# Mediation for Count Outcomes: Equivalence of the Mediated Effect


Stefany J. Coxe and David P. MacKinnon

Arizona State University


## Introduction

This study examines the combination of two methods that are commonly used in many areas of the social sciences: Poisson regression for count outcomes and statistical mediation analysis. Count outcomes, addressed within the generalized linear model (GLiM) framework with Poisson regression, are of increasing interest in psychology and related behavioral sciences (Coxe, West, and Aiken, 2009). Examples of count variables that are relevant to prevention research include number of depressive symptoms (Schaffer et al., 2000), number of alcoholic drinks per day (Armeli et al., 2005), and number of re-admissions to alcohol detoxification programs (Shanahan et al., 2005).

Mediation analysis (Judd and Kenny, 1981; MacKinnon, 2008) allows a researcher to examine the causal chain through which a one variable has its effect on another variable. The simplest case of a mediation analysis involves 3 variables: X, M, and Y. The variable M mediates the effect of X on Y; that is, it is hypothesized that X causes M and then M causes Y. Mediation models are used to evaluate the mechanism by which an intervention has an effect on an outcome; for example, an intervention is theorized to cause changes in beliefs, which in turn cause changes in behavior.

Mediation models can be estimated using standard regression software; three regression equations are estimated:

(1) Y = *c*X + e1

(2) Y = *b*M + *c'*X + e2

(3) M = *a*X + e3

When both M and Y are continuous and standard linear regression is used, the mediated effect is estimated with two equivalent expressions in the linear model framework: *a*b* or *c-c'* (MacKinnon, Warsi, & Dwyer, 1995). When the outcome Y is a count variable, equations (1) and (2) are estimated as GLiMs, namely Poisson regression models. MacKinnon, Lockwood, Brown, Wang, & Hoffman (2007) found that when Y is binary and logistic regression is used to estimate equations (1) and (2), *a*b* and *c-c'* are not algebraically equivalent.

## Method

This simulation study varied the a, b, and c' paths of the mediation model, as well as sample size. X and M were generated as continuous, conditionally normally-distributed variables, while Y was a conditionally Poisson-distributed count variable, as shown by the equations below.

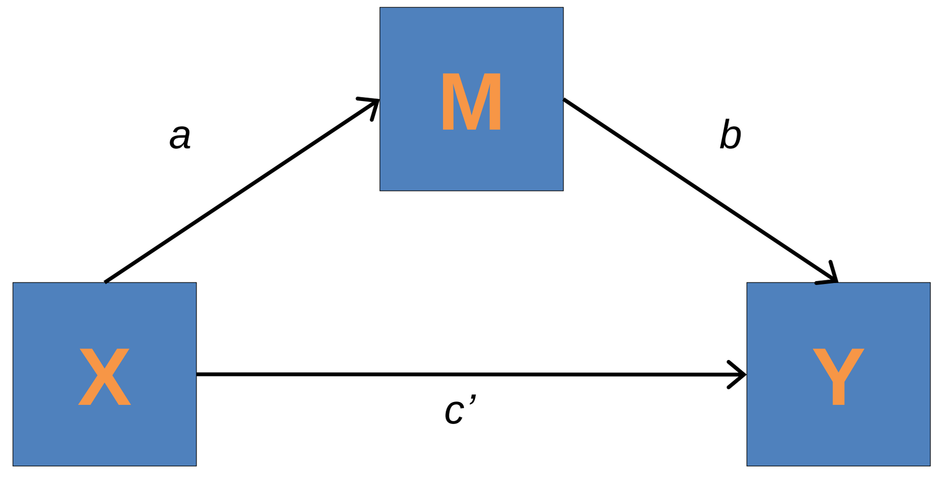
Zero, small, medium, and large effect sizes for the *a* path were 0, 0.14, 0.39, and 0.59. Zero, small, medium, and large effect sizes for the count outcome (*b* and *c'* paths) correspond to multiplicative changes of 1, 1.5, 2, or 5 times for a 1-unit change in X. Sample sizes of 100, 250, 500, and 1000 were examined. Each of the 256 conditions (4 a paths X 4 b paths X 4 c' paths X 4 sample sizes) were replicated 500 times.

