Presented at the annual convention of the Canadian Psychological Association, June 2025, St. John's NL. This is a slightly edited version of the original presentation.

Canadian Centre for Research Analysis and Methods
Haskayne | University of Calgary

Consequences of and Remedies for Unaccounted-for Random Measurement Error in Mediation Analysis of Clinical Trials and Other Two-Group Comparisons

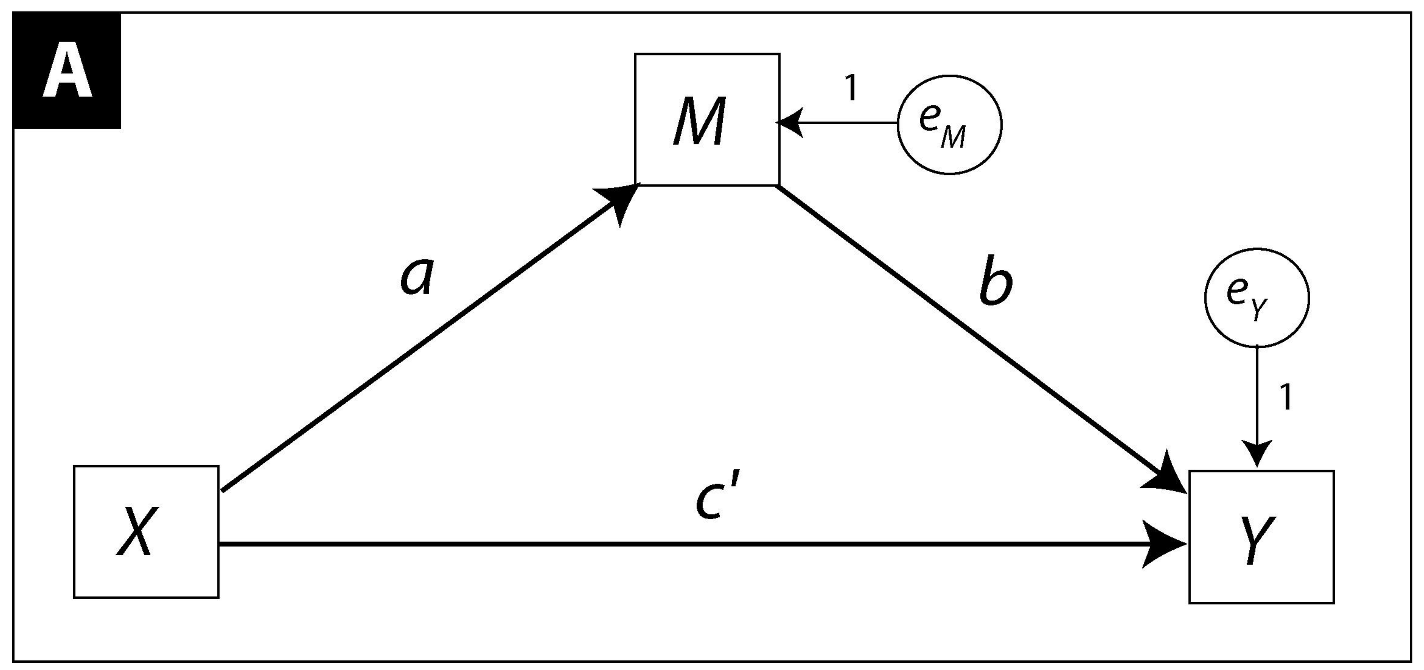
Andrew F. Hayes and Carole A. Lunney
University of Calgary

DCDL
DETERMINANTS OF CHILD DEVELOPMENT LAB

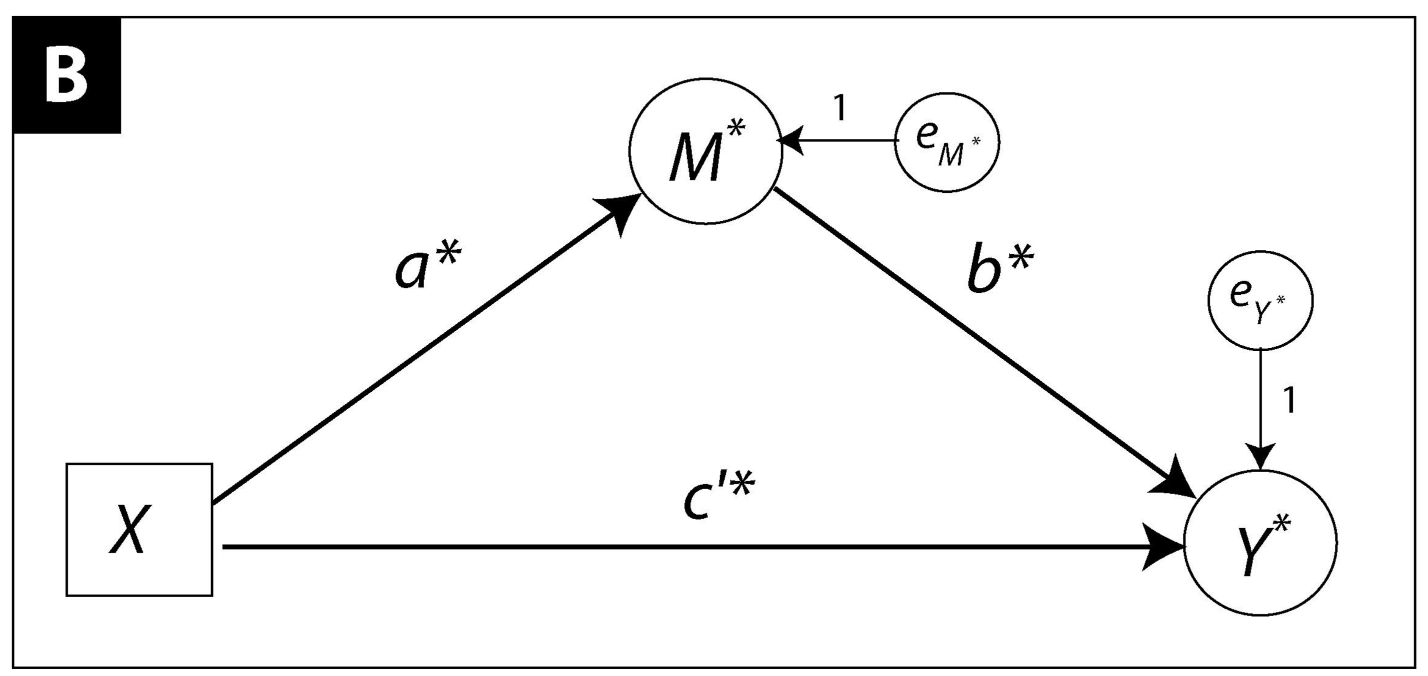
Indirect and direct effects are misestimated when random measurement error (RME) in the mediator is not acknowledged. PROCESS v5 can help.

