Lag as moderator meta-analysis: A methodological approach for synthesizing longitudinal data

Noel A. Card[1]

Abstract

Longitudinal data are common and essential to understanding human development. This paper introduces an approach to synthesizing longitudinal research findings called lag as moderator meta-analysis (LAMMA). This approach capitalizes on between-study variability in time lags studied in order to identify the impact of lag on estimates of stability and longitudinal prediction. The paper introduces linear, nonlinear, and mixed-effects approaches to LAMMA, and presents an illustrative example (with syntax and annotated output available as online Supplementary Materials). Several extensions of the basic LAMMA are considered, including artifact correction, multiple effect sizes from studies, and incorporating age as a predictor. It is hoped that LAMMA provides a framework for synthesizing longitudinal data to promote greater accumulation of knowledge in developmental science.

Introduction

Longitudinal research plays a critical role in advancing developmental science. In a survey of six leading developmental journals, Card and Little (2007) found that 41% of published studies used longitudinal designs. There are at least two important pieces of information that come from longitudinal data (for a fuller discussion, see Grimm, Davoudzadeh, & Ram, 2017). First, longitudinal data allow for the quantification of the stability of phenomena across time. Understanding the extent that a behavior or characteristic is stable versus changing is critical to understanding the emergence of individual differences, to identifying periods of greater or lesser stability across development (e.g., age differences in stability suggesting more or less critical periods of development), and to understanding the relative importance of various developmental periods (e.g., whether individual differences during adolescence persist into adulthood). A second major purpose of longitudinal data is to facilitate understanding of prediction across time. For phenomena in which experimental manipulation is not possible, our best proxy for understanding causal processes may come from data showing that levels of one variable (i.e., X, the presumed antecedent) predict later levels of a different variable (i.e., Y, the presumed consequent; see for example, Little, Card, Preacher, & McConnell, 2009). Documenting the strengths of these predictive relations, as well as understanding differences across ages (i.e., more or less sensitive periods for the process in question) and generalizability across studies, are critical goals of accumulating knowledge about human development.

Time Lags in Longitudinal Data

Although the value of longitudinal data has long been recognized, there has been less consideration of the time lags, the period of time between one measurement occasion and a second measurement occasion (c.f., Gollob & Reichardt, 1987). This lack of attention is unfortunate, as the time lags over which stability or longitudinal prediction are considered almost surely impacts the magnitudes of these effects.

When considering the stability of a single variable over time, the magnitude of this stability is inherently dependent on the time lag considered. In other words, it is less useful to talk of a construct as being stable or unstable than it is to quantify the magnitude of stability over a specific time period. Although the exact magnitude of stability over a particular time lag is an empirical question, the differences in stability over different time lags might be expected to follow certain patterns. Specifically, when stability follows from the simplex process (see Little, 2013, Chapter 6), the relative magnitude of stability is predictable across different time lags. In Figure 1a, the stability between measurement occasions 1 and 2, which we might say are one month apart, is some value (e.g., $\beta_{X_2,X_1} = 0.60$). If the stability between subsequent measurement occasions, which also have the same time lag (one month), is the same (e.g., $\beta_{X_3,X_2} = 0.60$), then the stability across two lags (i.e., two months) in knowable, namely $\beta_{X_3,X_1} = \beta_{X_2,X_1} * \beta_{X_3,X_2} = 0.60 * 0.60 = 0.36$. More generally, the expected stability for any number of time lags, t, is expected to equal $\beta_t = (\beta_1)^t$. In the example above where the one-month stability is 0.60, we might expect that the 4-month stability is 0.13 (0.60^4), or that the roughly two-week (half month) stability is 0.77 $(0.60^{1/2})$.

[1] University of Connecticut, USA

Corresponding author:
Noel Card, University of Connecticut, 348 Mansfield Road, Unit 1058, Storrs, CT 06269-1058, USA.
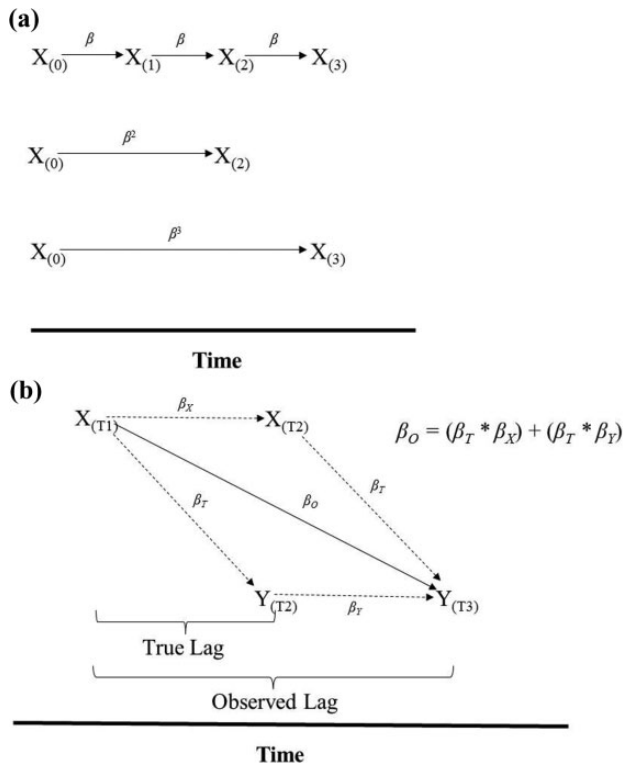Email: noel.card@uconn.edu

Figure 1. (a) Longitudinal stability across various time lags; and (b)(b) longitudinal prediction across various time lags.