<u>Data generation equations</u>

$$X = e_{Normal\,(0,1)}$$

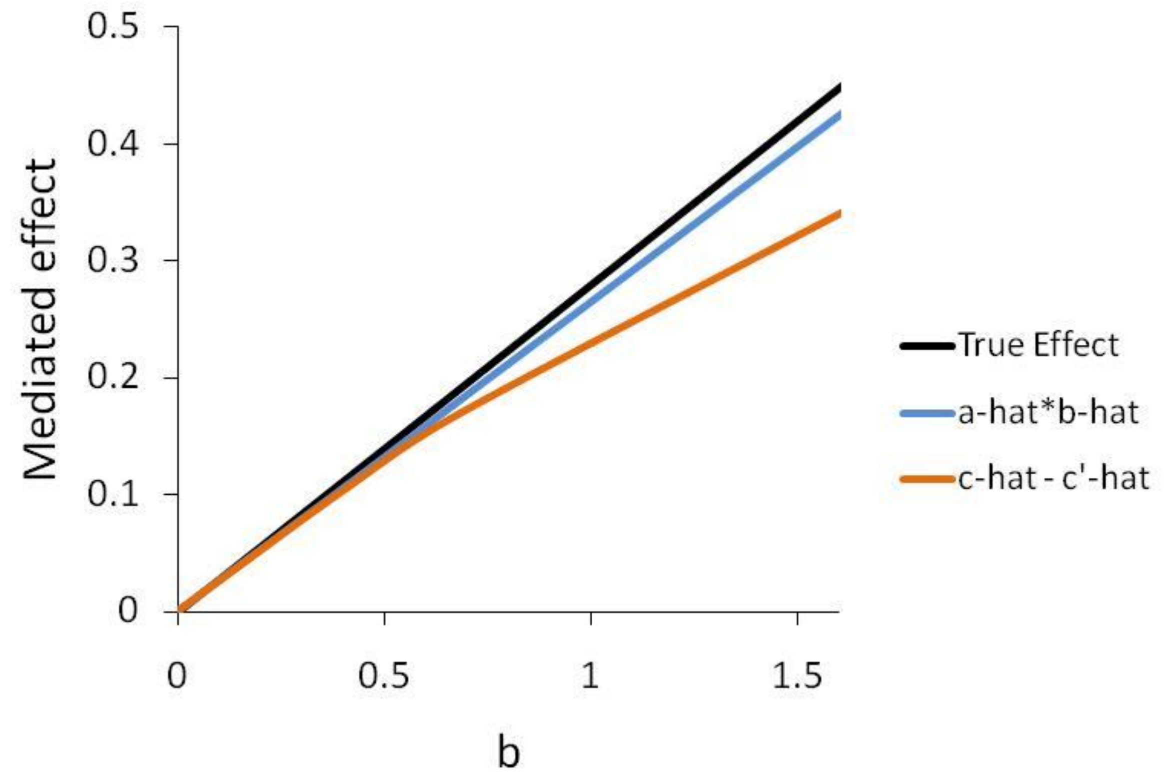
$$M = aX + e_{Normal\,(0,1)}$$

$$Y = bM + c'X + e_{Poisson}$$




## Results

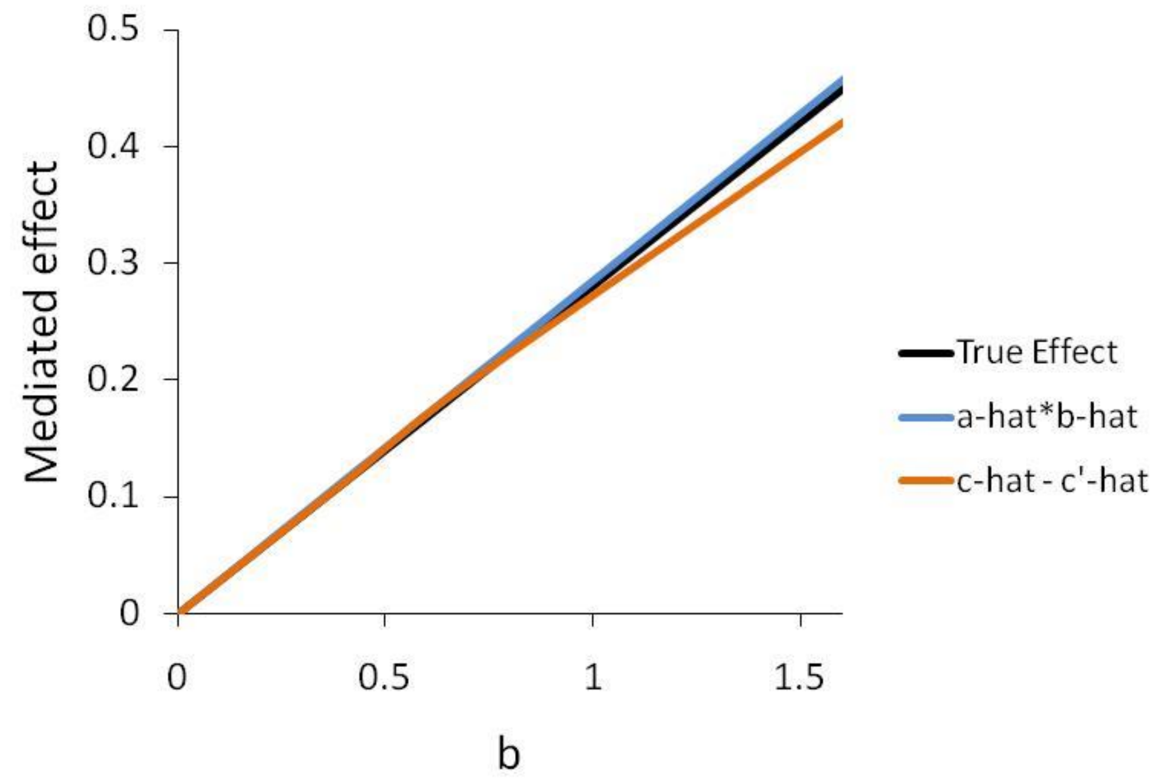
For standard Poisson outcomes, the *a*b* and *c-c'* estimates of the mediated effect were generally similar. However, compared to the true population values of the mediated effect, the *c-c'* method tended to under-estimate the mediated effect, particularly as the *b* path increased in magnitude; this effect was attenuated as sample size increased.




Sample size = 100

The *a*b* estimate of the mediated effect increases linearly with values of the *b* path and closely follows the true mediated effect value.

The *c-c'* estimate tends to underestimate the mediated effect by a large margin, particularly at larger values of the *b* path.