Mediation analysis

In a mediation analysis, the investigator is interested in the direct, indirect, and total effects of X on Y. ***In the typical clinical trial with X randomly assigned or known exactly, there is no measurement error in X****, but mediator M (and perhaps Y) are usually* ***observed scores*** *containing random measurement error (RME). The mediation model typically estimated is*



But what we usually care about are the direct, indirect, and total effects of X on Y involving the ***true scores*** *M* and Y*, had M and Y been measured perfectly without RME. In* **B** *below, b* and c'* are not the same as b and c' in* **A** *above when M in* **A** *contains RME.*



Derivations and Implications

It can be shown than when estimating **A** *instead of* **B** *with no measurement error in X but M measured with reliability ρ_M (and also making a few other assumptions people routinely make when conducting a mediation analysis)*

$$E(a) = a^*$$

$$E(b) = b^*(\rho_M - r_{XM}^2)/(1 - r_{XM}^2)$$

$$E(ab) = a^*b^*(\rho_M - r_{XM}^2)/(1 - r_{XM}^2)$$

$$E(c') = c'^* + a^*b^*[(1-\rho_M)/\ (1 - r_{XM}^2)]$$

$$E(c' + ab) = E(c') + E(ab) = c'^* + a^*b^*$$

This is all true regardless of the reliability of measurement of *Y*

This means that ignoring measurement error in M in the analysis by estimating **A** *will tend to result in…*

- …an estimate of the total effect of *X* that is correct. GOOD
- …an estimate of the indirect effect of *X* that is *typically* closer to zero than reality. It will *usually* be attenuated toward zero. NOT GOOD
- …an estimate of the direct effect of *X* that is under- or overestimated in magnitude, depending on the actual direct and indirect effects. NOT GOOD

Your estimate of the direct effect of X could be…

- …opposite in sign of the actual direct effect of *X*.
- …nonzero when the direct effect is actually zero.
- …zero when the direct effect is actually nonzero.
- …larger or smaller in magnitude than the actual direct effect.

MORAL: Expect that your estimates of direct and indirect effects of *X* in a mediation analysis will be wrong if you don't address random measurement error in the mediator.

Fixing the Problem

The hard way using structural equation modeling:

Here is an example program using lavaan in R that estimates a *single-indicator latent variable* model of the effect of cognitive behavioural therapy vs. therapy as usual ("cbt") on PTSD symptoms ("ptsd") through social support ("support"), with social support measured with reliability 0.7.

```
library(lavaan)
cbt<-read.table("cbt.csv", sep=",",header=TRUE)
library(lavaan)
model.silv<-"Lsupport=~support
             Lsupport~a*cbt
             ptsd~b*Lsupport+cp*cbt
             ab :=a*b
             c := a*b+cp
             support~~((1-0.70)*4.35)*support"
modelp<-sem(model.silv,data=cbt)
summary(modelp,rsquare=T)
set.seed(27654)
modelp<-sem(model.silv,data=cbt,se="bootstrap",bootstrap=5000)
parameterestimates(modelp,boot.ci.type="perc")
```

The easy way using the PROCESS macro, version 5

As of version 5, PROCESS for SPSS, SAS, and R **implements errors-in-variables regression** that fixes this problem with an accurate estimate of the reliability of your measurements of the mediator.

The PROCESS code below estimates the same model as the lavaan code above and generates (largely) identical results:

SPSS
```
process y=ptsd/x=cbt/m=support/relm=0.7.
```

SAS:
```
%process(data=cbt,y=ptsd,x=cbt,m=support,relm=0.7)
```

R
```
process(data=cbt,y="ptsd",x="cbt",m="support",relm=0.7)
```