The impact of time lag on longitudinal predictions is more complex (see Cole & Maxwell, 2003; Gollob & Reichardt, 1987). If we conceptualize longitudinal predictions as consistent with (if not definitively establishing) causal relations, then it is expectable that there exists some specific time lag over which these causal relations occur.[1] For instance, starting antibiotics might be expected to reduce symptoms of an infection within two days, beginning a cognitive behavioral intervention might be expected to reduce childhood depression in ten to fifteen weeks, and the impact of peer rejection might predict increases in victimization over several months. Developmental theories are generally silent about the time lag over which presumed causal processes occur, but it is certain that any causal process requires some amount of time to unfold. To consider the impact of selecting a suboptimal time lag to study a predictive relation, please see Figure 1b. Imagine that the actual time lag required for X to lead to increases in Y is 10 weeks, but a researcher is studying the process in a longitudinal study of 20 weeks. The observed longitudinal prediction relation (β_O) is going to be a biased estimate of the true longitudinal prediction relation (β_T; where "true" simply means the estimate of the process over the time lag over which the causal process unfolds). The direction and magnitude of this bias, however, are difficult to predict. Cole and Maxwell (2003) showed that the relation of the observed prediction relative to the true prediction depends on the stabilities of the antecedent (β_X) and consequence (β_Y), $\beta_O = (\beta_T * \beta_X) + (\beta_T * \beta_Y)$. Thus, the observed longitudinal prediction will overestimate the actual prediction when X and Y are highly stable but underestimate the actual longitudinal prediction if X and Y have low stability. To illustrate: imagine that the actual longitudinal prediction of X to Y 10 weeks later is $\beta_T = 0.30$, but the researcher investigates X

predicting later Y over 20 weeks. If the stabilities of X and Y are very strong, say $\beta_X = \beta_Y = 0.80$, then the longitudinal prediction in the 20-week longitudinal study is $\beta_O = (0.30 * 0.80) + (0.30 * 0.80) = 0.48$, a substantial over-estimation of the longitudinal prediction than if it had been studied over 10 weeks. In contrast, if the longitudinal stability of X and Y are low, say $\beta_X = \beta_Y = 0.20$, then the longitudinal prediction in the 20-week longitudinal study is $\beta_O = (0.30 * 0.20) + (0.30 * 0.20) = 0.12$, representing a substantial underestimation of the longitudinal prediction than if it had been studied over 10 weeks. Thus, a hypothetical study that has twice the time lag as the actual time for X to impact Y results in either over- or under-estimation of the effect size of the longitudinal prediction.

Lag as Moderator in Primary Studies

To model the impact of differences in time lag on longitudinal effects, Selig, Preacher, and Little (2012) introduced the lag as moderator (LAM) approach. The LAM approach capitalizes on individual differences in time lag *within* a primary study. For example, Selig and colleagues used the Early Head Start Research and Evaluation Study to demonstrate this approach. They used two waves in which data were collected at 14 and 24 months of age, and hence a 10-month lag between measurement occasions. However, there was substantial variability across families, with a range of 11 to 22 months at Wave 1 and 20 to 32 months at Wave 2, and an actual lag ranging from 5.9 to 17.0 months. This inter-individual variability in lag was included as a moderator of the effect of interest, home environment at approximately 14 months predicting the Bayley Mental Development Index at approximately 24 months. Specifically, the authors evaluated an interaction of individuals' home environment measure (X) and individuals' lag as predictors of later mental development using the following general equation

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 Lag_i + b_3 (X_i * Lag_i) \tag{1}$$

Selig and colleagues (2012) noted that lag moderation is likely to be nonlinear, such that the effect is likely to reach some peak at some optimal time lag (see also Wright, 1960). For this reason, Selig and colleagues' (2012) proposed modeling nonlinear lags used a quadratic lag term

$$\hat{Y}_i = b_0 + b_1 X_i + b_2 Lag_i + b_3 Lag_i^2 + b_4 (X_i * Lag_i) + b_5 (X_i * Lag_i^2) \tag{2}$$

Under this model, the association of the predictor with the dependent variable is allowed to differ both linearly with lag (i.e., becoming smaller or larger with longer time spans) and quadratically with lag (i.e., reaching some maximum or minimum effect).

Lag as Moderator Meta-Analysis

In this paper, I extend the logic of Selig et al.'s (2012) LAM to consider variability in time lag that exists *between* studies. Namely, I propose explicit coding and analysis of time lags of primary studies within the framework of existing meta-analysis techniques.[2] The proposed LAMMA represents a powerful method for modeling the impact of time lag on stability and longitudinal prediction. The strengths of this approach, relative to modeling within-study variability in lag in primary studies, are threefold. First, LAMMA will synthesize multiple longitudinal studies, thereby likely containing greater variability in time lag than that available in a single study.

Second, the aggregation of multiple studies will contain greater sample size than individual studies, thus increasing statistical power and precision. Third, aggregating across multiple studies allows for greater heterogeneity in sample characteristics (e.g., greater representation of multiple countries and/or multiple ages) than is typical in single studies, therefore providing more generalizable conclusions.

In the remainder of this paper, I briefly offer suggestions for literature searches and coding for LAMMA, and then detail the analyses of performing a LAMMA. I then briefly present results of an illustrative example, with full details (e.g., syntax and output) in an online document of Supplemental Materials. Finally, I will end with possible further extensions of LAMMA.

Searching and Coding for LAMMA

Prior to conducting analyses, meta-analysts engage in several steps including thoroughly searching the literature for relevant studies and systematically coding study features and results (see Cooper, 2009). In order to provide as broad an overview as possible for readers wishing to conduct a LAMMA, I next describe some aspects of searching and coding preliminary to data analysis that are unique to LAMMA.

The first unique aspect of conducting a LAMMA is at the stage of searching for relevant literature. Given that LAMMA focuses on longitudinal effect sizes, a necessary inclusion criterion will be that studies must be longitudinal in design. In many cases, meta-analysts might further limit studies to those with prospective designs; if retrospective designs are included, it is necessary that the retrospective reporting be of a specific enough timeframe that a meaningful lag, or duration between retrospective and current reports, can be coded. It will also be necessary that the same construct be measured at both times if one wants to synthesize stability coefficients; and the meta-analyst might reasonably include only studies using the same or a highly similar measure at both times (studies using a different measure will likely have lower stability estimates, as the correlation is attenuated to the extent that the two measures are not themselves correlated concurrently). For longitudinal prediction, minimum inclusion criteria are that the presumed antecedent is measured at Time 1, and the presumed consequent is measured at Time 2; many meta-analysts might also choose to only include studies that control for initial (Time 1) levels of the presumed consequent.

Other unique aspects of LAMMA arise during study coding. Unlike general meta-analyses that might use a wide range of possible effect sizes, LAMMA will use a more limited set of effect sizes. For meta-analyses of stability, the most common practice would likely be to use correlation coefficients to capture the inter-individual stability as an association between Time 1 and Time 2 values. If the variable considered is dichotomous, the meta-analyst might instead use the odds ratio (which is generally favorable over similar indices; see Fleiss, 1994). If the meta-analyst conceptualizes the construct of interest as continuous but a study reports the over-time contingency of a categorized variable, there exist equations for converting odds ratio to correlations (e.g., Card, 2012, pp. 118–119), and the meta-analyst might consider correcting the stability estimate for the artificial dichotomization (see below). Prior to analyses, most meta-analysts transform the correlation coefficient r using Fisher's transformation to improve the distributional properties for

analyses (see e.g.; Card, 2012; Rosenthal, 1991; for a contrasting view see Hunter & Schmidt, 2004).

