CCRAM Technical Report 022-04

Counterfactual/Potential Outcomes "Causal Mediation Analysis" with Treatment by Mediator Interaction Using PROCESS

Andrew F. Hayes
Haskayne School of Business and Department of Psychology
University of Calgary

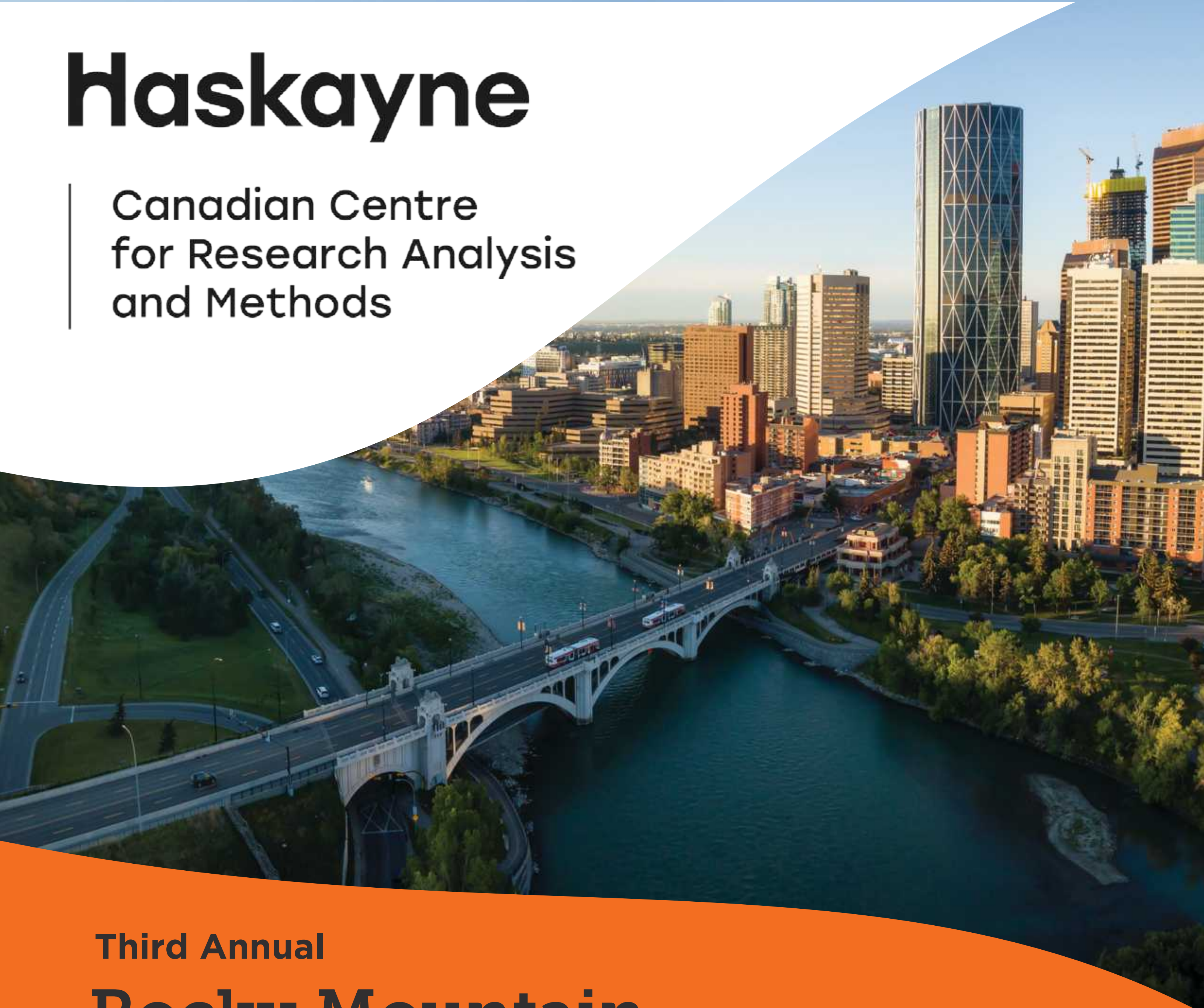
Draft date: 16 January 2025

Haskayne | Canadian Centre for Research Analysis and Methods

haskayne.ucalgary.ca/CCRAM

Haskayne

Canadian Centre
for Research Analysis
and Methods



Third Annual

Rocky Mountain Methodology Academy

July 15 – July 22, 2025

The Canadian Centre for Research Analysis and Methods hosts its annual Rocky Mountain Methodology Academy in Calgary, Alberta, Canada. Choose from ten courses taught by experts in social science data analysis and research methods. Between sessions and after class, explore Calgary and attend additional free lectures and events, including a day trip to Banff in the Canadian Rockies.

