



# New approaches for estimation of effect sizes and their confidence intervals for treatment effects from randomized controlled trials

Alan Feingold <sup>a</sup>,

<sup>a</sup>Oregon Social Learning Center

**Abstract** ■ Although Cohen’s *d* and the growth modeling analysis (GMA) *d* from linear models are common standardized effect sizes used to convey treatment effects, popular statistical software packages do not include them in their standard outputs. This article demonstrated the use of statistical software with user-prescribed parameter functions (e.g., Mplus) to produce *d* for treatment effects from both classical analysis and GMA—along with their associated standard errors (SEs) and confidence intervals (CIs). A Monte Carlo study was conducted to examine bias in the SE and CI for GMA *d* obtained with Mplus and found that both estimates were more accurate when calculated by the software with the standard bootstrap than with the delta method, but the delta method estimates were less biased than respective estimates from extant post hoc equations. Thus, users of many statistical software packages (including SAS, R, and LISREL) should obtain *d* or GMA *d* and associated CIs directly. Researchers employing less versatile software—and meta-analysts including *d*s and GMA *d*s in their syntheses of treatment effects—should continue to use the conventional post hoc equations. Biases in SEs and CIs for effect sizes obtained with them are ignorable and point estimates of *d* and GMA *d* are the same whether obtained directly from the software or with post hoc equations.

**Keywords** ■ effect sizes, confidence intervals, multilevel analysis, latent growth models.

**Acting Editor** ■ Denis Cousineau (Université d’Ottawa)

**Reviewers**  
■ Stephen Polizer-Ahles (The Hong Kong Polytechnic University)

[alanf@oslc.org](mailto:alanf@oslc.org)

AF: 0000-0003-0193-6305

10.20982/tqmp.15.2.p096

## Introduction

The need for effect sizes that communicate the potency of intervention effects is now well established (Grissom & Kim, 2012). There is also an increasing recognition of the importance of also providing confidence intervals (CIs) for these effect sizes (Cumming, 2013; Odgaard & Fowler, 2010; Preacher & Kelley, 2011).

Effect sizes can be unstandardized or standardized (Kelley & Preacher, 2012). Unstandardized effects sizes have an advantage over standardized effect size when making comparisons among findings from different studies that used the same outcome measure because there is no confounding of effect magnitude with sample homogeneity (Baguley, 2009). However, different studies examining the same hypothesis often use varying oper-

ationalizations of identical constructs, thus inextricably confounding homogeneity with instrumentation properties. This is a key reason standardized effects sizes are typically used in meta-analysis in the behavioral sciences (Feingold, 2017).

There are a number of standardized effect sizes in common use, and the choice of the effect size metric for a particular study is often based on the distributions of study variables: (1) the correlation coefficient (*r*), when the independent and dependent variables are both continuous, (2) the odds ratio (OR), when the outcome is categorical, and (3) the standardized mean difference (Cohen’s *d*), when the independent variable is categorical (e.g., treatment vs. control) and the dependent variable is continuous (Feingold, 2013). Thus, *d* is frequently used when reporting results from randomized controlled trials (RCTs)



examining efficacy of psychosocial interventions. However, the  $d$  statistic can vary as a function of design (Goulet-Pelletier & Cousineau, 2018), and the  $d$  relevant to RCTs (and thus this article) expresses the differences between independent groups (e.g., the treatment and control group)—whether observed in data from a completely randomized design, or from a mixed design that compares independent groups on repeated measures.

### Statistical Software for Standardized Effect Sizes

Although most statistical software packages (e.g., SPSS) output  $r$  or OR as a standardized effect size where appropriate, such programs do not report  $d$ . Thus, a two-step method is typically used to obtain  $d$  from reported descriptive (means and standard deviations) or inferential (e.g.,  $t$  ratios) statistics with post hoc equations typically found in meta-analytic texts (e.g. Borenstein, Hedges, Higgins, & Rothstein, 2009) to calculate  $d$  and its CI.

In addition, a growth model analysis (GMA)  $d$  can be derived from a linear multilevel or latent growth model that compares the random slopes of two groups over time to test intervention efficacy (Feingold, 2009). GMA  $d$  is a model-based estimate of the standardized mean difference between the two groups (e.g., treatment and control) at the end of a randomized study, and thus an equivalent of Cohen's  $d$  from a completely randomized design (Feingold, 2015). GMA  $d$  has now been reported in hundreds of RCTs (e.g. Chorpita et al., 2017; Felder, Epel, Lewis, Cunningham, Tobin, et al., 2017; Goodnight et al., 2017; Parra-Cardona et al., 2017; Stice, Rohde, Shaw, & Gau, 2017). As with classical  $d$ , GMA  $d$  is not reported in statistical outputs and has previously been obtained only with a two-step approach that uses a post hoc equation at step 2.

An alternative but rarely considered approach is to use statistical software with user-prescribed parameter functions— including lavaan in R (Rosseel, 2012), LISREL (Jöreskog & Sörbom, 2006), PROC CALIS in SAS (Inc, 2011), and Mplus (L. K. Muthén & Muthén, 2017) but not SPSS—to directly produce effect sizes, standard errors (SEs) and confidence intervals (CIs) for  $d$  and GMA  $d$  (Feingold, 2019). This article illustrates an application of this approach with Mplus, a versatile statistical package commonly used to conduct modeling analysis with observed and latent outcomes. Although Mplus—like other statistical programs—does not ordinarily produce  $d$  or GMA  $d$ , the software has the capability to create new parameters. This functionality would allow Mplus (and other programs with similar capabilities) to calculate  $d$  and GMA  $d$  directly, and to obtain their SEs and CIs with the same methods the program uses to produce SEs and CIs for standard parameters (e.g., regression coefficients). Thus, this article uses Mplus to demonstrate and validate the use of this new approach

for obtaining  $d$  and linear GMA  $d$  (with associated SEs and CIs), although adaptations to other software is relatively straightforward. (For effect size estimation for more complicated non-linear GMA models, see Feingold, 2019.)

### Calculation of $d$ in Classical Analysis

Cohen's  $d$  is the difference between the means of two independent groups divided by the pooled within-group standard deviation (SD).

$$\frac{M_1 - M_2}{SD}, \quad (1)$$

where  $M_1$  is the mean of one group and  $M_2$  is the mean of the other group.

Most mainstream statistical packages can be used to compare the means of two independent groups with commands specifying a  $t$  test, an analysis of variance (ANOVA), or a multiple regression analysis. However, Mplus requires the use of the regression framework to compare means, which entails coding the binary predictor ( $x$ ) capturing group and regressing the continuous outcome ( $y$ ) on those codes (Cohen, Cohen, West, & Aiken, 2003).

**Single covariate model (one-step method).** In a regression equation with a single dichotomous predictor (with a 1 unit difference between the codes used to create the  $x$  variable, e.g., "0" for control and "1" for treatment, or "-.5" and ".5"), the unstandardized regression coefficient ( $b$ ) of  $y$  on  $x$  is the raw score mean difference ( $M_1 - M_2$ ) between the groups—the numerator in the formula for  $d$  (see Equation 1). The square root of the residual variance from that model is the pooled within-group SD of the outcome—the denominator in Equation 1. The  $d$ , the SE of  $d$ , and the CI of  $d$  from a single covariate model (where group is the only predictor) can thus be calculated in Mplus in a single step with the MODEL CONSTRAINT command that creates new parameters from existing ones, such as regression coefficients and residual variances (see Listing 1 in Appendix A for the Mplus input that produces  $d$  from the single covariate model).

