Errors-in-Variables Mediation Analysis with PROCESS

Andrew F. Hayes
University of Calgary

Paul D. Allison
Statistical Horizons LLC

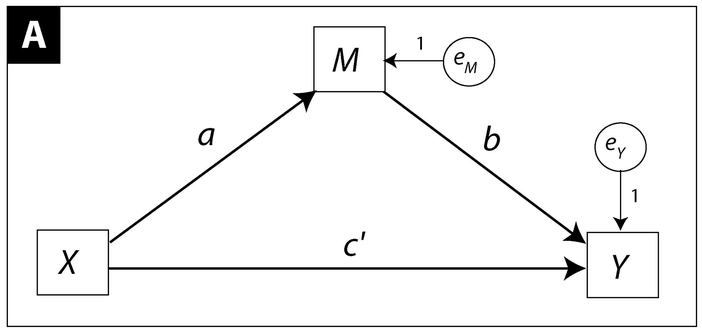
Sean Alexander
University of Calgary

With the release of PROCESS version 5, there is no excuse for not properly accounting for random measurement error in a mediation analysis.

What We (Usually) Do vs. What We Want

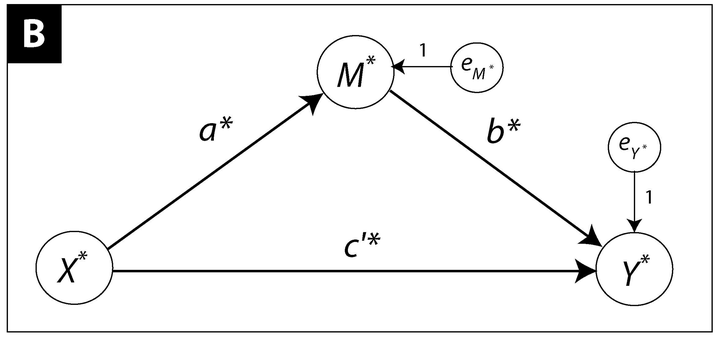
In a mediation analysis, the investigator quantifies the indirect (ab), direct (c'), and total effects (c' + ab) of X on Y, with X's effect operating at least in part through mediator M. In the typical analysis, X, M, and Y are ***observed scores*** *containing random measurement error (RME).*

A



But what we usually want to estimate are the effects of X on Y through M involving the ***true scores*** *X*, M*, and Y*. a*, b*, and c* in* **B** *below are not likely the same as a, b, and c' in* **A** *above when variables in* **A** *are contaminated with RME.*

B



Problems and Solutions

Indirect, direct, and total effects estimated using observed variable models as in **A** *are usually biased upward and/or downward when M and/or X is measured with less-than-perfect reliability, as they typically are.*

- Point estimates of effects, confidence intervals, and inferential tests of the effect we actually care about in **B** are likely to be inaccurate when estimating **A** instead of **B**.

There are better approaches to dealing with RME than just ignoring it (which is typical practice)

Latent Variable Structural Equation Modeling (SEM)

- Multiple indicator latent variable (MILV) modeling can be used when scores on more than one indicator of latent variables are available to estimate latent variable effects. No estimate of reliability is needed.
- Single indicator latent variable (SILV) modeling can be used, with the observed scores being the sole indicator of latent variables. A constraint is imposed in estimation to the error variance of indicators using an estimate of the reliability of the observed scores.

Errors-in-Variables Regression

- Unknown to most social scientists, but very effective and much easier to implement in your work (and teach in your classes) than latent variable SEM.
- Involves an (automated) modification of the data on the right-hand sides of equations, subtracting out RME from observed scores.
- Like the SILV approach, EIV regression requires an estimate of the reliability of the observed scores. Such estimates are readily available, and most researchers know how to generate them.
- **Now available in the freely-available PROCESS macro** for SPSS, SAS, and R already being widely used in mediation analysis.
- **EIV regression will produce largely the same results as latent variable SEM with much less effort.**

Which Would You Rather Do?

Hard(er): *Use an SEM program.*

Here is an example program using lavaan in R that estimates the parallel multiple mediator model example in the paper using the SILV approach.

```
library(lavaan)
model.silv<-"Lspa=~spa
             Lpes=~pes
             Lsobbs=~sobbs
             Lpacs=~pacs
             Lsobbs~i1*1+a1*Lpes
             Lpacs~i2*1+a2*Lpes
             Lspa~i3*1+b1*Lsobbs+b2*Lpacs+cp*Lpes
             Lsobbs~~Lpacs
            a1b1 :=a1*b1
             a2b2 :=a2*b2
             c := a1*b1+a2*b2+cp
             totind := a1*b1+a2*b2
             C1 := a1*b1-a2*b2
             #(1-reliability) multiplied by observed variances
             spa~~((1-0.91)*0.160231020753164)*spa
             pes~~((1-0.75)*0.547490624374402)*pes
             pacs~~((1-0.73)*0.590323019774451)*pacs
             sobbs~~((1-0.89)*0.540627166218156)*sobbs"
modelp<-sem(model.silv,data=nature)
summary(modelp,rsquare=T)
set.seed(7234)
modelp<-sem(model.silv,data=photo,se="bootstrap",bootstrap=5000)
parameterestimates(modelp,boot.ci.type="perc")
```

Easy: *Use the errors-in-variables regression routine in PROCESS as of version 5 (available for SPSS, SAS, and R).*

Just write your PROCESS command as usual, but include **estimates of the reliabilities** of the variables in the model. The code below estimates the same model as the SEM code above and produces comparable results.

SPSS:
```
process y=spa/x=pes/m=sobbs pacs/total=1/relx=0.75/relm=0.89,0.73
  /contrast=1.
```

SAS:
```
%process (data=photo,y=spa,x=pes,m=sobbs pacs,total=1,relx=0.75,
  relm=0.89 0.73,contrast=1)
```

R:
```
process(photo,y="spa",x="pes",m=c("sobbs","pacs"),relx=0.75,
  relm=c(0.89,0.73),contrast=1)
```

Presented at the annual convention of the Canadian Psychological Association, June 2025, in St. John's Newfoundland