For LAMMA of longitudinal predictions, one of two effect sizes are likely to be used. If one includes only studies of Time 1 X predicting Time 2 Y *without* controlling for initial levels, then this association is simply the correlation coefficient r (assuming continuous X and Y). However, if one has inclusion criteria of studies that control for initial levels, then the effect size of interest is a standardized regression coefficient. Specifically, the effect size is the regression coefficient of Time 1 X predicting Time 2 Y (β_1) in the following model: $Y_2 = \beta_1 X_1 + \beta_2 Y_1 + e$. Note that only studies reporting this exact model should be included within the meta-analysis; studies that include any additional covariates would produce an inaccurate estimate of β_1 unless it is assumed that the inclusion of covariates in the model does not impact results. Alternatively, the meta-analyst could compute the regression coefficient if the study reports the three component correlations among Time 1 X, Time 1 Y, and Time 2 Y (see Cohen, Cohen, West, & Aiken, 2003, p. 68)

$$\beta_1 = \frac{r_{X_1 Y_2} - (r_{X_1 Y_1} * r_{Y_1 Y_2})}{1 - r_{X_1 Y_1}^2} \qquad (3)$$

In a LAMMA including only studies controlling for initial levels of Y (i.e., the regression model shown earlier in this paragraph), the analysis of β_1 is straightforward in that one treats the regression coefficient as if it was a correlation coefficient (i.e., using Fisher's transformation if one would typically do so). However, if a LAMMA includes both studies controlling and studies not controlling for initial levels, and therefore includes both r and β_1, as effect sizes for prediction, methodological opinions vary in whether these can be combined (for a favorable opinion, see Peterson & Brown, 2005; for a less favorable opinion see Aloe, 2015). Given the presence of a large number of longitudinal studies, it might be possible to select one type of effect size over the other (if both are possible, it would generally be preferable to use those controlling for initial levels of Y). However, in most situations, there will be both types of effect sizes but not enough of either to exclude the other. In this latter situation, I recommend that users of LAMMA code whether or not the study controlled for initial levels, and test this between-study characteristic as a moderator of both conceptual and methodological interest.

In addition to coding the longitudinal effect size of interest, it is necessary to precisely code the Lag, or time between measurement occasions. In this paper, I generally consider Lag as measured in chronological time, such as months or years between measurement occasions. If a user of LAMMA adapts this conceptualization, then the task of coding Lag is one of exactly recording the time between measurement occasions in a common metric (e.g., months) across all studies. As will be shown below, I recommend analyzing Lag as a continuous variable; coders should retain the Lag of a particular study in as precise terms as possible, and should not artificially categorize this variable (e.g., short versus long lag). Although it is likely to be less common, it is possible that LAMMA could accommodate Lag using other time metrics. For example, those who meta-analyze clinical research might be interested in stability or prediction over a number of treatments or therapy sessions, and therefore record Lag as number of treatments or session between measurement occasions. Another example might be that an educational researcher is interested in stability or prediction across number of grades or another metric capturing amount of instruction

received. These and other (see Little, 2013, pp. 49–52) alternative metrics for time of the Lag are possibilities within LAMMA, so long as it is possible to consistently code these Lags from studies in the meta-analysis.

A final note regarding coding is that effect sizes are weighted during meta-analyses by a function of sample size. For the particular effect sizes of LAMMA, it is important to record the effective sample size used in estimating the longitudinal effect size. For instance, a study might have a sample size of 100 participants at Time 1 and a sample size of 80 at Time 2. If the difference in sample size is entirely due to attrition, then the stability coefficient is likely computed with 80 participants, and this sample size should be recorded. However, if Time 2 included newly-included participants, then perhaps only 70 participants provided data from both points used to estimate to stability coefficient. For longitudinal prediction of Time 2 Y from Time 1 X, controlling for Time 1 Y, the regression coefficient might be estimated from an even lower number if there are additional missing data on any of these three variables. Careful reading and coding of sample sizes used in analyses (sometimes derived from error degrees of freedom), combined with contacting study authors and making reasonable approximations from information available, create additional work not as often experienced in other meta-analyses.

LAMMA Analyses

The LAMMA combines the lag as moderator approach within studies (Selig et al., 2012) with an extensive literature on evaluating study characteristics as moderators (i.e., predictors of effect sizes, which are commonly two variable associations) in meta-analysis. In this section, I elaborate this combination in three ways of increasing complexity.

Lag as Fixed-Effects Moderators

The most straightforward way of considering continuous moderators within meta-analysis is within a weighted regression, sometimes termed meta-regression framework. Within this framework, the effect size for each study is regressed onto the coded study characteristic, weighted by a fixed-effects weight this represents the precision of the point estimate of the effect size in that study (most commonly, $w = 1 / SE_i$, where SE_i is the standard error of the effect size given by the particular effect size index used and a function of sample size; see for example, Card, 2012, pp. 176–178).

Applying this meta-regression approach to LAMMA, one could evaluate Lag as a predictor of effect sizes, either stability coefficients or longitudinal prediction coefficients, as follows

$$ES_i = b_0 + b_1 Lag_i + \varepsilon_i \qquad (4)$$

where ES_i is the stability coefficient or longitudinal prediction for study i, Lag_i is the time lag of study i, and ε_i is the residual for study i. The two coefficients estimated are b_0 and b_1, the interpretation of which I describe next.