**Multiple covariates model (two-step method).** Designs with one or more covariates in the model in addition to the binary treatment variable are common in program evaluations. In particular, a pretest score is often included as a covariate in an independent groups pretest-posttest control group design (MacKinnon, 2008; Morris & DeShon, 2002) to decrease the SE of  $b$  and increase power to detect the treatment effect.

When treatment is not the only independent variable in the regression model, the residual variance is no longer the pooled within-group SD because variance in  $y$  explained by other covariates is removed from the variance of  $y$  (Cohen et al., 2003). Therefore, in GMA models with multiple covariates, the pooled SD cannot be determined



from the residual variance but instead must be obtained from a prior analysis at step 1 and then specified in MODEL CONSTRAINT at step 2 (see Listing 2 in Appendix A for input producing  $d$  using a specified SD from a model with multiple covariates). Such a two-step method in which a predetermined SD of the outcome is specified in the input to calculate an effect size is also used in Mplus for mediation analysis with a continuous predictor and a dichotomous outcome (Feingold, MacKinnon, & Capaldi, 2019; B. O. Muthén, Muthén, & Asparouhov, 2016).

Given randomization, the expected value of the correlation between the treatment variable and other covariates in a regression model is zero. Thus,  $b$  and  $d$  would then both have the same expected values in single and multiple covariate models. However, the CIs may be narrower in the latter because of the reduced SE as a result of variance in  $y$  (e.g., posttest score) explained by other covariates (e.g., pretest scores).

By default, Mplus produces the SE for new estimates (e.g.,  $d$ ) with the same delta method (Benichou & Gail, 1989; Kendall & Stuart, 1977) used to obtain the SEs for standard parameters (e.g.,  $b$ ). Inclusion of an optional command in the input statement to report CIs will have Mplus produce the CIs for both  $b$  and  $d$  with the delta method.

Mplus also has a capability to generate SEs and CIs for both default (e.g.,  $b$ ) and new (e.g.,  $d$ ) parameters with the bootstrap—either the standard non-parametric (percentile) bootstrap (Efron & Tibshirani, 1993) or the Bollen-Stine residual parametric bootstrap (Bollen & Stine, 1992)—in lieu of SEs and CIs obtained by default with the delta method. However, bootstrap CIs are almost never reported instead of delta method CIs for either  $b$  or  $d$ , which indicates that researchers generally assume that  $b$  and  $d$  are normally distributed and have symmetric CIs.

The square of the SE of  $d$  is the variance of  $d$ , which is used in meta-analysis of study findings in the  $d$  metric (Borenstein et al., 2009; Feingold, 2017). Meta-analysts typically calculate this variance with post hoc equations that are approximations provided in meta-analytic texts that yield estimates that are close to the squares of the SEs of  $d$  obtained in Mplus with the input provided in Appendix A. Unlike primary researchers, meta-analysts must rely on post hoc equations (e.g., Equation 1) because Mplus can only produce  $d$  and its SE from raw data, whereas meta-analysts typically need to calculate them with statistics extracted from research documents.

### Growth Modeling Analysis (GMA)

GMA—including multilevel modeling/hierarchical linear models (Goldstein, 2011; Hedeker & Gibbons, 2006; Hox, Moerbeek, & van de Schoot, 2010; Raudenbush & Bryk, 2002) and latent growth modeling (Bollen & Curran, 2006; Preacher, Wichman, MacCallum, & Briggs, 2008)—is often used to compare trajectories (e.g., means of random linear slopes) between groups to examine differences in rate of growth on an outcome over the course of a longitudinal study, particularly to evaluate intervention efficacy. GMA has revolutionized approaches to the analysis of repeated measures data used to examine naturally occurring or experimentally induced changes in people's attitudes, health, and behaviors (Gueorguieva & Krystal, 2004; Kuljanin, Braun, & DeShon, 2011). GMA is now as familiar to evaluators of interventions as ANOVA and ordinary least squares regression.

### Calculating GMA $d$ from Extant Post Hoc Equations

Equation 2 is typically used to convert the unstandardized coefficient ( $b$ ) for the effect of group on slope (the treatment effect) to a standardized effect size (GMA  $d$ )—the model-estimated standardized mean difference between the two groups at the end of a randomized study,

$$\text{GMA } d = \frac{b \times \text{duration}}{SD}, \quad (2)$$

which estimates the same effect size parameter as Cohen's  $d$  (Feingold, 2013, 2015).

The  $b$  in the numerator of Equation 2 is the difference in the rate of change in the outcome between the two groups per unit of time (e.g., per week when time is coded in weeks), and duration is the length of the study based on units associated with  $b$  (e.g., number of weeks from baseline if  $b$  is the group difference in rate of change per week).<sup>1</sup> The numerator in Equation 2 (the product of  $b$  and duration) is thus the model-estimated raw score mean difference between the two groups at the end of the study (and analogous to  $M_1 - M_2$  in Equation 1 for  $d$ ). The SD (denominator of Equation 2) is the pooled within-group SD of the outcome ( $y$ ) that is an estimate of the same parameter as the SD in Equation 1. However, with GMA of data from multiple time points, the SD of  $y$  can be calculated from observed baseline or end-of-study within-group variation, depending on statistical and theoretical considerations (see discussion in Feingold, 2013). SD can also be ob-

<sup>1</sup>The magnitude of  $b$  can vary with the coding of time in a given GMA. Therefore, the duration term in Equation 2 can also vary with coding of time because the product of  $b$  and duration (i.e., the expected raw score mean difference at end of the study) must be the same regardless of the value of  $b$ . As an example, consider a 6-week study that includes 4 assessments, with 2 weeks between time points. If the GMA used time codes of 0, 2, 4, and 6 for T1 (baseline), T2, T3, and T4 assessments respectively,  $b$  is the difference in change rate per week, and thus duration is 6, because it is a 6-week study. However, if time codes were based on measurement occasions rather than week (e.g., 0, 1, 2, and 3),  $b$  would be twice as large because it would then be the difference in rate of change expected in a 2-week period. The duration in an analysis using this alternative coding for the same study would then be 3. When the time codes are based on occasions that differ by 1 point between them, duration is 1 less than the number of time points.



tained from the GMA in a single covariate model by summing the variance of the intercept growth factor and the Level 1 residual variance (Feingold, 2015, 2019).<sup>2</sup>

Recent work (Feingold, 2015) has derived and validated an equation for the estimation of the variance ( $v$ ) of GMA  $d$ ,

$$v = SE_b^2 \times \left( \frac{\text{duration}}{SD} \right)^2, \quad (3)$$

where  $SE_b$  is the SE of  $b$  and  $SD$  is the same statistic used in Equation 2 to calculate GMA  $d$ . The square root of  $v$  is thus the SE of the GMA  $d$ , which can be used to calculate the 95% CI of GMA  $d$ ,

$$CI = \text{GMA } d \pm 1.96 \times SE. \quad (4)$$

A mathematically equivalent approach for CI estimation uses Equation 2, but with the lower and upper confidence limits (CLs) of  $b$  replacing the point estimate to transform the CI for  $b$  into the CI for GMA  $d$  (Feingold, 2015).