UNIVERSITY OF CALGARY

Session 1: July 15-16, 2025

Longitudinal Data Analysis and Visualization

DR. ANDREA HOWARD, PHD (CARLETON UNIVERSITY)

Data are often collected longitudinally, meaning the same variables are measured repeatedly over time, with the goal of understanding how variables change within and between people over time. This course provides a broad overview of various methods of quantifying, modeling and visualizing change in variables over time and how to test hypotheses about intraindividual and interindividual change.

Introduction to Social Network Analysis

DR. JENNY GODLEY, PHD (UNIVERSITY OF CALGARY)

Social network analysis examines the patterning of relationships between individuals and groups to understand social action. This course will cover the design, collection, analysis and interpretation of both whole and ego-centred network data.

Mediation Analysis

DR. ANDREW F. HAYES, PHD (UNIVERSITY OF CALGARY)

Mediation analysis is among the most widely used data analysis techniques in the social, health, and business sciences. In this course, you will learn about the fundamentals of estimation and inference about direct, indirect, and total effects and how mediation analysis is used to study the mechanism(s) by which effects operate. Application focuses on use of the popular PROCESS macro for SPSS, SAS, and R invented by the course instructor.

Intersession 1: July 17, 2025

There is no additional charge to attend these events. A reception will follow the keynote address.

Guest Lecture

DR. AMANDA MONTOYA, UNIVERSITY OF CALIFORNIA LOS ANGELES

Keynote Address

DR. ANDREA HOWARD, CARLETON UNIVERSITY

Session 2: July 18-19, 2025

Introduction to Structural Equation Modeling

DR. DOUG BAER, PHD (UNIVERSITY OF VICTORIA)

This course introduces the fundamentals of structural equation modeling as a general analytical tool, including how to set up measurement and structural models, latent variables, path analysis, definitions and quantification of model fit and the implementation of structural equation modeling in statistical software.

Interactions in Regression Analysis

DR. ANDREW F. HAYES, PHD (UNIVERSITY OF CALGARY)

Effects that scientists study rarely are uniform across people, context, or stimuli. Moderation analysis is used to examine the extent to which an effect depends on another variable, meaning it is moderated or the two variables interact. In the context of regression analysis and emphasizing the PROCESS macro invented by the instructor, this class covers the fundamentals of moderation analysis, including model estimation, interpretation, visualization, and probing interactions.

Experience Sampling and Implementation

DR. SABRINA THAI, PHD (BROCK UNIVERSITY)

Experience sampling methods are a powerful approach that allows researchers to examine how psychological phenomena unfold in daily life. This course introduces experience sampling methods and provides an overview of issues to be considering in using these designs, including signal frequency, sample size, power, question wording, compensation and recruitment, data handling, and technology implementation.

Intersession 2: July 20, 2025

Canadian Rockies Day Trip

Included with your registration while space is available, enjoy a day trip to Banff, Alberta, one of the most popular tourist destinations in Canada. Seats are limited, so register for the academy now to reserve your space on the bus.

Session 3: July 21-22, 2025

Introduction to Multilevel Modeling

DR. JASON RIGHTS, PHD (UNIVERSITY OF BRITISH COLUMBIA)

This course provides an introduction to multilevel modeling, with a focus on its application within the social, education, health and business sciences. Participants will learn fundamental statistical principles underlying multilevel modeling, a variety of techniques and methods that can be used in many different research contexts and how to appropriately specify models and interpret results in practice.

Scale Development and Psychometrics

DR. JESSICA FLAKE, PHD (UNIVERSITY OF BRITISH COLUMBIA)

Researchers in the academic and private sectors often need to measure attitudes, intentions, satisfaction, or motivation. Because scale scores are used to make decisions and evaluate research outcomes, develop a product, or hire or promote an employee, researchers need to thoroughly evaluate their validity. This course covers how to develop, evaluate, and refine scales using modern psychometric methods.

Introduction to Mixed Methods Research

DR. CHERYL POTH, PHD (UNIVERSITY OF ALBERTA)

Mixed methods research requires specific integration of knowledge and skills that also leverage existing qualitative and quantitative skills. Participants in this course will learn how to distinguish credible mixed methods research and have opportunities to ask questions about recent integration practice advancements. Discussions of the many perceived (and real) integration challenges when designing, executing and disseminating mixed methods research will provide foundational understandings for participants to engage in the design of their own mixed methods research projects.