This basic meta-regression model can be estimated with any computer software capable of weighted regression analysis, including SPSS, SAS, and several packages within R (e.g., metaphor and robumeta). Although I refer readers to the online Supplemental Materials for syntax, and elsewhere for full details of performing these analyses (e.g., Card, 2012; Lipsey & Wilson, 2001), here I note three details. First, the overall model significance (listed

as regression sums of squares in SPSS) is typically denoted as $Q_{Regression}$ and evaluated as χ^2 with df = number of predictors, or 1 df in equation 4. Second, the regression coefficients are accurate, but the standard errors in some statistical software output are not, and therefore they must be corrected before use in significance testing or constructing confidence intervals (see Lipsey & Wilson, 2001). Third, the regression coefficients can be used to plot model-implied effect sizes (i.e., stability or prediction coefficients) at various time lags. The intercept (b_0) is interpreted as the stability or prediction when Lag = 0, which, if no transformations are made to coded lag, is a theoretical value. Alternatively, one might decide to transform Lag in order to be centered on a meaningful value so that the intercept is more meaningful (e.g., stability over a one-year lag) or so that it is the (weighted) average of lags across studies in the meta-analysis, and intercept represents average stability or prediction (see below regarding centering). The slope (b_1) represents the amount that the stability or prediction coefficient increases (or decreases, if the value is negative) across studies that are one unit of time longer. Here, the time unit of Lag is meaningful in terms of the magnitude of the coefficient (the value will be 12 times larger if Lag is scaled in years rather than months) but not the direction or statistical significance of the coefficient.

Modeling Nonlinear Lag

As mentioned earlier, it is likely advantageous to model nonlinear moderation of Lag in most cases. If considering stability coefficients, one would expect a quadratic association of Lag with stability (assuming a simplex structure, as described above). Similarly, with longitudinal prediction, we can follow the precedent of Selig and colleagues (2012) by modeling quadratic lag in the absence of any other known nonlinear function. Both situations suggest that a more preferable meta-regression equation for LAMMA is the following

$$ES_i = b_0 + b_1(Lag_i - \overline{Lag}) + b_2(Lag_i - \overline{Lag})^2 + \varepsilon_i \qquad (5)$$

The notation here is similar to equation 4, though two points merit elaboration. First, equation 5 now estimates three parameters: an intercept (b_0); a linear effect of Lag (b_1); and a quadratic effect of Lag (b_2). This meta-regression will now be evaluated by $Q_{Regression}$ with two df, so in addition to overall significance one would likely examine the significance of the linear and quadratic effects individually by evaluating those regression coefficients. The second point is that the Lag predictors in equation 5 are now centered by subtracting the average Lag ($\overline{Lag}$) from each study's Lag (Lag_i). The reason for this centering is the same as when evaluating quadratic effects in traditional analyses, to remove nonessential collinearity between the linear and quadratic effects (Aiken & West, 1991, p. 35) so as to most clearly evaluate the unique linear and quadratic effects. In the context of LAMMA, however, one must keep in mind that equation 5 is a weighted regression, so the average Lag subtracted must be computed using the study weights. The simplest way to compute this is to regress Lag onto a constant in a weighted regression using the same fixed-effects weights as used in equation 5 (see example within online Supplementary Materials).

Centering Lag around its weighted average has the desirable impact of reducing the collinearity between the linear and quadratic effects of Lag. This centering also changes the interpretation of the intercept, so that b_0 estimates the stability or prediction effect at the average lag across studies. One could still compute model-implied

effects at any given time lag, but if one desires significance tests or confidence intervals for stability at any particular time lag that is not the average, it might be preferable to consider alternative methods of reducing collinearity (e.g., residual centering; Lance, 1988; Little, Bovaird, & Widaman, 2006).

Lag within mixed-effects models. To this point, my presentation of LAMMA has been under rather stringent conditions assuming that all between-study variability was due to the time lag of the longitudinal study and to sampling error. In other words, once a researcher designs a study in terms of the time lag between data collection occasions, the stability or longitudinal prediction of that study (ES_i in equation 5 above) is a function of a population parameter specified by intercept and time lag ($b_0 + b_1$) and sampling error of drawing a sample of a particular size from the population (ε_i). This model, termed a fixed effects model, may be unrealistic in many circumstances. In addition to time lag and sampling error, it is plausible that studies vary in their effect sizes for other reasons, such as sample characteristics (e.g., age, ethnicity, and socioeconomic status), measurement (e.g., using self-, parent, or teacher reports; features of measurement instructions such as whether a specific timeframe is specified), and study design (e.g., whether an experimental intervention or a naturally occurring transition occurred between measurement occasions). To allow for these additional between-study differences, even if the specific sources of variance are not specified, it is useful to consider mixed effects models, as specified in the following equation

$$ES_i = b_0 + b_1(Lag_i - \overline{Lag}) + b_2(Lag_i - \overline{Lag})^2 + \xi_i + \varepsilon_i \quad (6)$$

Equation 6 differs from equation 5 in that an addition term, ξ_i (Greek lowercase Xi; see Card, 2012, p. 240), is added. This term denotes that the population-level variability of studies like study i (those studying the same population, using the same measures and methodology, etc.) separate from sampling error. Figure 2 illustrates this concept. The top of Figure 2 shows a hypothetical example where the effect size, that is the stability or longitudinal prediction, varies non-linearly with time lag. One particular study had time lag L (e.g., 5 months) between measurement occasions. The middle of Figure 2 shows a distribution of studies with that same time lag, L (5 months). The average of this distribution is defined by the LAMMA estimates, $\widehat{ES_i} = b_0 + b_1(Lag_i - \overline{Lag}) + b_2(Lag_i - \overline{Lag})^2$. However, rather than being a fixed point (as in the fixed-effects model of equation 5), this model allows variability around this predicted stability or prediction coefficient. We might imagine that some studies using the time lag L (5 months) sampled adolescents using teacher reports of the construct within a single school year (i.e., a set of particular methodological features), whereas other studies, which also used time lag L (5 months), sampled children using behavioral observations and had a grade transition between measurement occasions (i.e., a different set of methodological features). Regardless of the impact of any specific methodological features, it might be the case that some study features cause different types of studies to be at the higher or lower end of this distribution. This source of variability is denoted by ξ_i in equation 6 and shown as a negative ξ_A and positive ξ_B in the middle of Figure 2. A final source of between-study variability is that each study is a sample from a population, denoted by ε_i in equation 6. This sampling error results in two studies, even if they have the same lag and identical methodological features, to have differences in their effect sizes (e.g., Study 3 versus 4 at the bottom of Figure 2). Although the direction and
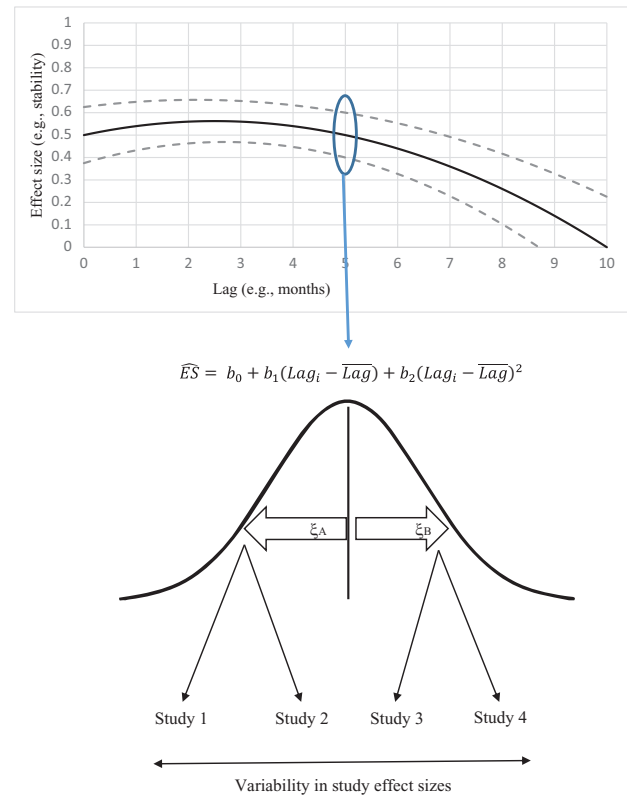