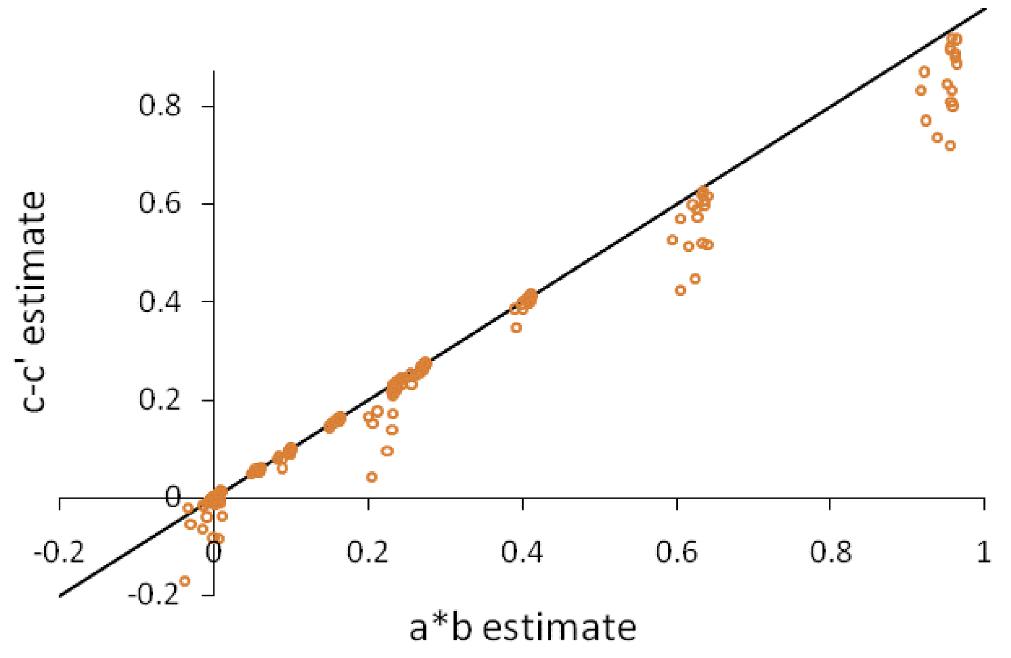

Sample size =1000

The *a*b* estimate continues to be very close to the true mediated effect.

The *c-c'* estimate continues to underestimate the mediated effect, but the discrepancy is somewhat attenuated.

The *a*b* and *c-c'* estimates of the mediated effect were generally similar, but the *c-c'* estimate of the mediated effect is consistently lower than the corresponding *a*b* estimate, across all conditions.

Points below the diagonal black line are conditions for which *a*b* > *c-c'*.
Points above the diagonal black line are conditions for which *a*b* < *c-c'*.
Note that all conditions are on or below the diagonal line.




## Discussion

These findings for mediation with a count outcome largely parallel those of MacKinnon et al. (2007) for mediation with a binary outcome. We found that the *a*b* and *c-c'* methods of estimating the mediated effect can lead to different estimates of the mediated effect. For example, for the largest *b* path value and a sample size of 100, ***a*b* = 0.43** and ***c-c'* = 0.34**. Additionally, we found that the degree of discrepancy between the *a*b* and *c-c'* methods depends on both the size of the *b* path (larger values of *b* → larger discrepancy) and the sample size (larger sample size → smaller discrepancy).

Although we found a discrepancy between the *a*b* and *c-c'* estimates of the mediated effect for count outcomes, it is unclear how the magnitude of this discrepancy relates to that found for binary outcomes. The effect size for the logistic regression models is based on the <u>linear, additive effect</u> of the predictors on the latent variable underlying the observed binary outcome; the effect size for the Poisson regression models is based on the <u>non-linear, multiplicative effect</u> of the predictors on the count outcome. At this point, it is unclear how to completely equate the effect sizes across logistic regression and Poisson regression models (i.e., across linear and non-linear models), so it is not clear whether the magnitude of the difference between *a*b* and *c-c'* is similar.

MacKinnon et al. (1993; 2007) showed that the discrepancy between *a*b* and *c-c'* for binary outcomes was due to the fixed residual variance of the logistic regression model; the variance was fixed to the same value in both equations (1) and (2), leading to different scaling for the coefficients in those equations (i.e., for *c* versus *b* and *c'*). They offered a rescaling solution based on the ratios of residual variances to equate *a*b* and *c-c'* for binary outcomes. Despite the fact that Poisson regression and logistic regression both belong to the family of GLiMs, a similar rescaling solution is not yet available for Poisson regression. Unlike logistic regression, Poisson regression does not have a continuous latent variable interpretation with homoscedasticity of variance; all versions of Poisson regression are heteroscedastic, and the non-constant variance does not allow for the rescaling shown for logistic regression.

## References


Armeli, S., Mohr, C., Todd, M., Maltby, N., Tennen, H., Carney, M. A., et al. (2005). Daily evaluation of anticipated outcomes from alcohol use among college students. *Journal of Social and Clinical Psychology, 24, 767–792.*

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data. *Journal of Personality Assessment*, 91, 121-136.

Judd, C. M. & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602-619.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Mahwah, NJ: Erlbaum.

MacKinnon, D. P. & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17, 144-158.

MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, 4, 499-513.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41-62.

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of American Academy of Child and Adolescent Psychiatry*, 39, 28-38.

Shanahan, C., Lincoln, A., Horton, N., Saitz, R., Winter, M., & Samet, J. (2005). Relationship of depressive symptoms and mental health functioning to repeat detoxification. *Journal of Substance Abuse Treatment*, 29, 117–123.



Presented at the 18th Annual meeting of the Society for Prevention Research.