Latent Profile Analysis

DR. MATTHEW MCLARNON, PHD (MOUNT ROYAL UNIVERSITY)

Latent profile analysis is a family of statistical models that can be used to identify unobserved, heterogenous and qualitatively distinct subgroups in one's data. This course will provide participants with the theoretical and conceptual background and applied analytical skills needed to specify an appropriate analytical model, interpret the results and thoroughly address research questions using latent profile analysis.

Register now! Seats are limited.

To register for courses and for more information, visit **haskayne.ucalgary.ca/CCRAM/academy**

All courses will take place on the main campus of the University of Calgary.

While space is available, your registration includes a day trip by chartered bus to the town of Banff in the Canadian Rockies.

The more courses you attend the more you save!

→ Courses are $950 (CAD) each

→ Enroll in 2 courses: $1,800 (CAD) **SAVE 5%**

→ Enroll in 3 courses: $2,500 (CAD) **SAVE 10%**

→ Longitudinal Toolkit: $2,280 (CAD) **SAVE 20%**

→ Graduate students are eligible for an **additional 10% discount.**

Note: As courses run from 9am to 5pm each day, you can register for only one course per session. The Longitudinal Toolkit includes longitudinal data analysis and visualization, experience sampling and implementation, and introduction to multilevel modeling. Prices above do not include 5% federal goods and services tax (GST). Consult your bank for current exchange rates.

HASKAYNE SCHOOL OF BUSINESS
2500 University Drive NW
Calgary, AB T2N 1N4, CANADA

Canadian Centre for Research Analysis and Methods (CCRAM)

haskayne.ucalgary.ca/CCRAM
ccram@ucalgary.ca

Counterfactual/Potential Outcomes "Causal Mediation Analysis" with Treatment by Mediator Interaction Using PROCESS

Andrew F. Hayes
University of Calgary
Haskayne School of Business and Department of Psychology

By default, PROCESS estimates a simple mediation model (i.e., one mediator) using two ordinary least squares (OLS) linear regression equations, as below

$$\hat{M} = i_M + aX \tag{1}$$

$$\hat{Y} = i_Y + c'X + bM \tag{2}$$

where *X* is dichotomous or continuous and *M* and *Y* are continuous or otherwise deemed properly modeled using OLS regression. The direct effect of *X* of *Y* is estimated by *c*', the indirect effect of *X* on *Y* through *M* is estimated by *ab*, and the total effect of *X* is estimated as the sum of the direct and indirect effects, *c*' + *ab*. Extensions or variations on this basic model are implemented by PROCESS when there is more than one mediator, *X* is multicategorical, or when covariates are included to account for potential confounding of the relationships between variables in the model by other variables. For a detailed discussion of this model and its extensions as well as their implementation in PROCESS, see *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3rd edition).

The proper estimation of the direct and indirect effects as described above using equations 1 and 2 assumes the effect of *M* on *Y* is invariant across values or levels of *X*. In other words, it assumes no moderation of *M*'s effect on *Y* by *X*, or no *X by M interaction*. And given the symmetry property of interactions, this is equivalent to assuming that *X*'s direct effect on *Y* doesn't depend on *M*. As *X* is often some kind of experimental manipulation in which participants in a study are randomly assigned to values of *X* representing a "treatment" or a "control" condition, this assumption is sometimes called "no treatment by mediator interaction." The discussion that follows does not require that *X* is an experimental manipulation or even that it is categorical. *X* could be continuous or categorical with two or more values, and random assignment is not a requirement of the math.

As of version 3.4 released in August 2019, PROCESS can test the no *X* by *M* interaction assumption with the use of **xmtest** option described in *Introduction to Mediation, Moderation, and Conditional Process Analysis*. But it could not estimate a model that allows *X* by *M* interaction until the release in October 2022 of version 4.2. As of version 4.2, *X* by *M* interaction can be specified in a mediation model by adding **xmint=1** to the PROCESS command. This option is available **only** when estimating model 4 (i.e., for simple/single mediator or parallel multiple mediator models). When this option is used, the model of *Y* will include the product (or products, depending on the model) of *X* and *M* in the model of *Y*, resulting in a mediation model specified (in the single-mediator case) as

$$\hat{M} = i_M + aX \tag{3}$$

$$\hat{Y} = i_Y + c'X + bM + b_{XM}XM \tag{4}$$

Including the product of X and M in equation 4 changes the mathematical definition and labeling of the direct and indirect effects of X. This document uses the language common in the so-called "causal mediation analysis" literature, guided by the counterfactual or potential outcomes approach to computation of the effects of X in a mediation model.