$$\widehat{ES} = b_0 + b_1(Lag_i - \overline{Lag}) + b_2(Lag_i - \overline{Lag})^2$$

Figure 2. Illustration of mixed-effects lag as moderator meta-analysis.

magnitude of any particular study's sampling error is indeterminate, it is possible to estimate the expectable amount of sampling error across studies in a meta-analysis (based on well-known standard error formulas and study sample sizes) to in turn estimate the amount of variability between study effect sizes due to population-level differences (commonly denoted with the Greek lower-case tau, so this variance is τ^2).

Although the idea of how to estimate population-level variance is relatively straightforward, in practice this can be challenging. The reason is that τ^2 represents the variability around a regression line in which stability or longitudinal prediction coefficients are predicted by a nonlinear function of lag, and this nonlinear function of lag is itself predicted based on study effect sizes differentially weighted by precision of point estimates. However, the τ^2 represents an additional source of imprecision of point estimates (i.e., we do not know if any one study is at the lower or upper end of the distribution in the middle section of Figure 2), so this additional uncertainty should be used within the weighting of studies. One solution to this issue is to solve the mixed-effects model of equation 6 using an iterative matrix algebra solution described next.

The weighted regression used in meta-regression analyses is an elaboration of the traditional matrix representation of regression, with an additional matrix **V** included for differential weighting of studies (see Card, 2012, p. 241)

$$\mathbf{B} = (\mathbf{X'V^{-1}X})^{-1}\mathbf{X'V^{-1}Y} \quad (7)$$

In equation 7 **B** is a m x 1 vector where m = number of predictors. In the context of the LAMMA with lag effects represented by intercept, linear, and quadratic terms, this vector would have three rows. Here **X** is a k x m matrix where k is the number of

studies and m is again the number of predictors. In LAMMA with quadratic function of lag, the first column would contain the constant value of 1 in the first column, each study's centered lag in the second column, and each study's centered and squared lag in the third column. Here $\mathbf{Y}$ is a k x 1 vector of effect sizes from the k studies. In the context of LAMMA, these values are either stability coefficients of longitudinal prediction effects. Finally, $\mathbf{V}$ is a k x k matrix with values on the diagonal representing each study's imprecision in estimating the stability or longitudinal prediction (and zeros for the off-diagonal elements). In a fixed-effects LAMMA, these values are the squared standard errors for the k studies.

The mixed-effects LAMMA begins with the $\mathbf{V}$ of equation 7 containing only study squared standard errors and uses these values to derive an initial estimate of the coefficients for intercept, linear, and quadratic coefficients of effect sizes (stability or longitudinal prediction) regressed on Lag. However, the model-implied effect sizes from this initial iteration are then used to compute deviations of each study's observed effect size (within a vector $\mathbf{Y}$) relative to the effect size predicted by study lag (within a vector $\hat{\mathbf{Y}}$), represented as a vector $\mathbf{D}$ with values containing each study's $\xi_i + \varepsilon_i$

$$\mathbf{D} = (\mathbf{Y} - \hat{\mathbf{Y}})^2 - (\mathbf{VU}) \tag{8}$$

(where $\mathbf{U}$ is a k x 1 vector constant with all cells $= 1$). One then separates the amount of variability expected from sampling error (contained in $\mathbf{V}$) to estimate population-level variance, τ^2

$$\tau^2 = \mathbf{U}'\mathbf{V}^{-2}\mathbf{D}(\mathbf{U}'\mathbf{V}^{-2}\mathbf{U})^{-1} \tag{9}$$

This value of τ^2 is then added to the squared standard error for each study to represent the uncertainty in the point estimate of stability or longitudinal prediction for each study, and this combined term is inserted into the values on the diagonal of $\mathbf{V}$ in equation 7. Then, using equation 7, the three regression coefficients of a quadratic representation of Lag are re-estimated. These new regression coefficients are used to predict new values of stability or prediction based on each study's time lag ($\hat{\mathbf{Y}}$), which are in turn used to recompute $\mathbf{D}$ and τ^2. This cycle continues until the change in τ^2 becomes trivially small between subsequent iterations (see Overton, 1998).

Illustrative Example

To demonstrate the LAMMA analyses described here, I next re-analyze data regarding the longitudinal stability of peer victimization among children and adolescents. In the original analysis (Card, 2003), 51 studies provided estimates of this stability. In this re-analysis, I have not updated the literature search, as the goal is to demonstrate the LAMMA techniques rather than to revisit the substantive conclusions. However, in the current re-analysis, one study using retrospective reports over an unclear time span was excluded; thus, the current analysis is based on 50 studies.

The stability estimates from these 50 studies are shown in the online Supplementary Materials. The stability estimates were coded as r, which were transformed via Fisher's transformation to Z_r for analysis, and overall results are back-transformed to r for reporting. In the example presented here, I focus on effect sizes not corrected for artifacts for simplicity (see below; (Card, 2003) presented both corrected and uncorrected). Therefore, the fixed-effects weights of studies are a simple function of sample size ($N - 3$; Card, 2012).