### Producing GMA $d$ for Linear Models in Mplus: An Illustrative Analysis

As with Cohen's  $d$ , the GMA  $d$  and associated statistics can be obtained directly with Mplus. Example 6.10 in the Mplus user's guide (L. K. Muthén & Muthén, 2017) consists of an input statement for a linear GMA—with 4 equidistant time points (coded 0, 1, 2, and 3 for  $y_{11}$ ,  $y_{12}$ ,  $y_{13}$ , and  $y_{14}$ , at T1, T2, T3, and T4, both respectively), 2 continuous time-invariant covariates ( $x_1$  and  $x_2$ ), and a single time-varying covariate ( $a_{31}$ - $a_{34}$ )—used for an illustrative GMA of an accompanying dataset included with the Mplus software (ex6.10.dat,  $N = 500$ ). The current illustration uses this example as a foundation for demonstrating the calculation of GMA  $d$ —and the three different SEs and CIs for GMA  $d$ —from Mplus, and affords comparisons with respective statistics obtained with the widely used post hoc equations that include delta method statistics (Feingold, 2009, 2015).

Because GMA  $d$  is used only with binary covariates, a DEFINE command was added to Example 6.10 input to dichotomize the continuous  $x_1$  (based on a mean split) covariate in the accompanying dataset, and the MODEL CONSTRAINT command was included for the program to calculate GMA  $d$ . With 4 equidistant time points differing by 1 unit between them, duration = 3. Two types of models are considered, a multiple covariates model (where there is one or more covariates in the model in addition to the covariate for condition) and a single covariate model (where group is the only time-invariant covariate in the analysis).

**Multiple covariates model.** As with classical analysis, a two-step method is needed to obtain effect sizes from a GMA model with multiple covariates (e.g., Mplus user's guide Example 6.10). The input requires specifications of: (a) duration, and (b) a predetermined SD for the within-group variation of  $y$  (see MODEL CONSTRAINT code in the Listings 3 and 4 in Appendix B). The SD of 1.478 specified for this example of a model with multiple covariates was the within-group SD at baseline ( $y_{11}$  POOLED), which is commonly used for SD estimation in RCTs because it ensures that the SD is unbiased by effect of treatment or attrition—and is also used to evaluate the effectiveness of the randomization when comparing the two groups at baseline (see Feingold, 2009, 2013). This SD was obtained by regressing  $y_{11}$  on the binary  $x_1$  covariate and taking the square root of the residual variance.

Three input statements—all modifications of Mplus Example 6.10—were used to conduct the illustrative analysis to produce the standard GMA statistics plus the GMA  $d$ s, SEs, and CIs from the multiple covariates model with: (a) the default delta method (see Listing 3 in Appendix B), (b) the standard bootstrap (Listing 4), and (c) the residual bootstrap (Listing 5). (Specification for Mplus to use a bootstrap instead of the default delta method to estimate SEs and CIs is the same in this example as in all Mplus input statements, and 500 draws were used in the illustrative analysis—the standard number of draws when the bootstrap was requested in input statements in Mplus user's guide examples.)

The observed GMA  $d$  of .872 was the same in the three analyses using the delta method and two bootstraps to produce the SEs and CIs for GMA  $d$  because point estimates are not affected by the method used to estimate the SEs and CIs. The Mplus-generated SEs and CIs from the multiple covariates model are reported in last three columns in the top half of Table 1 (under the heading "All 3 Covariates" for "Mplus Outputted CI") for each CI estimation method; respective SEs and CIs were nearly identical across the three estimation methods. The last the three columns in bottom half of the table ("Transformations of CI of  $b$  to CI of  $d$ ") reports CIs obtained for the GMA  $d$ s obtained with the post hoc equations approach. That is, the bottom half of the table reports corresponding results obtained by transforming the CI of  $b$  to the CI for GMA  $d$  for each type of CI with Equation 2 (with CLs of  $b$  substituted for  $b$ ). These CIs were virtually identical to respective CIs for GMA  $d$  calculated directly by Mplus (reported in the top half of the table), irrespective of the Mplus estimation method.

<sup>2</sup>In multilevel modeling (MLM) approaches to GMA (e.g., Raudenbush & Bryk, 2002), there is a single Level 1 variance in the model output. In the competing latent growth/structural model equation modeling framework for GMA used by Mplus, there is a separate residual variance associated with the  $Y$  at each time point, and the average of these residuals is the Level 1 variance in the MLM approach.





**Table 1** ■ SEs and CIs for GMA *ds* from Linear Models as a Function of Estimation Methods

Method	Covariates in Example 6.10 Dataset Included in the Linear GMA							
	x1 Only				All 3 Covariates			
	SE	95% CI	SE	95% CI	SE	95% CI	SE	95% CI
<b>Mplus Outputted CI</b>								
delta method	.134	[.767, 1.294]	.136	[.746, 1.279]	.133	[.734, 1.255]	.105	[.666, 1.078]
pbootstrap	.132	[.756, 1.291]	.134	[.741, 1.278]	.131	[.724, 1.252]	.104	[.679, 1.074]
rbootstrap	.131	[.764, 1.270]	.133	[.762, 1.273]	.129	[.748, 1.255]	.096	[.676, 1.070]
<b>Transformation of CI of <i>b</i> to CI of <i>d</i></b>								
delta method	NA	[.767, 1.293]	NA	[NA, NA]	NA	[NA, NA]	NA	[.666, 1.078]
pbootstrap	NA	[.755, 1.291]	NA	[NA, NA]	NA	[NA, NA]	NA	[.680, 1.074]
rbootstrap	NA	[.763, 1.271]	NA	[NA, NA]	NA	[NA, NA]	NA	[.676, 1.070]

*Note.* GMA = growth modeling analysis,  $N = 500$ . SE = standard error; CI = 95% confidence interval, pbootstrap = percentile (standard) bootstrap, rbootstrap = residual bootstrap, SD1 = 1.748, SD2 = SD estimated with  $y_{11}$  residual variance, SD3 = SD estimated with mean of all  $y$  ( $y_{11}$ - $y_{14}$ ) residual variances, NA = not applicable. CIs for time-varying GMA *ds* from the single covariate model ( $x_1$  only) cannot be compared with respective CIs from the multiple covariates model (using 3 covariates) because point estimates differ between the two types of models.

**Single covariate model.** An analysis was first conducted with only  $x_1$  as a covariate to illustrate calculation of GMA *d* from a single covariate model with a specified SD. However, in a GMA with a single covariate capturing group, SD does not have to be specified but can be estimated from the model using the variance of the intercept growth factor and residual variances of  $y$  with either of two equations. The first equation for estimating SD from the model takes the square root of the sum of the intercept growth factor variance and the mean of all the residual variances of  $y$  (Feingold, 2015)—and is most appropriate when the residual  $y$  variances are assumed homogeneous (or are specified to be equal in a Monte Carlo study). The second equation takes the square root of the sum of the intercept growth factor variance and the residual  $y$  variance associated with a time code of 0 ( $y_{11}$  in this example).

Listings 6, 7 and 8 from Appendix C provide the input for a single covariate model for each SD estimation approach: (1) specified SD (see Listing 6), (2) SD estimated from mean of all residual variances of  $y$  (Listing 7), and (3) SD estimated using the  $y_{11}$  (baseline) residual variance (Listing 8). Note that Listings 1 and 2 from Appendix A indicate the input for use of the default delta method for CI estimation for each SD estimation method. To obtain bootstrap CIs instead of delta method CIs, the bootstrap must be specified by adding an analysis command between the MODEL CONSTRAINT and OUTPUT commands, and the bootstrap specification must be added to the output CINTERVAL command (as shown in Listing 4 and Listing 5 in Appendix B for the input for the multiple covariates model).