In the presumed media influence example in Chapter 3 of *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3rd edition), the direct, indirect, and total effects of newspaper article placement (front page or interior page) on reactions to the article are estimated, with presumed media influence the mediator, using equations 1 and 2. That is, in that example, it is assumed that the effect of presumed media influence on reactions to the article is the same regardless of whether participants are told the article was to appear on the front page of the paper or in an internal supplement.

This assumption can be relaxed using the PROCESS command below in SPSS (black box), SAS (white box), and R (grey box). This command is identical to the one on page 93 of *Introduction to Mediation, Moderation, and Conditional Process Analysis* (3rd edition) but adds the **xmint=1** option (while excluding the **normal** option, as the Sobel test is not available in conjunction with the use of the **xmint** option).

(Example 1)

```
process y=reaction/x=cond/m=pmi/model=4/total=1/xmint=1/seed=31216.
```

```
%process(data=pmi,y=reaction,x=cond,m=pmi,model=4,total=1,xmint=1,seed=31216)
```

```
process(data=pmi,y="reaction",x="cond",m="pmi",model=4,total=1,xmint=1,
   seed=31216)
```

In this example, X, M, and Y in equations 3 and 4 are the variables in the pmi data file named `cond`, `pmi`, and `reaction`, respectively.

Whereas using equations 1 and 2, the indirect effect of X on Y through M is estimated as ab, with this specification allowing X by M interaction, the indirect effect of X on Y through M is estimated as

$$NIE = a(b + b_{XM}X_{cf})(X_{cf} - X_{ref}) \tag{5}$$

and called the *natural indirect effect* (*NIE*) or, by some, the *total natural indirect effect*. In equation 5, X_{cf} is the counterfactual state for X and X_{ref} is the reference state for X. In addition to the point estimate of the natural indirect effect, PROCESS will produce a percentile bootstrap confidence interval for inference.

The direct effect of X on Y is not c', as it would be using equations 1 and 2, but, instead,

$$NDE = [c' + b_{XM}(i_M + aX_{ref})]\,(X_{cf} - X_{ref}) \tag{6}$$

and is now called the natural direct effect of X (*NDE*) or, by some, the *pure natural direct effect*. Inference for the natural direct effect is provided with a standard OLS regression standard error, *t*- and *p*-value, and confidence interval.

The total effect of X (*TE*)is the sum of the natural indirect and direct effects:

$$TE = NIE + NDE$$

When X is dichotomous and there are no covariates in the model, the total effect *TE* is equivalently estimated using c in

$$\hat{Y} = i_Y + cX$$

as $c(X_{cf} - X_{ref})$, and PROCESS produces inference with standard OLS regression theory standard errors, *t*- and *p*-values and confidence intervals. But when X is continuous or the model includes covariates, $c(X_{cf} - X_{ref})$ is not equivalent to the sum of the natural direct and indirect effects. In this case, the total effect is estimated by adding the natural direct and indirect effects and PROCESS generates a bootstrap confidence interval for inference about this sum.

Notice that if we assume no X by M interaction by excluding XM from equation 4 (or, in other words, assuming its weight is zero) and that $X_{cf} = X_{ref} + 1$, then NIE and NDE in equations 5 and 6 simplify to the familiar *ab* and *c*′ that PROCESS would generate when the **xmint** option is not used.

Specification of Reference and Counterfactual Values of *X*

In the counterfactual tradition, *X*'s effect on *Y* is determined by the values of X that are used to define the reference state and the counterfactual state of X. When X is dichotomous, PROCESS by default sets X_{ref} to the numerically smallest code representing the two groups and X_{cf} is set to the numerically largest code representing the two groups. For example, if in the data a control and treatment group are coded $X = 0$ and 1, respectively, then PROCESS sets X_{ref} to 0 and X_{cf} to 1, meaning that the control group is the reference state and the treatment group is the counterfactual state. In the presumed media influence example above, PROCESS treats the interior condition as the reference state and the front page condition as the counterfactual state because these are coded 0 and 1 in the data, respectively.

If preferred, the reference group state can be specified with the use of the **xrefval** option, including the desired value of X_{ref} following an equals sign. For example, to make the front page condition the reference state in the example above, add **xrefval=1** to the PROCESS command. PROCESS will automatically set X_{cf} to the other group code when X is dichotomous, which in this example is $X = 0$, or the interior page condition.