The first analysis was a fixed-effects regression of the linear effect of lag. I analyzed this model as a weighted regression using SPSS (see online Supplementary Materials for syntax and output), though readers might select another general software package (e.g., SAS) or specialized meta-analysis software such as Comprehensive Meta-Analysis (see Borenstein, Hedges, Higgins, & Rothstein, 2009) or one of several R packages (see Polanin, Hennessey, & Tanner-Smith, 2017). This result replicated the earlier analysis of Card (2003), showing that there was a negative relation between the lag of the study and the stability estimate, $b = -0.17$, $p < 0.001$. Inspection of model-implied stabilities indicated a six-month stability of $\hat{r} = 0.59$, a one-year stability of $\hat{r} = 0.54$, and a two-year stability of $\hat{r} = 0.41$ (see figure in the online Supplementary Materials).

The second analysis added a quadratic Lag to evaluate the non-linear relation of stability with time lag. It was necessary to center Lag prior to creating a quadratic term, but even in this case there was substantial collinearity ($r = 0.51$) due to skewed distribution of Lag across studies. Nevertheless, the model with both (centered) linear and quadratic Lag predicted 39% of the heterogeneity in stability of victimization ($Q_{(2)} = 522.07$, $p < 0.001$). There was again a linear decline of stability with time lag ($b = -0.19$, $p < 0.001$), but this was qualified by a quadratic effect of lag ($b = 0.023$, $p < 0.001$). The model-implied stabilities of this model at 6-months, one-year, and two-years were $\hat{r} = 0.62$, 0.54, and 0.38, respectively. For sample SPSS syntax, output, and full details of analyses, please see the online Supplementary Materials.

The third illustrative analysis estimated a mixed-effects model. The rationale for this model was that the previous two models exhibited significant residual heterogeneity, so a more realistic model is of a distribution of population effect sizes around a regression line (rather than only sampling error around a regression line; see e.g., Overton, 1998). The online Supplementary Materials show Mplus syntax (see Cheung, 2008) to perform these mixed effects analyses with the illustrative dataset. The results indicated once again a negative linear effect of lag ($b = -0.13$, $p < 0.01$), but the quadratic effect was no longer significant ($b = 0.009$, $p = 0.39$). The reason for this difference from the fixed-effects LAMMA is that the mixed-effects LAMMA uses lower weights for studies to account for imprecision due both to sampling error and to population-level uncertainty of where a study falls in the distribution of effect sizes around the regression line.

Extensions of LAMMA

The LAMMA model presented to this point can be extended in numerous ways. Although actual application of this model to longitudinal data will invariably lead to new challenges and opportunities, I here describe three likely extensions of this approach.

Artifact Correction

Some meta-analysis approaches emphasize the fact that effect sizes as reported within individual studies are often affected (typically attenuated) by various methodological imperfections, termed artifacts (see Hunter & Schmidt, 2004). The presence of artifacts within studies of a meta-analysis has two consequences. First, if study effect sizes are reduced by these artifacts, then the aggregation of these study effect sizes in meta-analysis will produce an attenuated estimate of the overall effect. Second, if studies differ in their degree of methodological imperfections, then the effect sizes will vary across studies for reasons that are likely less

interesting than other study features (e.g., lag and age). Fortunately, there exist techniques for correcting effect sizes for study artifacts that provide an unbiased estimate of what effect sizes would have been if the imperfections had not been present. Here I describe two artifact corrections that are likely useful in LAMMA.

The correlation between two constructs is reduced when the measures of those constructs have imperfect reliability. However, if a study reports reliability estimates (typically Cronbach's alpha for internal consistency of multi-item composite variables), then we can use those estimates to correct the correlation to what it would have been if estimated between perfectly reliable measured variables (or latent variables) using this formula:

$$r_{Corrected} = \frac{r_{Observed}}{\sqrt{\alpha_1 \alpha_2}} \qquad (10)$$

Where $r_{Observed}$ is the correlation reported in the study, α_1 and α_2 are the internal consistencies of the two variables comprising the correlation, and $r_{Corrected}$ is the estimate of what the correlation would have been with perfectly reliable measures. This estimate is unbiased, but not unerringly precise, so it is also necessary to adjust the standard error of the effect size to represent this additional uncertainty (see e.g., Card, 2012, pp. 129–131). In the context of LAMMA of stability coefficients, one of the studies in the illustrative example (see above and online Supplementary Materials) serves as an example: Juvonen, Nishina, and Graham (2000) estimated the stability of victimization over a one-year lag using a sample of 106 early adolescents. They estimated the stability as $r = 0.37$, but this estimate was based on a self-report instrument with internal consistency estimates $\alpha = 0.77$ and 0.79 at the two measurement occasions. From this information, we can estimate the actual stability, if estimated with perfectly reliable measures, as $r_{Corrected} = \frac{0.37}{\sqrt{.77*.79}} = .47$. However, the use of this artifact correction estimate reduces the weight of this study in the meta-analysis; rather than a weight of 103 the corrected stability is weighted by about 63, to account for the uncertainty in the artifact correction.

Another artifact that is likely to be encountered in LAMMA is when studies artificially dichotomize a continuous variable (see above discussion of coding effect sizes for LAMMA). With artificial dichotomization, the magnitude of attenuation is a function of where the dichotomization occurred: median splits result in just over 20% reduction in magnitude, and the attenuation becomes greater with more extreme splits. Rather than presenting formulas, I refer readers to a table (Card, 2012, p. 138). As with the correction for unreliability, the artifact correction for artificial dichotomization adds imprecision to the point estimate, and consequently lower weighting of the study.

These artifact corrections are relatively straightforward when applied to correlation coefficients, such as used to represent stability estimates or prediction *without* control for initial levels. In a LAMMA of longitudinal prediction *with* control for initial level, there do not exist straightforward artifact corrections of the regression coefficient. The reason for this absence is because imperfect reliability or artificial dichotomization of any of the three variables impacts this effect size. The only tractable method of artifact correction seems to be to correct each of the three correlations among Time 1 X, Time 1 Y, and Time 2 Y, and then to compute the regression coefficient from these corrected correlations (see equation 3 above).

