	-0.7978	17.1712	1.79	0.0741
Percentage Eliminated	17.4680	4.8956	7.8728	27.0632	3.57	0.0004

Figure 9 High Initial Delinquency

Summary of Effects: Delinquency0=5

	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.5613	0.0420	-0.6436	-0.4790	-13.37	<.0001
Controlled Direct Effect (CDE)	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
Natural Direct Effect (NDE)	-0.4471	0.0580	-0.5607	-0.3334	-7.71	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	20.3539	6.1013	8.3957	32.3122	3.34	0.0008
Percentage Due to Interaction	7.5180	4.3369	-0.9823	16.0182	1.73	0.0830
Percentage Eliminated	16.8669	4.7634	7.5308	26.2029	3.54	0.0004

Figure 10 and Figure 11 show the conditional causal mediation results in the low-risk and high-risk subpopulations. Clearly, the total program effect for the low-risk subpopulation (–0.6645) is larger in magnitude than that of the high-risk subpopulation (–0.4510). The percentage of program effect that is mediated by the moral and social values in the low-risk subpopulation (17%) is smaller than that of the high-risk subpopulation (25%). This is perhaps conceivable because on average the low-risk subpopulation already has a higher level of **Value0**, so the program effect is less likely to be mediated by further enhancing the level of **Value**.

Figure 10 Low-Risk Subpopulation

Summary of Effects: Delinquency0=2 Support=4 Value0=21

	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.6645	0.0541	-0.7704	-0.5585	-12.29	<.0001
Controlled Direct Effect (CDE)	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
Natural Direct Effect (NDE)	-0.5502	0.0450	-0.6385	-0.4620	-12.22	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	17.1935	4.0870	9.1831	25.2040	4.21	<.0001
Percentage Due to Interaction	21.8778	10.1947	1.8965	41.8590	2.15	0.0319
Percentage Eliminated	29.7750	9.0158	12.1043	47.4458	3.30	0.0010

Figure 11 High-Risk Subpopulation

Summary of Effects: Delinquency0=5 Support=1 Value0=15						
	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.4510	0.0828	-0.6132	-0.2888	-5.45	<.0001
Controlled Direct Effect (CDE)	-0.4666	0.0511	-0.5668	-0.3664	-9.13	<.0001
Natural Direct Effect (NDE)	-0.3368	0.1054	-0.5434	-0.1302	-3.20	0.0014
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	25.3315	10.5041	4.7439	45.9191	2.41	0.0159
Percentage Due to Interaction	-15.0986	10.6265	-35.9262	5.7289	-1.42	0.1554
Percentage Eliminated	-3.4635	11.4229	-25.8519	18.9250	-0.30	0.7617

To make these conditional comparisons easier, PROC CAUSALMED provides keywords for specifying the covariate levels in the EVALUATE statement. For example, instead of finding the actual minimum and maximum values for the continuous covariates in the preceding specification, you can use the MIN and MAX keywords in the EVALUATE statements, as shown in the following examples:

```
evaluate 'Low Delinquent Subpopulation'  Delinquency0=min;
evaluate 'High Delinquent Subpopulation' Delinquency0=max;
evaluate 'Low-Risk Subpopulation'        Delinquency0=min Support=max Value0=max;
evaluate 'High-Risk Subpopulation'       Delinquency0=max Support=min Value0=min;
```

Example 4. Interpreting Controlled Direct Effects and Percentage Eliminated

One novel concept in the counterfactual framework is the definition of the controlled direct effect (CDE) (Robins and Greenland 1992; Pearl 2001). Whereas traditional mediation analysis emphasizes more about the decomposition of the total effect, the controlled direct effect (CDE) and percentage eliminated (PE, originally an abbreviation of “portion eliminated”) are more useful constructs from a policy-making or intervention perspective. Examples 1 and 2 mainly covered effect decompositions. This example focuses more on the interpretations of the CDE and the PE.

By default, PROC CAUSALMED evaluates all causal mediation effects at the sample mean value of the continuous mediator (or the baseline/reference level of the categorical mediator). Because the mediator, **Value**, in Example 2 is a continuous variable, PROC CAUSALMED uses its sample mean value, 20.08375 (see Figure 2), to evaluate the CDE shown in Figure 5. You can obtain a simple verification of this default behavior by explicitly setting the controlled mediator level in the first EVALUATE statement of the following specification, which includes other EVALUATE statements for computing the CDE and the PE given other mediator and covariate levels:

```
proc causalmed data=Youth;
   class Program(descending) Gender SES Introversion;
   model    Delinquency = Program Value Program*Value;
   mediator Value       = Program;
   covar    Support GPA Value0 Delinquency0 Gender SES Introversion;
   evaluate 'Value=20.08375'  Value=20.08375;
   evaluate 'Value=15'        Value=15;
   evaluate 'Value=25'        Value=25;
   evaluate 'Value=25 Delinquency0=2 Support=4 Value0=21'
             Value=25 Delinquency0=2 Support=4 Value0=21;
   evaluate 'Value=25 Delinquency0=5 Support=1 Value0=15'
             Value=25 Delinquency0=5 Support=1 Value0=15;
run;
```