When X is continuous (which PROCESS assumes if there are more than 2 values of X in the data and X is not specified as multicategorical using the **mcx** option), you must specify at least the reference value of X using the **xrefval** option in the PROCESS command. When only a single value is provided, PROCESS will automatically set the counterfactual value to $X_{cf} = X_{ref} + 1$. Alternatively, you can specify both the reference and counterfactual values of X, in that order. For example, using the economic stress mediation analysis example in Chapter 3 *of Introduction to Mediation, Moderation, and Conditional Process Analysis*, the PROCESS command

(Example 2)

```
process y=withdraw/x=estress/m=affect/model=4/xmint=1/total=1/xrefval=2,4.
```

```
%process(data=estress,y=withdraw,x=estress,m=affect,model=4,xmint=1,total=1,
   xrefval=2 4)
```

```
process(data=estress,y="withdraw",x="estress",m="affect",model=4,xmint=1,total=1,
   xrefval=c(2,4))
```

allows for interaction between economic stress and business-related depressed affect and sets the reference and counterfactual values for `estress` to X_{ref} = 2 and X_{cf} = 4. A failure to specify at least the reference value of X when X is continuous will produce an error that PROCESS will insist you fix before it will estimate the model.

Covariates

In mediation analysis, it is common to include covariates to account for potential confounding of the effects due to shared causal influence of confounders, with the covariates being measures of those potential confounding variables. This is accomplished by including the confounders on the right-hand side of the equations for M and Y, as in

$$\hat{M} = i_M + aX + \sum_{j=1}^{q} d_j U_j$$

$$\hat{Y} = i_Y + c'X + bM + b_{XM}XM + \sum_{j=1}^{q} f_j U_j$$

where q is the number of covariates and U_j denotes a covariate. The covariates can be included in the PROCESS command in the usual way by using the **cov=** option, listing the covariates in an arbitrary order following the equals sign as described in the PROCESS documentation.

The NIE is still

$$NIE = a(b + b_{XM}X_{cf})(X_{cf} - X_{ref})$$

but the NDE becomes dependent on specific values of the covariate(s), as the covariates appear in the formula for the natural direct effect:

$$NDE = [c' + b_{XM}(i_M + aX_{ref} + \textstyle\sum_{j=1}^{q} d_j U_j)]\,(X_{cf} - X_{ref})$$

Because the total effect is the sum of the direct and indirect effects, the total effect is therefore also conditioned on specific values of the covariates used in the computation of the NDE.

By default, PROCESS will set covariates to their sample means when calculating the natural direct and total effects of X, meaning that these represent the effect of X among cases average on the covariates. It is possible to condition the estimation of the direct and total effects of X on different values of the covariates by including the **coval** option in the PROCESS command, with a value for each of the covariates following the equals sign and the specific values of the covariates listed in the same order following **coval=** that the covariates are listed following **cov=**. For example, in Chapter 4 *of Introduction to Mediation, Moderation, and Conditional Process Analysis*, the economic stress example is extended to include entrepreneurial self-efficacy (`ese`), sex (`sex`), and length of time in business (`tenure`) as covariates. In addition to allowing for interaction between economic stress and business-related depress affect, the PROCESS command

(Example 3)

```
process y=withdraw/x=estress/m=affect/cov=ese sex tenure/model=4/xmint=1/total=1
  /xrefval=2,4/coval=3,1,5.
```

```
%process(data=estress,y=withdraw,x=estress,m=affect,cov=ese sex tenure,model=4,
   xmint=1,total=1,xrefval=2 4,coval=3 1 5)
```

```
process(data=estress,y="withdraw",x="estress",m="affect",cov=c("ese","sex",
   "tenure"),model=4,xmint=1,total=1,xrefval=c(2,4),coval=c(3,1,5))
```

estimates the indirect, direct, and total effects of economic stress (for reference and counterfactual values of economic stress of 2 and 4, respectively), with the direct and total effects conditioned on entrepreneurs who score 3 on entrepreneurial self-efficacy (`ese`=3), are male (`sex`=1), and who have been in business for 5 years (tenure=5).

When using the **xmint** option, all covariates are assigned to all equations. It is not possible to assign some covariates to one equation and others to a different equation, as is possible when using the **cmatrix** option in the absence of the **xmint** option.