The two model-estimated approaches for SD estimation

produce identical SDs (and thus the same GMA *ds* calculated using that SD as the denominator) when the residual  $y$  variance associated with a time code of 0 equals the mean of all the other residual  $y$  variances. In addition, with the equation using a single residual  $y$  variance, the GMA *ds* (but not their CIs) are nearly the same as the GMA *ds* obtained with the predetermined SD using the previously described two-step approach. Because there is no specified value for SD when SD is estimated from the model, the coefficient for the group difference in slopes (*b*) and the GMA *d* derived from it are obtained simultaneously in a single step.

In the illustrative GMA for this model that included only  $x_1$  as a covariate and specified SD = 1.478 in MODEL CONSTRAINT (Listing 6 input in Appendix C), GMA *d* = 1.030. When the SD was estimated from the GMA using the average of the 4  $y$  residual variances (Listing 7), GMA *d* = .994. When the SD was estimated from the GMA using the  $y_{11}$  residual variance (Listing 8), GMA *d* = 1.010. Thus, a GMA *d* of about 1.00 was obtained regardless of the approach used to estimate SD in the single covariate model, and was thus larger than the GMA *d* of .87 that was observed when the other covariates were included in the model in the previous illustrative analysis of the same data.

The first 9 columns in Table 1 reports the SEs and CIs calculated with the different methods of SD and CI estimation, including the CIs for the post hoc approach that were not calculated directly by Mplus but were obtained by transforming delta method or bootstrap CIs for *b* to CIs for GMA *d* (see bottom half of table headed "Transformation of CI of *b* to CI of *d*.".) The observed differences among



the different CIs for respective GMA  $d$ s were not meaningful.

### **Monte Carlo Study of the Validity of the Mplus Estimates for GMA $d$**

The validity of the estimates obtained with different methods used by Mplus for SE and CI for GMA  $d$  needs to be compared with the validity of the respective statistics obtained with the widely used post hoc equations (reported in Feingold, 2015). Errors in parameter estimates (bias) can be assessed with Monte Carlo simulation studies (e.g. Cheung, 2009; Hedges, Pustejovsky, & Shadish, 2012; Lau & Cheung, 2012; MacKinnon, Lockwood, & Williams, 2004). A Monte Carlo study was previously conducted by Feingold (2015) to examine bias in the estimates of the SE and CI for the GMA  $d$  obtained with post hoc equations (Equations 2 and 3, which used the  $SE_b$  estimated by the delta method). Bias was found to be small, particularly in large samples.

### **Objectives of Current Article**

A key purpose of this article is to illustrate an approach for obtaining effect sizes (and estimation of their SEs and CIs) from classical analysis and linear GMA that uses statistical software (specifically, Mplus) to create new parameters. However, this approach produces different types of SEs and CIs, raising questions about which statistics should be reported in program evaluations. Previous research using this approach with non-linear GMA models found that directly produced CIs obtained by Mplus with the delta method were less biased than respective bootstrap CIs (Feingold, 2019). Thus, a major objective of this article is to determine whether that finding generalizes to the GMA  $d$  obtained with a linear growth model. The prior study also found that sample sizes greater than 150 were needed for relatively unbiased effect sizes from quadratic GMAs. Thus, the current study examines whether a similar  $N$  is needed for linear GMA effect sizes. Evaluations of bias in bootstrap SEs and CIs for  $b$  for linear GMA were not possible in an earlier study (Feingold, 2015) because the Mplus version then available did not have its current capability of producing bootstrap statistics in a Monte Carlo study. Thus, bootstrap SEs and CIs for  $b$  are examined here to afford comparisons with respective biases in bootstrap SEs and CIs for GMA  $d$  that are derived from  $b$ .

### **Method**

Feingold (2015) used Monte Carlo analysis to examine bias in the SE and the 95% CI for  $b$  for the treatment effect computed with the delta method in 10 Monte Carlo simulations—each using 10,000 replications and specifying two parameters for the slope differences (.10 and .20) crossed with five sample sizes (ranging from 50 to 500).

Each replication manufactured and analyzed data for a balanced linear GMA with a dichotomous time-invariant covariate (i.e., two groups of equal size, 4 equidistant time points differing by 1 point between them) and a continuous outcome (for a complete sample Mplus input statement, see Appendix A in Feingold, 2015, or the non-bolded text in Listing 9 in Appendix D of this article).

Biases in the delta method SEs and CIs for  $b$ , and in the GMA  $d$  obtained with post hoc equations using delta method CIs, were both examined in Feingold (2015) following conventional practices for interpreting Monte Carlo results (L. K. Muthén & Muthén, 2002). The current study of bias in GMA  $d$  and associated statistics used the same 10 GMA models and input statements as the previous study to afford meaningful comparisons between biases in SEs and CIs obtained directly by Mplus in this study vs. the previously reported biases in respective statistics obtained with post hoc equations. However, commands and options were added to the earlier inputs to also generate GMA  $d$  within Mplus, along with its SE and CI. Thus, the expanded input statement (see Listing 9 in Appendix D, with added text in bold) used in this new Monte Carlo study obtained (a) the prior results for biases in  $b$  with the delta method, (b) additional results for biases in  $b$  obtained with the bootstrap, and (c) bias in SE and CI for GMA  $d$  calculated by with both the delta method and the standard bootstrap.

The commands and options used in the current study were Monte Carlo counterparts to the Mplus inputs presented in the introduction for the illustrative study (i.e., the expansions of the input statement in Example 6.10 in the Mplus user's guide), with the SD in the parameter creation equation calculated with the single-step method by Mplus using the mean  $y$  residual variances to estimate SD in each replication (see Listing 6 in Appendix C). The key differences between the two types of input statements are that the Monte Carlo study inputs include specifications of the effect size parameters for both  $b$  and GMA  $d$  but omit the CINTERVAL option. Given the intercept growth factor and residual variances, the  $b$  of .10 for the smaller treatment effect is associated with a GMA  $d$  of .3464, and the  $b$  of .20 for the larger effect corresponds to a GMA  $d$  of .6928 in these models (Feingold, 2015).

### **Input Statements for Current Monte Carlo Simulations**

For the 5 simulations evaluating the smaller effect size ( $b = .10$  and GMA  $d = .3464$ ), the bolded text in Listing 9 in Appendix D was added to the Feingold (2015) input statements to conduct Monte Carlo simulations examining SEs and CIs for the GMA  $d$  produced with the default delta method by Mplus. In the input statements for the 5 simulations specifying the larger effect size, .20 replaced .10 in the first line added to the MODEL COMMAND, and .6928



replaced .3464 in the first line in MODEL CONSTRAINT. To examine the standard bootstrap estimates instead of delta method estimates, the same ANALYSIS command specifying bootstrap was included that is used in an empirical study (see Listing 4 and Listing 5 in Appendix B for examples of bootstrap specification in Mplus).

A preliminary Monte Carlo analysis with the smallest specified sample size ( $N = 50$ , where the bias in the SE was the greatest with the default delta method) found that the Bollen-Stine residual bootstrap CIs evinced greater bias than respective delta method CIs, which had also been observed with effect sizes for quadratic GMA (Feingold, 2019). Accordingly, only biases in SEs and CIs obtained with the delta method and standard percentile bootstrap were examined in all 10 analyses,

Note that there is no option for CIs specified in the input for a Monte Carlo study via a CINTERVAL command because Mplus evaluates bias in CIs with coverage: the proportion of the replications in which the CI contains the parameter (L. K. Muthén & Muthén, 2002). Thus, perfect coverage for the 95% CI is .950, and the smaller the bias in the SE, the closer the coverage is to .950. In addition, Feingold (2015) proposed a CI bias statistic obtained by subtracting .950 from the coverage value, which is also reported in the results of CI bias in the current Monte Carlo study.