The first EVALUATE statement generates a summary table of effects (not shown here) that is exactly the same as that in Figure 5, verifying the default evaluation of the CDE and the PE.

The next two EVALUATE statements compute the CDE and the PE by using the sample minimum (`Value=15`) and maximum (`Value=25`), respectively, of the mediator variable. When this variable is evaluated at the minimum mediator value, Figure 12 shows that the CDE is –0.2678 and the PE is 52%. This CDE means that if an intervention sets everyone’s moral and social values at the minimum sample level (`Value=15`), the program effect would have been only –0.2678. Compared with the total effect, which is –0.5621, such an intervention on the mediator would make it less effective in reducing the delinquent behavior. Hence, there would be a reduction in the program effect by (1–(–0.2678/(–0.5621)))100% = 52%, which is what the "Percentage Eliminated" indicates in the output. Certainly, such an intervention is for illustration purposes only here—no one should seek a way to do such an intervention.

Figure 12 Controlled at Low Moral and Social Values

Summary of Effects: Value=15

	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.5621	0.0413	-0.6430	-0.4812	-13.62	<.0001
Controlled Direct Effect (CDE)	-0.2678	0.1386	-0.5395	0.003824	-1.93	0.0533
Natural Direct Effect (NDE)	-0.4479	0.0572	-0.5600	-0.3357	-7.83	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	20.3246	6.0692	8.4291	32.2201	3.35	0.0008
Percentage Due to Interaction	43.0183	23.6626	-3.3595	89.3961	1.82	0.0691
Percentage Eliminated	52.3537	22.7402	7.7837	96.9237	2.30	0.0213

In contrast, if an intervention sets everyone's moral and social values at the maximum sample level (`Value=25`), Figure 13 shows that the CDE is –0.6589, which is larger in magnitude than the total effect of –0.5621. The corresponding "Percentage Eliminated" is –17%. In fact, this negative PE should now be interpreted as the "percentage gained." That is, if an intervention on the mediator could be achieved to set everyone's moral and social values at this highest level, then you could expect an increase of 17% in the program effect, which means that it would reduce delinquent behavior even more. Certainly, such an intervention is desirable, but it would be up to the program administrator or policy maker to find a way to implement it. Causal mediation analysis merely provides theoretical answers to hypothetical situations.

Figure 13 Controlled at High Moral and Social Values

Summary of Effects: Value=25

	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.5621	0.0413	-0.6430	-0.4812	-13.62	<.0001
Controlled Direct Effect (CDE)	-0.6589	0.0826	-0.8208	-0.4970	-7.98	<.0001
Natural Direct Effect (NDE)	-0.4479	0.0572	-0.5600	-0.3357	-7.83	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	20.3246	6.0692	8.4291	32.2201	3.35	0.0008
Percentage Due to Interaction	-26.5506	14.6326	-55.2299	2.1287	-1.81	0.0696
Percentage Eliminated	-17.2152	16.2305	-49.0263	14.5959	-1.06	0.2888

The CDE and the PE concepts apply to *conditional* causal mediation effects as well. The last two EVALUATE statements in the preceding specification define low-risk and high-risk subpopulations in a way similar to what is done in Example 3 (see Table 1). However, the theoretical question of interest in the current example is how the CDE and the PE would differ in the low-risk and high-risk subpopulations if intervention could set the moral and social values at the highest level in both subpopulations.

For the low-risk subpopulation, Figure 14 shows that the CDE is –0.6589 and the PE is less than 1%. The small PE indicates that the program administrator does not need to worry about finding a more effective intervention for this subpopulation. Such an intervention, if it existed and were implemented, would not result in any meaningful changes in delinquent behavior (that is, small PE).

This is in sharp contrast to the high-risk subpopulation: Figure 15 shows that the CDE is still –0.6589 but the PE is now –46%. This large PE results from the fact that the conditional total program effect is much smaller than the CDE in magnitude for this subpopulation. This means that if an intervention could set the moral and social values of everyone in the high-risk subpopulation at the highest level, then an increase of 46% in program effect for reducing delinquent behavior would be expected. Again, it is up to the program administrator or policy maker to find ways to implement such an intervention. Unfortunately, the 95% confidence interval for the PE is quite wide and covers the zero point. Thus, the estimated PE is not reliable enough to recommend such an intervention. More data must be collected.