More than One Mediator

The **xmint** option is available in only model 4, which is used for both simple mediation (one mediator) and multiple mediators arranged in parallel form. When using the **xmint** option with *k* mediators following **m=**, PROCESS estimates the model with *k* equations for *M* (one for each mediator)

$$\hat{M}_j = i_{M_j} + a_j X$$

and a single equation for *Y*

$$\hat{Y} = i_Y + c'X + \sum_{j=1}^{k} b_j M_j + \sum_{j=1}^{k} b_{XM_j} XM_j$$

In this parallel multiple mediator model, the specific natural indirect effect of *X* on *Y* through M_j is

$$NIE_j = a_j(b_j + b_{b_{XM_j}} X_{cf})(X_{cf} - X_{ref})$$

and PROCESS produces a bootstrap confidence interval for inference for each specific natural indirect effect. With no covariates in the model, the natural direct effect of *X* is

$$NDE = [c' + \sum_{j=1}^{k}[b_{XM_j}\,(i_{M_j} + a_j X_{ref})]]\,(X_{cf} - X_{ref})$$

The total effect of *X* is the sum of natural direct effect and the *k* specific natural indirect effects of *X*, with ordinary OLS theory used for inference unless the model contains covariates or *X* is continuous, in which case inference for the total effect is provided by PROCESS with a bootstrap confidence interval.

Covariates can be used in a multiple mediator model just as described for the simple mediation case earlier, and the **coval** option is available for multiple mediator models as well. Inclusion of covariates will make the NDE dependent on the covariates, as discussed earlier, as the formula for the natural direct effect becomes

$$NDE = [c' + \sum_{j=1}^{k}[b_{XM_j}\,(i_{M_j} + a_j X_{ref} + \sum_{j=1}^{q} d_j U_j)]]\,(X_{cf} - X_{ref})$$

with the default being the estimation of the natural direct effect when the covariates as set to the means.

When using the **xmint** option in a multiple mediator model, the *X* by *M* interaction is estimated for all mediators. PROCESS does not allow the specification of interaction between *X* and a subset of the mediators.

Controlled Direct Effect

When a mediation model includes *X* by *M* interaction, the direct effect of *X* is dependent on the mediator(s). Because a mediator is dependent on *X* in a mediation model, this complicates the estimation and interpretation of the direct effect. The definition of the natural direct effect as it is typically defined in the counterfactual mediation analysis literature (and as implemented in PROCESS and discussed in this document) is the direct effect of *X* when the mediator is set to the estimated mean of the mediator conditioned on $X = X_{ref}$. Another kind of direct effect, the *controlled* direct effect (*CDE*), is frequently defined as the direct effect of *X* when the mediator is set to the overall mean of the mediator, without conditioning that mean on a value of *X*.

PROCESS generates this controlled direct effect automatically, defined as

$$CDE = [c' + b_{XM}\bar{M}]\,(X_{cf} - X_{ref})$$

in a simple mediation model, or

$$CDE = [c' + \sum_{j=1}^{k} b_{XM_j}\,\bar{M}_j]\,(X_{cf} - X_{ref})$$

in a parallel multiple mediator model with *k* mediators. OLS standard errors, *t*- and *p*-values, and confidence intervals are provided for inference.

It is possible to estimate the controlled direct effect setting the mediator(s) to any desired value rather than relying on the default. To do so, use the **cdeval** option, specifying the desired value of the mediator or mediators following the equals sign. If the model contains more than one mediator, the values for the mediators in the **cdeval** option must be listed in the same order as the mediators are listed in the PROCESS command following **m=**.

For example, using the parallel multiple mediator model example in Chapter 5 of *Introduction to Mediation, Moderation, and Conditional Process Analysis*, the command below generates the controlled direct effect when `import`=4.5 and `pmi`=5.3.

(Example 4)

```
process y=reaction/x=cond/m=import pmi/model=4/total=1/xmint=1/cdeval=4.5 5.3.
```

```
%process(data=pmi,y=reaction,x=cond,m=import pmi,model=4,total=1,xmint=1,
  cdeval=4.5 5.3)
```

```
process(data=pmi,y="reaction",x="cond",m=c("import","pmi"),model=4,total=1,
  xmint=1,cdeval=c(4.5,5.3))
```

Because X is dichotomous with the two groups coded 0 (interior page condition) and 1 (front page condition), using the values of 3.9077 and 5.3769 in the **cdeval** option would produce a controlled direct effect that is equivalent to the natural direct effect because these are the means of `import` and `pmi` in the interior page condition, which would be treated as the reference by PROCESS using this code.