## Results

The Monte Carlo analysis found that the GMA  $d$  obtained by Mplus in each model was identical to the previously reported respective GMA  $d$  calculated with Equation 2 using statistics from the GMA (Feingold, 2015). Thus, bias in the point estimates was the same whether GMA  $d$  was calculated directly by Mplus or with Equation 2 (see Feingold, 2015, for demonstration that the bias in the point estimate of the GMA effect size parameter is ignorable).

Table 2 reports the results from the Monte Carlo analysis evaluating the bias in the SE and CI for the GMA  $d$  obtained with each of three different approaches: (a) post hoc equations (Equations 2 and 3) with a delta method  $SE_b$  (from Feingold, 2015), (b) delta method calculated in Mplus, and (c) bootstrap in Mplus. The first three columns in the table report (a) the specified  $N$  for the Monte Carlo results in that row, (b) the effect size parameter (small or medium delta), and (c) the empirical distribution of the generated GMA  $d$ s in each analysis, calculated as the SD of the GMA  $d$ s across the 10,000 replications used in each simulation. The next three columns (4-6) report the averages of the SEs of GMA  $d$ s across the same replications that were calculated using (a) Equation 2, as previously reported (Feingold, 2015), (b) the delta method, and (c) the percentile bootstrap, respectively. The coverage values for each CI estimation method are reported in the same order in columns 7-9.

The next six columns report the biases in the SEs obtained using different methods, with columns 10-12 reporting raw bias and columns 13-15 reporting percent bias. For respective biases in the point estimates, see Feingold (2015). The raw biases in columns 10-12 were calculated by the standard practice of subtracting the empirical distribution of SEs in column 3 from the corresponding average SEs in columns 4-6 (L. K. Muthén & Muthén, 2002). These raw biases were then divided by the SD of GMA  $d$ s across replications (column 3) and multiplied by 100 to obtain the percent biases that are reported in columns 13-15. The final 3 columns of the table report the CI bias index, calculated by subtracting .950 from coverage values in columns 7-9.

The last row in Table 2 reports the medians of the coverage values, percent biases, and CI biases across the 10 simulations. These statistics indicated that the delta method produced less biased SEs and CIs for the GMA  $d$  than the post hoc equations that used the delta method  $SE_b$ , and the percentile bootstrap afforded less biased SEs and CIs than the delta method. Indeed, the median coverage for the bootstrap was a perfect .950.

An examination of the rows in the table indicates that the bias in statistics obtained with both post hoc equations (as previously reported in Feingold, 2015) and the delta method in Mplus were the greatest at the smallest sample size and decreased rapidly as  $N$  increased. With the bootstrap, by contrast, minimal bias was found at the smallest sample size and there was no evident trend in bias related to  $N$ , with all observed variations in biases across the different simulations likely ascribable to sampling errors in the simulation analysis. Thus, the benefits of using the bootstrap over the other two approaches diminished as sample size increased.

Most important, the advantage of the bootstrap over the delta method for estimation of SE and CI for GMA  $d$  was also observed for  $b$  (see Table 3). However, unlike with GMA  $d$ , where the benefits of the bootstrap were appreciable at most sample sizes, the reduction in the bias in CI for  $b$  found with the bootstrap CI compared to the delta method CI was meaningful only with the smallest sample size ( $N = 50$ ).

A Monte Carlo simulation was also conducted with a very large sample size ( $N = 2000$ ) for each estimation method but with an otherwise identical input statement. Essentially zero bias was observed in the point estimate, the SE, and the CI obtained with all methods for both  $b$  and GMA  $d$  when  $N$  was very large, suggesting bias in the equation and delta method statistics was small sample size bias. That the estimation of both the effect size parameter and its SE improved with increases in sample size indicated that the GMA  $d$  meets the important effect size criterion of consistency (Preacher & Kelley, 2011).



**Table 2 ■** Some caption

N	Conditions		Monte Carlo Results										Bias Estimates					
			Avg					Coverage					Raw			Percent		
			Eq	Delta	BTSP	Eq	Delta	BTSP	Eq	Delta	BTSP	Eq	Delta	BTSP	Eq	Delta	BTSP	
50	.3464	.4506	.4325	.4355	.4626	.936	.942	.954	-.0181	-.0151	.0120	4.02	3.35	2.66	.014	.008	-.004	
50	.6928	.4539	.4325	.4391	.4664	.935	.941	.954	-.0214	-.0148	.0125	4.71	3.26	2.75	.015	.009	-.004	
100	.3464	.3095	.3040	.3055	.3142	.942	.945	.952	-.0055	-.0040	.0047	1.78	1.29	1.52	.008	.005	-.002	
100	.6928	.3117	.3040	.3079	.3167	.943	.947	.952	-.0077	-.0038	.0050	2.47	1.22	1.60	.007	.003	-.002	
150	.3464	.2532	.2477	.2487	.2532	.940	.942	.946	-.0055	-.0045	.0000	2.17	1.78	.00	.010	.008	.004	
150	.6928	.2551	.2477	.2506	.2552	.939	.942	.946	-.0074	-.0045	.0001	2.90	1.76	.04	.011	.008	.004	
250	.3464	.1938	.1915	.1922	.1941	.947	.948	.950	-.0023	-.0016	.0003	1.19	.83	.15	.003	.002	.000	
250	.6928	.1952	.1915	.1937	.1956	.945	.948	.949	-.0037	-.0015	.0004	1.90	.77	.20	.005	.002	.001	
500	.3464	.1373	.1353	.1357	.1364	.944	.945	.948	-.0020	-.0016	-.0009	1.46	1.17	.66	.006	.005	.002	
500	.6928	.1383	.1353	.1367	.1375	.944	.946	.947	-.0030	-.0016	-.0008	2.17	1.16	.58	.006	.004	.003	
— Median —						.942	.945	.950				2.17	1.25	.62	.008	.005	.000	

*Note.* Avg = average SEd across replications, Coverage = 95% coverage for b, CI = 95% confidence interval, Eq = equations approach with delta statistics, delta = delta method, BTSP = percentile bootstrap. Numbers in Eq columns were from (Feingold, 2015).





## Discussion

The Monte Carlo study found that the GMA  $d$  calculated by Mplus was identical to the GMA  $d$  obtained with the use of Equation 2 following the GMA (the conventional post hoc approach). However, the bias in the CIs for the GMA  $d$  calculated by the standard bootstrap with was smaller than the bias in CIs obtained with the delta method, although the latter was smaller than the bias in CIs obtained with post hoc equations that transformed delta method CIs for  $b$  to CIs for GMA  $d$ .

Bootstrap CIs have been found to have advantages in estimation over conventional approaches to CI estimation for other statistics as well (Banjanovic & Osborne, 2016), especially the indirect effects in mediation analysis (Hayes, 2013; MacKinnon, 2008; Shrout & Bolger, 2002). However, the delta method yielded better time-varying GMA  $d$ s than the bootstrap in quadratic GMA, where the effect sizes are determined from effects of group on linear and quadratic slopes (Feingold, 2019).

The observed biases decreased rapidly as the specified sample size increased, as did the differences among methods in manifested bias in estimates. With an extremely large sample size (e.g.,  $N = 2000$ ), there was essentially zero bias in point estimates, SEs, and CIs for GMA  $d$ , irrespective of estimation method. Because the illustrative study using a modified Mplus user's guide example had a large sample size ( $N = 500$ ), it was no surprise that SEs and CIs for GMA  $d$  were essentially identical across methods used in that example.