Managing Multiple Lags Per Study

When studies included within LAMMA contain three or more measurement occasions, these will provide multiple estimates of

stability or longitudinal prediction of varying time lags. Treating these multiple effect sizes as if they were from separate studies is problematic for two reasons (see Borenstein et al., 2009; Wilson, Polanin, & Lipsey, 2016). First, ignoring the statistical interdependence among these effect sizes (i.e., the likely similarity among these effect sizes because they share a sample and methodologies) results in an artificially small standard error of overall results, resulting in increased Type I error rates and/or overly narrow confidence intervals. Second, this approach would unduly give more representation to studies with many effect sizes. A better, and far more common, approach to handling multiple effect sizes per study is to average multiple effect sizes to produce one effect size for that study. However, this approach may be problematic in LAMMA for two reasons. First, it is likely that most meta-analyses of longitudinal effect sizes rely on modest numbers of studies; aggregating information within studies is therefore problematic in that valuable information is lost. Second, aggregating within a study loses the variability in lags within a study (e.g., stability estimates across 6, 12, 18, and 24 months aggregated to a single 15-month stability), and in some LAMMA contexts the result might be little between-study variability with which to find differences.

Instead of aggregating or ignoring multiple effect sizes per study, it would be valuable to consider more sophisticated ways of managing these multiple stability or prediction coefficients of studies with three of more measurement occasions within LAMMA. An approach to doing this would be to adapt the matrix equation (equation 7 above; see Borenstein et al., 2009; Gleser & Olkin, 2009). In this adaptation, we keep the multiple effect sizes separate, so the number of cases in the LAMMA is the number of effect sizes (i.e., stability or longitudinal prediction) coefficients. However, rather than ignoring the interdependence among effect sizes, one accounts for the covariance among studies within the **V** matrix. That is, rather than having diagonal elements of **V** represent variances of parameter estimates (i.e., squared standard errors), this adaption uses **V** that is a block diagonal matrix with number of rows and columns equal to number of effect sizes (which is larger than number of studies). For example, Study 1 in a LAMMA might have two time-points, and therefore one stability coefficient, r_{12}. For this study, there is simply one diagonal element in **V** containing the variance (squared standard error) of that effect size (note that r is commonly transformed via Fisher's transformation prior to analysis to yield, Z_r, the variance of which is $\frac{1}{N-3}$; see e.g., Card, 2012, p. 90). If Study 2, however, has three time-points, it will contain three stability coefficients: r_{12}, r_{23}, and r_{13}. The next three diagonal elements, therefore, will contain the variance of these three effect sizes. Importantly, the off-diagonal elements in these rows/columns will no longer be 0, as when effect sizes come from independent studies. Instead, these off-diagonal elements will contain the covariance's among these estimates. The covariance of the stability between times 1 and 2 (r_{12}) and stability between times 2 and 3 (r_{23}) depends on the stability between times 1 and 3 (r_{13})

$$Cov_{r_{12}, r_{23}} = r_{13} \sqrt{Var_{r_{12}}} \sqrt{Var_{r_{23}}} \qquad (11)$$

Equation 11 can be used to construct off-diagonal elements of **V** for all interdependent effect sizes from a study. Once this **V** is constructed for all effect sizes in a LAMMA, as well as the **Y** containing all effect sizes and the **X** containing all functions (e.g., intercept, linear, and quadratic) of lag for each effect size, one can use equations 7–9 to accommodate multiple effect sizes per

study. For details of this approach, see Borenstein et al. (2009) and Gleser and Olkin (2009).

Lag and Age as Simultaneous and Interactive Predictors

Developmental science frequently highlights similarities and differences in age, and the simultaneous consideration of age and lag offer interesting possibilities in synthesizing longitudinal effect sizes.

Conceptually, it might be of interest to understand if stability differs across age, as developmental periods of lower stability (i.e., higher instability) might be important periods to study a phenomenon. Given the earlier argument that stability must always be considered in the context of the time lag investigated, it follows that meta-analytic investigation of age differences in stability should control for Lag. Thus, a reasonable LAMMA model for evaluating age differences in stability might be the following

$$ES_i = b_0 + b_1(Lag_i - \overline{Lag}) + b_2(Lag_i - \overline{Lag})^2 + b_3(Age_i) + \varepsilon_i \tag{12}$$

Note that I have shown this model as a fixed-effects model for brevity, but it could readily be adapted to a mixed-effects model by adding ξ_i. Here, Age_i is the sample age for study i., which is likely the average age of individuals within that study. Equation 12 allows one to estimate the effect of Age, b_3, controlling for the effects of lag (i.e., the likely possibility that studies of different ages also differ in the time lags considered). One could evaluate more elaborate models of nonlinear prediction of age as desired (e.g., adding a quadratic function of age to identify ages of highest or lowest stability).

It is also possible to model interactions of Lag and Age, such as in the following equation

$$\begin{aligned}ES_i = \ & b_0 + b_1(Lag_i - \overline{Lag}) + b_2(Lag_i - \overline{Lag})^2 + b_3(Age_i - \overline{Age}) \\ & + b_4(Lag_i - \overline{Lag})(Age_i - \overline{Age}) \\ & + b_5(Lag_i - \overline{Lag})^2(Age_i - \overline{Age}) + \varepsilon_i\end{aligned} \tag{13}$$

Note that equation 13 also centers age around the weighted average, which reduces collinearity of the main effects of age and lag with their interactions (Aiken & West, 1991). Although potentially applicable to either stability or predictive effect sizes within LAMMA, equation 13 seems especially relevant for longitudinal prediction studies. That is, it might be conceptually plausible to expect that predictive processes occur at different rates across development. In many cases, one might expect processes to operate more rapidly at younger ages (i.e., represented by the linear interaction effect, b_4). However, evaluation of age moderation of both the linear and quadratic effects of lag (i.e., b_4 and b_5) would allow meta-analysts to plot our prediction lags across development, potentially aiding (along with theory) the identification of possible critical periods of development.

The inclusion of age, or any other relevant study characteristic, is possible within the LAMMA framework. Two caveats merit consideration, however. First, the ability to test independent effects (such as in equation 12) and even more so, interactive effects (such as in equation 13) may in many cases be limited by the number of studies available. Power considerations in testing meta-analysis moderator effects are complex (involving number of studies, sample sizes within studies, distribution of study characteristics, residual heterogeneity, and other factors), but it seems safe to predict that many meta-analyses of longitudinal effect sizes will have low power to detect complex interactions. While low power limits hypothesis testing, it need not preclude regular fitting of such models. Instead, regular consideration and reporting of stability and prediction effects across a combination of lag and age within LAMMA should aid in building an accumulative database useful for developmental science.