Multicategorical *X*

When *X* is a multicategorical variable with *g* groups and so specified using the **mcx** option in PROCESS, a mediation model allowing *X* by *M* interaction is estimated with the following equations when the **xmint** option is used:

$$\hat{M} = i_M + \sum_{j=1}^{g-1} a_j D_j$$

$$\hat{Y} = i_Y + \sum_{j=1}^{g-1} c'_j D_j + bM + \sum_{j=1}^{g-1} b_{D_jM} D_j M$$

where the $g - 1$ variables D_j are constructed using a multicategorical coding system representing the *g* groups (e.g., indicator coding). This will produce g –1 relative natural indirect and direct effects, relatively total effects, and relative controlled direct effects of *X*. For a discussion of relative effects, see Chapter 6 *of Introduction to Mediation, Moderation, and Conditional Process Analysis*. These relative effects are labelled in the PROCESS output for g –1 variables "X1", "X2", and so forth, up to "X_{g-1}". Each of these effects capture a comparison between a reference state and a counterfactual state, with those comparisons being dependent on the coding system used.

With indicator coding (option **mcx=1** in PROCESS), there are g –1 counterfactual groups, each of which is compared to a common "baseline" group that serves as the reference state group (with indicator coding, this "baseline group" is often called the "reference group", though use of that term is confusing in this context so that term is reserved here to refer to the reference state rather than the counterfactual state) That common reference state or baseline group is the group that receives all zeros on the g – 1 indicator variables. The g – 1 counterfactual groups are the remaining groups, where counterfactual group k is the group set to 1 on indicator variable D_k.

Sequential coding (option **mcx=2** in PROCESS), sometimes called adjacent categories coding, is useful when X is ordinal and so the groups can be ordered from low to high on something the grouping dimension variable X represents (e.g., levels of education completed; "low," "moderate," and "high" on some experimentally-manipulated dimension). There are g –1 counterfactual groups and g – 1 reference groups, but which group serves which role depends on the comparison being represented. If we order the g groups from low (ordinal position 0) to high (ordinal position g – 1) on the dimension X represents, the reference group for the comparison represented with X_k in the PROCESS output is the group in ordinal position k – 1 and the counterfactual group is the group in ordinal position k. For example, with three groups, the part of the effect of X captured by X1 in the PROCESS output represents the comparison between group 0 (reference state group) and 1 (counterfactual group) and the part of X's effect captured by X2 represents the comparison between group 1 (reference group state) and 2 (counterfactual group).

The relative natural indirect effect, relative natural direct effect, relative controlled direct effect, and relative total effect of X for the component of X's effect represented by X_k (where k = 1 to g –1) are estimated as

$$\text{relative } NIE_k = a_k[b + \sum_{j=1}^{g-1} b_{D_jM}(D_j|\, cf)]$$

$$\text{relative } NDE_k = c'_k + \sum_{j=1}^{g-1} b_{XM_j}\,[i_M + \sum_{j=1}^{g-1} a_j(D_j|\, ref)]$$

$$\text{relative } CDE_k = c'_k + \sum_{j=1}^{g-1} b_{XM_j}\,\bar{M}$$

$$\text{relative } TE_k = NIE_k + NDE_k$$

where (D_j | *ref)* and (D_j | *cf*) j = 1 to g – 1, are the values of D_j for the reference group and the counterfactual group for that comparison, respectively.

PROCESS provides inferential statistics using standard OLS-theory *t*- and *p*-values and confidence intervals for the relative direct and relative controlled direct effects of X, as well as for the relative total effects when there are no covariates in the model. Inference is available through bootstrapping for the relative natural indirect effects, as well as the relative total effects in models with covariates.

The discussion and equations above assume a single mediator and no covariates. PROCESS will also estimate a parallel multiple mediator model with a multicategorical X as well as models with covariates. This modifies the equations above, which become substantially more complex, but the modifications are consistent with the earlier discussion in this document of these topics. The **coval** and **cdeval** options are available when X is multicategorical and are used in the PROCESS command just as described earlier.

Note that **mcx** options 3 (Helmert coding), 4 (effect coding), and 5 (custom codes) are not available with the use of the **xmint** option.

For example, using the example in Chapter 6 of *Introduction to Mediation, Moderation, and Conditional Process Analysis*, the command below estimates the relative natural direct effects, relative natural indirect effects, relative total effects, and relative controlled direct effects of protesting on evaluation of the attorney through perceived appropriateness of her response.