However, even at the smallest sample size examined in the Monte Carlo study ( $N = 50$ ), the bias in the SE was always less than the 5% threshold for acceptable bias proposed by L. K. Muthén and Muthén (2002), whereas a sample size of 100 was insufficient to yield SEs with ignorable bias for effect sizes from quadratic GMA (Feingold, 2019). In addition, coverage was always excellent for the linear GMA  $d$  (94–96).

Thus, the bias in the statistics obtained with Equations 2 and 3 should not be problematical for researchers using software that cannot output the more accurate SEs and CIs for GMA  $d$  produced directly by GMA software like Mplus that has user-prescribed parameter functions. Also, the post hoc equations are needed for meta-analysis, where it is necessary to calculate the  $v$  for the GMA  $d$  from reported statistics rather than from raw data (Feingold, 2017).

Although previous examinations of the validity of estimates from post hoc equations included delta method SEs or CIs in those equations (Feingold, 2015), the terms used in those equations can include bootstrap SEs and CIs reported for  $b$  (see example in Table 1), which would be expected to yield less biased SEs and CIs for GMA  $d$  than when these

equations included delta method statistics. Indeed, the illustrative study found that transforming the bootstrap CI for  $b$  to the CI for GMA  $d$  yielded essentially the same CI as the bootstrap CI obtained directly in Mplus. Thus, the transformations equations could be used with software that provides bootstrap CIs or SEs for  $b$  but cannot directly produce GMA  $d$ . counterpart. For example, when an empirical researcher using GMA has reported a bootstrap SE or CI for  $b$ , a meta-analyst should have no qualms about using it to calculate the  $v$  of GMA  $d$  with extant methods (Feingold, 2015, 2017). Indeed, the Monte Carlo findings indicate that meta-analysts should calculate  $v$  with the bootstrap CI for GMA  $d$  rather than the delta method CI when retrieved studies reported both CIs.

GMA  $d$ s from a linear models are model-estimated standardized mean differences (Cohen's  $d$  equivalents) at the end of the study only when the design uses randomization (e.g., in an RCT) or matching to ensure that the expected mean difference between the two independent groups at baseline is zero. Because Equation 2 does not include a term for the effect of group on the random intercepts from the GMA, GMA  $d$ s are derived exclusively from differences between the groups in rate of growth from their respective—and potentially different—baselines. As a result, the GMA  $d$  is effectively adjusted for baseline differences, as in ANOVA (see Feingold, 2019, for an extended discussion of this issue, which applies to GMA  $d$ s from both linear and quadratic models).

Cohen's  $d$  and the GMA  $d$  from a linear model are examples of standardized effect sizes. However, there are circumstances in which unstandardized effect sizes are preferable (Baguley, 2009). The equations for unstandardized effect sizes for classical analysis (raw score mean difference) and linear GMA (model-estimated mean difference at end of study) are simply the numerators in Equations 1 and 2. Moreover, only minor modifications to the input statements for  $d$  and GMA  $d$  would be needed for Mplus to produce respective unstandardized effect sizes and their CIs. Specifically, the denominator in the each equation specified in MODEL CONSTRAINT would be eliminated, and labels applied to parameters for residual variances used to estimate SD are unnecessary.

Although the focus of this study was on the use of  $d$  and GMA  $d$  for findings from RCTs, where the standardized effect size is for the difference between the treatment and control groups (or between two different treatment groups), the methods are applicable to comparisons between any two independent groups. For example, the classical  $d$  that compares treatment and control groups is the same classical  $d$  that would be used to compare men on women to examine sex differences (e.g., Feingold, 1994). The GMA  $d$  can also be used in research that compares men



**Table 3 ■ Monte Carlo Analyses of Unstandardized Coefficients (bs) for the Group Difference in Slopes for a Linear Latent Growth Model as a Function of Sample Size and Estimation Method**

parameters		Monte Carlo Results				Bias Estimates					
		Avg		Coverage		SE		Percent		CI	
<i>N</i>	SD	Delta	BTSP	Delta	BTSP	Delta	BTSP	Delta	BTSP	Delta	BTSP
50	.1258	.1216	.1250	.937	.942	-.0042	-.0008	3.34	.64	.013	.008
100	.0878	.0866	.0876	.943	.944	-.0012	-.0002	1.37	.23	.007	.006
150	.0720	.0708	.0713	.942	.943	-.0012	-.0007	1.67	.97	.008	.007
250	.0554	.0550	.0552	.947	.949	-.0004	-.0002	.72	.36	.003	.001
500	.0394	.0390	.0391	.945	.946	-.0004	-.0003	1.02	.76	.005	.004
— Median —				.943	.944			1.37	.64	.007	.006

Note. Avg = average SE of *b* across replications, Coverage = 95% coverage for *b*, CI = 95% confidence interval, Delta = delta method, BTSP = bootstrap. Unlike in Table 1, results are not reported separately for small and medium effect sizes because findings did not vary by effect size for *b*, and the equations approach is not applicable.

and women in outcome trajectories (e.g., Huttenlocher, Haight, Bryk, & Seltzer, 1991; Leahey & Guo, 2001). Thus, methods of calculation (including the Mplus code to conducts such calculations) of *d* and GMA *d* are applicable to a broader range of research areas than evaluation of intervention efficacy.

In summary, users of GMA software with the appropriate capability to obtain GMA *d* and its CI directly should obtain and report the GMA *d* and its bootstrap CI, although the default delta methods CIs are only slightly more biased than the bootstrap CIs, especially when both are calculated directly by software. However, meta-analysts who do not have access to raw data, and empirical researchers who use a less versatile statistical software package than Mplus—and thus must rely on the post hoc equations (i.e., Equations 1-3) to calculate the SE and/or CI for *d* or GMA *d*—need not be unduly concerned about the bias in the SEs and CIs for the GMA *d* obtained with those equations. This is particularly true when the statistics included in those equations are bootstrap CIs for *b* that can be transformed to CIs for GMA *d* with a simple modification of Equation 2 that replaces the *b* with the CLs for *b*.

**Authors' note**

This work was supported by National Institutes of Health (NIH)/National Institute on Alcohol Abuse and Alcoholism (NIAAA) grant R01AA025069. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH or NIAAA.

**References**

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100, 603–617. doi:10.1348/000712608X377117

Banjanovic, E. S., & Osborne, J. W. (2016). Confidence intervals for effect sizes: Applying bootstrap resampling. *Practical Assessment, Research & Evaluation*, 21, 1–18.

Benichou, J., & Gail, M. H. (1989). A delta method for implicitly defined random variables. *The American Statistician*, 43, 41–44.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, 21, 205–229. doi:10.1177/0049124192021002004

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York: Wiley.

Cheung, M. W. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, 41, 425–438. doi:10.3758/BRM.41.2.425

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., . . . Krull, J. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology*, 85, 13–25. doi:10.1037/ccp0000133

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral analysis (3rd ed.)* ahwah, NJ: Erlbaum.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.



- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*, 429–456. doi:<http://dx.doi.org/10.1037/0033-2909.116.3.429>
- Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*, 43–53. doi:[10.1037/a0014699](https://doi.org/10.1037/a0014699)
- Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology*, *17*, 111–121. doi:[10.1037/a0030048](https://doi.org/10.1037/a0030048)
- Feingold, A. (2015). Confidence interval estimation for standardized effect sizes in multilevel and latent growth modeling. *Journal of Consulting and Clinical Psychology*, *83*, 157–168. doi:[10.1037/a0037721](https://doi.org/10.1037/a0037721)
- Feingold, A. (2017). Meta-analysis with standardized effect sizes from multilevel and latent growth models. *Journal of Consulting and Clinical Psychology*, *85*, 262–266. doi:[10.1037/ccp0000162](https://doi.org/10.1037/ccp0000162)
- Feingold, A. (2019). Time-varying effect sizes for quadratic growth models in multilevel and latent growth modeling. *Structural Equation Modeling*, *26*, 418–429. doi:[10.1080/10705511.2018.1547110](https://doi.org/10.1080/10705511.2018.1547110)
- Feingold, A., MacKinnon, D. P., & Capaldi, D. M. (2019). Mediation analysis with binary outcomes: Direct and indirect effects of pro-alcohol influences on alcohol use disorders. *Addictive Behaviors*, *94*, 26–35. doi:[10.1016/j.addbeh.2018.12.018](https://doi.org/10.1016/j.addbeh.2018.12.018)
- Felder, J. N., Epel, E., Lewis, J. B., Cunningham, S. D., Tobin, J. N., et al. (2017). Depressive symptoms and gestational length among pregnant adolescents: Cluster randomized control trial of centering/pregnancy plus group prenatal care. *Journal of Consulting and Clinical Psychology*, *85*, 574–584. doi:[10.1037/ccp0000191](https://doi.org/10.1037/ccp0000191)
- Goldstein, H. (2011). *Multilevel statistical models (4th ed.)*. Hoboken, NJ: Wiley.
- Goodnight, J. A., Bates, J., E., H.-M., A., P., S., G., Ballard, R. H., et al. (2017). Dispositional, demographic, and social predictors of trajectories of intimate partner aggression in early adulthood. *Journal of Consulting and Clinical Psychology*, *85*, 950–965. doi:[10.1037/ccp0000226](https://doi.org/10.1037/ccp0000226)
- Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, part i: The cohen's d family. *The Quantitative Methods for Psychology*, *14*, 242–265.
- Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate application (2nd ed.)*. New York: Routledge.
- Gueorguieva, R., & Krystal, J. H. (2004). Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. *Archives of General Psychiatry*, *61*, 310–317. doi:[10.1001/archpsyc.61.3.310](https://doi.org/10.1001/archpsyc.61.3.310)
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford.
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224–239. doi:[10.1002/jrsm.1052](https://doi.org/10.1002/jrsm.1052)
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.
- Huttenlocher, J., Haight, W., Bryk, A., & Seltzer, M. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–248. doi:[10.1037/0012-1649.27.2.236](https://doi.org/10.1037/0012-1649.27.2.236)
- Inc, S. I. (2011). *Sas/stat® 9.3 user's guide*. Cary, NC: SAS Institute Inc.
- Jöreskog, K. G., & Sörbom, D. (2006). *Lisrel 8.80 for windows*. Lincolnwood, IL: Scientific Software International, Inc.
- Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*, 137–152. doi:[10.1037/a0028086](https://doi.org/10.1037/a0028086)
- Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics: Volume 1 (4th edition)*. New York: MacMillan.
- Kuljanin, G., Braun, M. T., & DeShon, R. P. (2011). A cautionary note on modeling growth trends in longitudinal data. *Psychological Methods*, *16*, 249–264. doi:[10.1037/a0023348](https://doi.org/10.1037/a0023348)
- Lau, R. S., & Cheung, G. W. (2012). Estimating and comparing specific mediation models in complex latent variable models. *Organizational Research Methods*, *15*, 3–16. doi:[10.1177/1094428110391673](https://doi.org/10.1177/1094428110391673)
- Leahey, E., & Guo, G. (2001). Gender differences in mathematical trajectories. *Social Forces*, *80*, 713–732. doi:[10.1353/sof.2001.0102](https://doi.org/10.1353/sof.2001.0102)
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Routledge.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128. doi:[10.1207/s15327906mbr3901\\_4](https://doi.org/10.1207/s15327906mbr3901_4)
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups design. *Psychological Methods*, *7*, 105–125. doi:[10.1037/1082-989X.7.1.105](https://doi.org/10.1037/1082-989X.7.1.105)



- Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 4, 599–620. doi:[10.1207/S15328007SEM0904\\_8](https://doi.org/10.1207/S15328007SEM0904_8)
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide (8th ed)*. Los Angeles, CA: Muthén & Muthén.
- Odgaard, E. C., & Fowler, R. L. (2010). Confidence intervals for effect sizes: Compliance and clinical significance in the journal of consulting and clinical psychology. *Journal of Consulting and Clinical Psychology*, 78, 287–297. doi:[10.1037/a0019294](https://doi.org/10.1037/a0019294)
- Parra-Cardona, J. R., Bybee, D., Sullivan, C. M., Rodríguez, M. M. D., Tams, L., & Bernal, G. (2017). Examining the impact of differential cultural adaptation with latina/o immigrants exposed to adapted parent training interventions. *Journal of Consulting and Clinical Psychology*, 85, 58–71. doi:[10.1037/ccp0000160](https://doi.org/10.1037/ccp0000160)
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, 16, 93–115.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth modeling*. doi:[10.1037/a0022658](https://doi.org/10.1037/a0022658)
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed)*. Thousand Oaks, CA: Sage.
- Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. doi:[10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02)
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422–445. doi:[10.1037//1082-989X.7.4.422](https://doi.org/10.1037//1082-989X.7.4.422)
- Stice, E., Rohde, P., Shaw, H., & Gau, J. M. (2017). Clinician-led, peer-led, and internet-delivered dissonance-based eating disorder prevention programs: Acute effectiveness of these delivery modalities. *Journal of Consulting and Clinical Psychology*, 85, 883–895. doi:[10.1037/ccp0000211](https://doi.org/10.1037/ccp0000211)

## Appendix A: Mplus Input for Calculating d

### Listing 1: Input for Computing d from Single Covariate Model

```
TITLE: Example 1 of computation of Cohen's d with Mplus
DATA: FILE IS example.dat;
VARIABLE: NAMES ARE x y;
MODEL: y ON x (b);
       y (r);
MODEL CONSTRAINT:
       new(d);
       d = b/sqrt(r);
OUTPUT: CINTERVAL;
```

### Listing 2: Input for Computing d from Multiple Covariates Model

```
TITLE: Example 2 of computation of Cohen's d with Mplus
DATA: FILE IS example.dat;
VARIABLE: NAMES ARE x1 x2 y;
MODEL: y ON x1(b)
       x2;
MODEL CONSTRAINT:
       new(d);
       d = b/SD;
OUTPUT: CINTERVAL;
```

Note to Listing 2. The SD in the MODEL CONSTRAINT command in the multiple covariates model is the pooled within-group SD of y obtained in prior analysis that must be specified. In other words, the numerical value of SD replaces "SD" in the input. So if SD is, say, 1.5, the second line under MODEL CONSTRAINT would be: d=b/1.5;





## Appendix B: Expanding Mplus Example 6.10 to Produce GMA $d$ for Multiple Covariates Model

### Listing 3: Input for Delta Method for CI Estimation

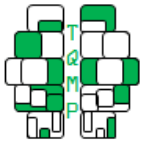
```
TITLE: Computation of GMA  $d$  for  $x_1$  in multiple covariate model
with delta method for CIs
DATA: FILE IS ex6.10.dat;
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
DEFINE: IF ( $x_1$  GE  $-.073$ ) THEN  $x_1=1$ ;
          IF ( $x_1$  LT  $-.073$ ) THEN  $x_1=0$ ;
MODEL: i s | y11@0 y12@1 y13@2 y14@3;
          i s ON x1 x2;
          y11 ON a31;
          y12 ON a32;
          y13 ON a33;
          y14 ON a34;
          s on x1(b);
MODEL CONSTRAINT:
          new(d);
          d = (b*3)/1.478;
OUTPUT: SAMPSTAT CINTERVAL;
```