Figure 14 Controlled at High Moral and Social Values for the Low-Risk Subpopulation

Summary of Effects: Value=25 Delinquency0=2 Support=4 Value0=21						
	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.6645	0.0541	-0.7704	-0.5585	-12.29	<.0001
Controlled Direct Effect (CDE)	-0.6589	0.0826	-0.8208	-0.4970	-7.98	<.0001
Natural Direct Effect (NDE)	-0.5502	0.0450	-0.6385	-0.4620	-12.22	<.0001
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	17.1935	4.0870	9.1831	25.2040	4.21	<.0001
Percentage Due to Interaction	-7.0552	3.3749	-13.6699	-0.4405	-2.09	0.0366
Percentage Eliminated	0.8421	5.8392	-10.6025	12.2868	0.14	0.8853

Figure 15 Controlled at High Moral and Social Values for the High-Risk Subpopulation

Summary of Effects: Value=25 Delinquency0=5 Support=1 Value0=15						
	Estimate	Standard Error	Wald 95% Confidence Limits		Z	Pr > |Z|
Total Effect	-0.4510	0.0828	-0.6132	-0.2888	-5.45	<.0001
Controlled Direct Effect (CDE)	-0.6589	0.0826	-0.8208	-0.4970	-7.98	<.0001
Natural Direct Effect (NDE)	-0.3368	0.1054	-0.5434	-0.1302	-3.20	0.0014
Natural Indirect Effect (NIE)	-0.1142	0.0312	-0.1754	-0.05314	-3.66	0.0002
Percentage Mediated	25.3315	10.5041	4.7439	45.9191	2.41	0.0159
Percentage Due to Interaction	-57.7260	39.7249	-135.59	20.1334	-1.45	0.1462
Percentage Eliminated	-46.0908	39.8859	-124.27	32.0842	-1.16	0.2479

Summary

This paper introduces the CAUSALMED procedure and illustrates how you can use it to perform causal mediation analysis. Causal mediation analysis enables you to understand the mechanisms of a causal process and to decompose causal effects. The CAUSALMED procedure implements regression-based estimation methods, and it estimates causal mediation and related effects that are defined within a counterfactual framework.

A simple causal mediation analysis like the one performed in Example 1 gives you information about the direct and indirect effects of a treatment on an outcome. Further decompositions of the treatment effects, as illustrated in Example 2, provide additional information about the causal process that can be used to determine where a policy maker might attempt to intervene in the causal system. In considering potential interventions, it is important to look at the effects within the appropriate population. As Example 3 and Example 4 show, you can use the EVALUATE statement in the CAUSALMED procedure to investigate questions within custom-defined subpopulations or in different hypothetical situations.

For more information about the features of PROC CAUSALMED and for examples of causal mediation with binary responses and binary mediators, see the documentation of the CAUSALMED procedure (SAS Institute Inc. 2017). For theoretical development and interpretations of these scales and decompositions, see VanderWeele (2014, 2015). Causal mediation analysis is a fast-growing topic. The CAUSALMED procedure covers some of the most fundamental and reliable techniques in the field. For more information about additional topics, such as sensitivity analysis or different kinds of estimation methods, see Hong (2015) and VanderWeele (2015), among others.

REFERENCES

Baron, R. M., and Kenny, D. A. (1986). "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51:1173–1182.

Hong, G. (2015). *Causality in a Social World: Moderation, Mediation, and Spill-Over*. New York: John Wiley & Sons.

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.

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

SAS Institute Inc. (2017). *SAS/STAT 14.3 User's Guide*. Cary, NC: SAS Institute Inc. `http://go.documentation.sas.com/?docsetId=statug&docsetTarget=titlepage.htm&docsetVersion=14.3&locale=en`.

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. (2014). "A Unification of Mediation and Interaction: A 4-Way Decomposition." *Epidemiology* 25:749–761.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

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`.

ACKNOWLEDGMENTS

The authors are grateful to Bob Rodriguez, Clay Thompson, and Ed Huddleston of the Advanced Analytics Division at SAS for their valuable assistance in the preparation of this paper. The first author also thanks Dr. Linda Valeri and Dr. Tyler VanderWeele of Harvard University for their support and assistance in the development of the CAUSALMED procedure.

CONTACT INFORMATION

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

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