(Example 5)

```
process y=liking/x=protest/m=respappr/model=4/total=1/xmint=1/mcx=1.
```

```
%process(data=protest,y=liking,x=protest,m=respappr,model=4,total=1,xmint=1,
  mcx=1)
```

```
process(data=protest,y="liking",x="protest",m="respappr"),model=4,total=1,
  xmint=1,mcx=1)
```

Features Not Available in Conjunction with the xmint Option

Some of the options ordinarily available when estimating a mediation model are not, as of the release of version 4.2, available for use in conjunction with the **xmint** option. These include **normal** (Sobel test), **contrast** (comparing indirect effects), **effsize** (standardized scaling of effects), **mc** (Monte Carlo confidence intervals), and options for assignment of covariates to equations including **cmatrix** and **covmy**. Models with an *X* by *M* interaction cannot be edited using the procedures discussed in Appendix B of *Introduction to Mediation, Moderation, and Conditional Process Analysis*.

Equivalence between PROCESS, SAS PROC CAUSALMED, mediate in Stata, and Mplus

Comparable code for the models in examples 1, 2, 3, and 5 using SAS, Mplus, and/or Stata are provided below. Point estimates of effects will largely be identical. Small differences in standard errors will exist as a result of differences in estimation methods used compared to PROCESS (maximum likelihood versus ordinary least squares). In all the Mplus examples above, the exclamation points in the Mplus code should be eliminated when bootstrap inference is desired.

<u>Example 1</u>

SAS:

```
proc causalmed data=pmi pmedmod poutcomemod;
  model reaction=cond pmi cond*pmi;
  mediator pmi=cond;
  evaluate 'effect' _control=0 _treatment=1;
  bootstrap bootci (perc) nsamples=5000;
run;
```

Mplus:

```
DATA:
  file=C:\data\pmi.csv;
VARIABLE:
```



```
  names are cond pmi import reaction gender age;
  usevariables are cond pmi reaction mx;
DEFINE:
  mx=cond*pmi;
ANALYSIS:
  !bootstrap=5000;
MODEL:
  pmi on cond;
  reaction on cond pmi mx;
MODEL INDIRECT:
  reaction MOD pmi mx cond(1 0);
OUTPUT:
  !cinterval (bootstrap)
```

Stata:

```
mediate (reaction)(pmi)(cond),aequations
estat cde,mvalue(5.6016)
```

Example 2

SAS:

```
proc causalmed data=estress pmedmod poutcomemod;
  model withdraw=estress affect estress*affect;
  mediator affect=estress;
  evaluate 'effect' _control=2 _treatment=4;
  bootstrap bootci (perc) nsamples=5000;
run;
```

Mplus:

```
DATA:
  file=C:\data\estress.csv;
VARIABLE:
  names are tenure estress affect withdraw sex age ese;
  usevariables are estress affect withdraw mx;
DEFINE:
  mx=estress*affect;
ANALYSIS:
  !bootstrap=5000;
MODEL:
  affect on estress;
  withdraw on estress affect mx;
MODEL INDIRECT:
  withdraw MOD affect mx estress(4 2);
OUTPUT:
  !cinterval (bootstrap);
```

Stata:

```
mediate (withdraw)(affect)(estress,continuous(2 4)),aequations
estat cde,mvalue(1.5981)
```

Example 3

SAS:

```
proc causalmed data=estress pmedmod poutcomemod;
  model withdraw=estress affect estress*affect;
  mediator affect=estress;
  covar ese sex tenure;
  evaluate 'effect' _control=2 _treatment=4 ese=3 sex=1 tenure=5;
  bootstrap bootci (perc) nsamples=5000;
run;
```

Mplus:

```
DATA:
  file=C:\data\estress.csv;
VARIABLE:
  names are tenure estress affect withdraw sex age ese;
  usevariables are estress affect withdraw mx esec tenurec sexc ;
DEFINE:
  mx=estress*affect;
  esec=ese-3;
  sexc=sex-1;
  tenurec=tenure-5;
ANALYSIS:
  !bootstrap=5000;
MODEL:
  affect on estress esec sexc tenurec;
  withdraw on estress affect mx esec sexc tenurec;
MODEL INDIRECT:
  withdraw MOD affect mx estress(4 2);
OUTPUT:
  !cinterval (bootstrap);
```

Stata:

Stata cannot estimate controlled and natural direct effects when covariates are set to user specified values.

Example 5

Mplus and PROC CAUSALMED have no options for specifying a multicategorical X.

Stata:

```
mediate (liking)(respappr)(protest),aequations
estat cde,mvalue(4.8663)
```