### Listing 4: Input for Standard Bootstrap for CI Estimation

```
TITLE: Computation of GMA  $d$  for  $x_1$  in multiple covariate model
with standard bootstrap for CIs
DATA: FILE IS ex6.10.dat;
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
DEFINE: IF ( $x_1$  GE  $-.073$ ) THEN  $x_1=1$ ;
          IF ( $x_1$  LT  $-.073$ ) THEN  $x_1=0$ ;
MODEL: i s | y11@0 y12@1 y13@2 y14@3;
          i s ON x1 x2;
          y11 ON a31;
          y12 ON a32;
          y13 ON a33;
          y14 ON a34;
          s on x1(b);
MODEL CONSTRAINT:
          new(d);
          d = (b*3)/1.478;
ANALYSIS:
          BOOTSTRAP=500;
OUTPUT: SAMPSTAT CINTERVAL(BOOTSTRAP);
```

### Listing 5: Input for Residual Bootstrap for CI Estimation

```
TITLE: Computation of GMA  $d$  for  $x_1$  in multiple covariate model
with residual bootstrap for CIs
DATA: FILE IS ex6.10.dat;
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
DEFINE: IF ( $x_1$  GE  $-.073$ ) THEN  $x_1=1$ ;
          IF ( $x_1$  LT  $-.073$ ) THEN  $x_1=0$ ;
```



```
MODEL:  i s | y11@0 y12@1 y13@2 y14@3;
        i s ON x1 x2;
        y11 ON a31;
        y12 ON a32;
        y13 ON a33;
        y14 ON a34;
        s on x1(b);
```

```
MODEL CONSTRAINT:
        new(d);
        d = (b*3)/1.478;
```

```
ANALYSIS:
        BOOTSTRAP=500 (RESIDUAL);
```

```
OUTPUT:  SAMPSTAT CINTERVAL(BCBOOTSTRAP);
```

Note to Listing 5. Bold type indicates an addition or modification to the input statement for Example 6.10 in the Mplus user's guide to produce effect sizes in addition to standard statistics.

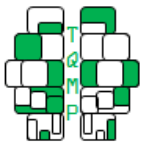
### Appendix C: Modifying Mplus Example 6.10 to Produce GMA *d* for Single Covariate Model

Listing 6: Input Using a Specified SD in Model Constraint

```
TITLE:  Computation of GMA d for x1 in single covariate model with
        specified SD
DATA:  FILE IS ex6.10.dat;
VARIABLE:  NAMES ARE y11-y14 x1 x2 a31-a34;
           USEVARIABLES=y11-y14 x1;DEFINE:  (*IF (x1 GE -.073) THEN x1=1;
           IF (x1 LT -.073) THEN x1=0;
MODEL:  i s | y11@0 y12@1 y13@2 y14@3;
        i s ON x1;
        s on x1(b);
        i(v1);
        y11-y14(r1-r4);
MODEL CONSTRAINT:
        new(d);
        d = (b*3)/sqrt(1.478);
OUTPUT:  SAMPSTAT CINTERVAL;
```

Listing 7: Input Using SD Estimated from All Y Residual Variances

```
TITLE:  Computation of GMA d for x1 in single covariate model
        using all Y residual variances
DATA:  FILE IS ex6.10.dat;
VARIABLE:  NAMES ARE y11-y14 x1 x2 a31-a34;
           USEVARIABLES=y11-y14 x1;
DEFINE:  IF (x1 GE -.073) THEN x1=1;
        IF (x1 LT -.073) THEN x1=0;
MODEL:  i s | y11@0 y12@1 y13@2 y14@3;
        i s ON x1;
        s on x1(b);
        i(v1);
        vy11-y14(r1-r4);
MODEL CONSTRAINT:
        new(d);
```



```
d = (b*3)/sqrt(v1 + r1/4 + r2/4 + r3/4 + r4/4);
```

```
OUTPUT: SAMPSTAT CINTERVAL;
```

Listing 8: Input Using SD Estimated from Y11 Residual Variance

```
TITLE: Computation of GMA d for x1 in single covariate model  
using Y11 residual variance
```

```
DATA: FILE IS ex6.10.dat;
```

```
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
```

```
USEVARIABLES=y11-y14 x1;
```

```
DEFINE: IF (x1 GE -.073) THEN x1=1;
```

```
IF (x1 LT -.073) THEN x1=0;
```

```
MODEL: i s | y11@0 y12@1 y13@2 y14@3;
```

```
i s ON x1;
```

```
s on x1(b);
```

```
i(v1);
```

```
y1(r1);
```

```
MODEL CONSTRAINT:
```

```
new(d);
```

```
d = (b*3)/sqrt(v1 + r1);
```

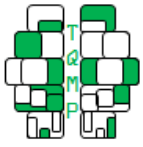
```
OUTPUT: SAMPSTAT CINTERVAL;
```

Note to Listings 6-8. GMA = Growth Modeling Analysis. An analysis command can be added to each model to request either standard bootstrap or residual bootstrap be used in SE and CI estimation instead of the default output produced by these input statements. Bold type indicates an addition or modification to the input statements for Example 6.10 in the Mplus user's guide.

Appendix D: Mplus Input for Monte Carlo Study for GMA  $d = .3464$  and  $n = 250$

Listing 9:

```
MONTECARLO: NAMES ARE y1-y4 x;  
CUTPOINTS = x (0);  
NOBSERVATIONS = 250;  
NREPS = 10000;  
SEED = 53487;  
CLASSES = C(1);  
GENCLASSES = C(1);  
ANALYSIS: TYPE = MIXTURE;  
ESTIMATOR = ML;  
MODEL MONTECARLO:  
%OVERALL%  
[x@0]; x@1;  
i BY y1-y4@1;  
s BY y1@-3 y2@-2 y3@-1 y4@0;  
[y1-y4@0];  
[i*0 s*.2];  
i*.25;  
s*.09;  
i WITH s*0;  
y1-y4*.5;  
  
i ON x*.3;
```



```

s ON x*.1;

%C#1%
[i*0 s*.2];

MODEL:

%OVERALL%
i BY y1-y4@1;
s BY y1@-3 y2@-2 y3@-1 y4@0;
[y1-y4@0];
[i*0 s*.2];
i*.25;
s*.09;
i WITH s*0;
y1-y4*.5;

i ON x*.3;
s ON x*.1;

%C#1%
[i*0 s*.2];
s on x.1 (b);
i (v1);
y1-y4 (r1-r4);

```

**MODEL CONSTRAINT:**

```

new(d*.3464);
d = (b*3)/sqrt(v1 + r1/4 + r2/4 + r3/4 + r4/4);

```

**OUTPUT:** TECH9;

Note to Listing 9. Bold type indicates input added to the input statement in Appendix A of Feingold (2015) to examine bias in the GMA d produced by Mplus. The specified GMA d of .3464 was for the small standardized effect size associated with a b of .10 and the specified residual variances. This input yields SEs and CIs for the default delta method. A bootstrap command must be added to produce bootstrap statistics.

**Citation**

Feingold, A. (2019). New approaches for estimation of effect sizes and their confidence intervals for treatment effects from randomized controlled trials. *The Quantitative Methods for Psychology, 15*(2), 96–111. doi:10.20982/tqmp.15.2.p096

Copyright © 2019, Feingold. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 14/03/2019 ~ Accepted: 05/05/2019