LAMMA as a Piece of Science Considering Lag

Before ending, I want to emphasize that LAMMA is one piece of a comprehensive developmental science considering time lags. Namely, LAMMA allows researchers to understand *between-study* differences in time lag in relation to between-study differences in stability or longitudinal prediction. This approach differs from Selig and colleagues' (2012) *within-study* focus on *between-person* differences in lag. Yet another approach would be to identify *within-person* differences in lag (e.g., intervention effects for a child occurring faster during the academic year than during summer vacation) through idiographic studies with varying time lags. It has been argued that effects from idiographic studies (focused on within-person variability) and nomothetic studies (focused on between-person variability) should not be assumed the same (Molenaar, 2004), and it remains to be seen whether effects of lag on between-study results in LAMMA analyses are comparable to other effects of lag. So, while this paper aims to propose LAMMA with a focus on between-study variability, a healthy state of developmental science would also consider between-person and within-person variability in lag in nomothetic and ideographic primary research. Collectively, this approach would represent a concerted consideration of time lag that is generally lacking currently.

Conclusions

The goal of this paper has been to introduce and explain the use of LAMMA as a tool for meta-analyzing longitudinal study results. This framework would allow developmental scientists to synthesize a sizable body (41% of developmental studies, in a survey by Card & Little, 2007) of longitudinal research that has been largely neglected in previous meta-analyses. By synthesizing stability estimates, we can obtain better information about the comparative stabilities of different constructs, the ages at which stability is lower or higher, and quantified estimates of the stabilities over different lengths of time. Similarly, synthesis of longitudinal prediction allows more precise and generalized statements about possible causal relations (in cases when experimental manipulation is not available), and LAMMA can inform the time period over which such prediction takes place. All of these goals build upon the recognized value of longitudinal data (e.g., Grimm et al., 2017; Little et al., 2009) by providing a framework for meta-analytically synthesizing these results.

The LAMMA approach emphasizes the role of time lag in longitudinal studies, emphasizing the dependency of stability and prediction relations on the specific time span considered. Unfortunately, most developmental theory has not reached the degree of specificity about the time over which a process operates. In the short term, LAMMA allows an empirical approach to identify these time lags. With the accumulation of empirical results,

I hope that developmental theories will incorporate meta-analytically derived empirical data so as to become more precise about these timeframes.

Although the general framework of LAMMA is reasonably straightforward, application of these techniques in meta-analyses of longitudinal data invariably will pose unique challenges. I have considered some of these issues, including challenges in coding, and extensions of the basic LAMMA for artifact correction, with multiple effect sizes per study, and in combination with age. As these and other issues arise, I anticipate that the methodology of LAMMA will also need to respond to the needs of developmental applications of this approach. However, despite any potential uncertainties of using LAMMA to synthesize longitudinal developmental data, the opportunities for doing so are abundant. My hope is that the techniques described in this paper will prove fruitful in contributing to a more accumulative approach to developmental science.

Funding

Supplemental Materials

Supplemental material for this article is available online.

Notes

1. This contention that there exists a specific time lag is debatable. If we consider a causal process as having a sequence of unfolding events, such as A causing B which in turn causes C, which causes D, etc. until the observed consequence, then the "true' time lag is the sum of time for each step to unfold. However, human development is a complex rather than mechanistic process, and so this view is likely qualified by variability in the amount of time for each step in the sequence, multideterminacy (i.e., that each step in the sequence is probabilistically influenced by other causal agents), and multifinality (i.e., that each step in the sequence may probabilistically lead to other outcomes that may in turn influence the causal process). The possibility that a specific, or "true" lag might not exist does not invalidate subsequent consideration, as one can still consider an average lag or simply differences in comparing multiple lags.

2. An early paper by Thorndike (1933) considered time lags and stability estimates from multiple studies reporting test–retest stability of the Binet intelligence test. Although Thorndike's paper anticipated some of the issues regarding diminishing stability, it was published far earlier than the emergence of meta-analysis in the late 1970s and has not been commonly integrated into this literature.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: SAGE Publications.

Aloe, A. M. (2015). Inaccuracy of regression results in replacing bivariate correlations. *Research Synthesis Methods*, *6* (1), 21–27.

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

Card, N. A. (2003, April). Victims of peer aggression: A meta-analytic review. In N. A. Card & A. Nishina (Chairs), *Whipping boys and other victims of peer aggression: Twenty-five years of research, now where do we go?* Poster symposium presented at the biennial meeting of the Society for Research in Child Development, Tampa, FL.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford.

Card, N. A., & Little, T. D. (2007). Longitudinal modeling of developmental processes. *International Journal of Behavioral Development*, *31* (4), 297–302.

Cheung, M. W.-L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses in structural equation modeling. *Psychological Methods*, *13* (3), 182–202.

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

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112* (4), 558–577.

Cooper, H. M. (2009). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: SAGE Publications.

Fleiss, J. H. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York, NY: Russell Sage Foundation.

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 357–376). 2nd edition. New York, NY: Russell Sage Foundation.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58* (1), 80–92.

Grimm, K. J., Davoudzadeh, P., & Ram, N. (2017). Developments in the analysis of longitudinal data. In N. A. Card (Ed.), Developmental Methodology, *Monographs of the Society for Research in Child Development*, *82*, 46–66.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. 2nd edition. Thousand Oaks, CA: SAGE Publications.

Juvonen, J., Nishina, A., & Graham, S. (2000). Peer harassment, psychological adjustment, and school functioning in early adolescence. *Journal of Educational Psychology*, *92* (2), 349–359.

Lance, C. E. (1988). Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. *Applied Psychological Measurement*, *12* (2), 163–175.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford.

Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling*, *13* (4), 497–519.

Little, T. D., Card, N. A., Preacher, K. J., & McConnell, E. (2009). Modeling longitudinal data from research on adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 15–54). 3rd edition. New York, NY: Wiley.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2* (4), 201–218.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, *3* (3), 354–379.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, *90* (1), 175–181.

Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2017). A review of meta-analysis packages in R. *Journal of Educational and Behavioral Statistics*, *42* (2), 206–242.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: SAGE Publications.

Selig, J. P., Preacher, K., & Little, T. (2012). Modeling time-dependent association in longitudinal data: A lag as moderator approach. *Multivariate Behavioral Research*, *47* (5), 697–716.

Thorndike, R. L. (1933). The effect of the interval between test and retest on the constancy of the IQ. *Journal of Educational Psychology*, *24* (7), 543–549.

Wilson, S. J., Polanin, J. R., & Lipsey, M. W. (2016). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods*, *7* (2), 121–139.

Wright, S. (1960). The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrics*, *16* (3), 423–445.