Introducing Continuous Time Meta-Analysis (CoTiMA)

Christian Dormann and Christina Guthier
Johannes Gutenberg-University Mainz

Jose M. Cortina
Virginia Commonwealth University

in press: Organization Research Methods

1

Abstract

Meta-analysis of panel data is uniquely suited to uncovering phenomena that develop over time, but extant approaches are limited. There is no straightforward means of aggregating findings of primary panel studies that use different time lags and different numbers of waves. We introduce Continuous Time Meta-Analysis (CoTiMA) as a parameter-based approach to meta-analysis of cross-lagged panel correlation matrices. CoTiMA enables aggregation of studies using two or more waves, even if there are varying time lags within and between studies. CoTiMA thus provides meta-analytic estimates of cross-lagged effects for a given time lag regardless of the frequency with which that time lag is used in primary studies. We describe the continuous time underpinnings of CoTiMA, its advantages over discrete time correlation-based meta-analysis of structural equation models (MASEM), and how CoTiMA would be applied to meta-analysis of panel studies. An example is then used to illustrate the approach. We also conducted Monte Carlo simulations demonstrating that bias is larger for time category-based MASEM than for CoTiMA under various conditions. Finally, we discuss data requirements, open questions, and possible future extensions.

[1] Christian Dormann, Faculty of Law & Economics, Johannes Gutenberg-University Mainz; Christina Guthier, Faculty of Law & Economics, Johannes Gutenberg-University Mainz; Jose Cortina, School of Business, Virginia Commonwealth University. Correspondence concerning this article should be addressed to Christian Dormann, Faculty of Law & Economics, Johannes Gutenberg-University Mainz, Jakob-Welder-Weg 9, D-55128 Mainz, Germany. E-mail: cdormann@uni-mainz.de

Meta-analysis has been widely accepted as a useful tool to aggregate effect sizes obtained in primary studies, such as effect sizes obtained in experimental, correlational, or prospective studies. For decades, the number of meta-analyses published per year has been rising and today exceeds the number of articles using common statistical methods such as structural equation models (SEM; cf. Cheung, 2015). Even though this development is mainly driven by meta-analyses of experimental studies published in medical journals (Cheung), meta-analyses of correlational studies have become more common as well. More recently, a number of meta-analyses of cross-lagged effects reported in organizational studies has been published (e.g., Ford, et al., 2014; Mathieu, Kukenberger, D'Innocenzo, & Reilly, 2015; Nohe, Meier, Sonntag, & Michel, 2015; Riketta, 2008), and this number is likely to increase further because panel studies allow for more valid causal inferences compared to cross-sectional studies. The continuous time meta-analysis (CoTiMA) approach introduced in the current article solves several issues that researchers face when they aim at synthesizing longitudinal effects.

In panel studies, researchers collect at least two variables from the same individuals at two or more measurement occasions (waves) separated by one or more different time lags (e.g., Finkel, 1995; Kessler & Greenberg, 1981). Although the selection of variables is often driven by theory, the length of time lags is not. As Mitchell and James noted, "Decisions about when to measure and how frequently to measure critical variables are left to intuition, chance, convenience, or tradition" (2001, p. 533; see also Cole & Maxwell, 2003). Given the arbitrariness of this decision, it is not too surprising that time lags vary considerably across primary studies aiming to answer the same research question. A meta-analysis of effects from such studies, unlike meta-analyses of cross sectional effects, must contend with differing numbers of waves and differing time lags that are used across primary studies. As it is well known that lag length and effect size are related to one another (e.g., Cohen, Cohen, West, & Aiken, 2003; Dormann & Griffin, 2015; Voelkle, Oud, Davidov, & Schmidt, 2012), a meta-analyst must choose a strategy for dealing with varying numbers of waves and lags.

Previous meta-analyses of panel studies sometimes ignore this variability (e.g., Anderson et al., 2010; Nielsen et al., 2017; Rugulies, Aust, & Madsen, 2017). As we show later, across those meta-analyses in which variability was considered, categorization of time lags into, for example, short-term, medium-term, and long-term (e.g., Riketta, 2008) is perhaps the most common strategy. Nevertheless, all of the approaches used to cope with this variability have significant drawbacks. An additional problem arises when meta-analyzing studies that contain more than two time points because, using traditional approaches, one must choose whether to use correlations between Time 1 and Time 2, between Time 2 and Time 3, between Time 1 and Time 3, or all of the above, with each of these choices creating its own challenges.

Given the shortcomings of existing procedures for meta-analyzing results from panel studies, the challenge is fourfold. First, there is a need to meta-analyze effects from panel studies without artificial categorization of time lags. Second, there is a need to estimate model parameters for time lags of particular interest (e.g., optimal time lags, time lags that reflect practical constraints on measurement) regardless of whether those time lags were represented among the primary studies being analyzed. Third, there is a need is to estimate the rate at which model parameters change as one moves away from these time lags of interest. And fourth, there is a need to provide a coherent framework for meta-analyzing effects from multi-wave studies of all kinds. All these needs can be precisely addressed with CoTiMA.

The present article is structured as follows. First, we review approaches used in previous longitudinal meta-analyses to deal with differing time lags of primary studies. Second, we briefly review the principles of continuous time modeling. Third, we explain how continuous time modeling can be done using structural equation modeling (SEM), and how to conduct a CoTiMA, which essentially is a multi-sample continuous time SEM. Fourth, we demonstrate a CoTiMA and compare the findings with the typical correlation-based meta-analysis of longitudinal relations between job satisfaction and in-role performance by Riketta (2008). This allows for comparison of results based on different time lag categories with continuous time effects. Finally, we discuss data requirements and provide recommendations for presenting CoTiMA results.

Common Meta-Analytic Treatments of Cross-lagged Panel Data

The simplest data used for meta-analyses of primary panel studies involve a 4 × 4 cross-lagged panel correlation (CLPC) matrix of X and Y measured at Time 0 and Time 1 (i.e., X0, Y0, X1, and Y1), the sample size, and the time lag between the two time points. Many meta-analyses of longitudinal studies (e.g., Maricuțoiu, Sulea, & Iancu, 2017; Riketta, 2008) have estimated the six population correlations among the four variables as a starting point (Step 1). These population correlations are typically aggregated using traditional meta-analytic approaches such as that described in Borenstein, Hedges, Higgins, and Rothstein (2009). The aggregated correlations can then serve as input for further analyses (Step 2), for example, SEM, which would yield cross-lagged regression coefficients. This approach is called correlation-based meta-analytic structural equation modelling (correlation-based MASEM, Cheung & Cheung, 2016). Problems with this approach include missing correlations in primary studies, ambiguity in taking into account possible random effects computed in Step 1 in Step 2, and possibly incorrect test statistics and standard errors by using correlations rather than covariances (Cheung & Cheung). For our present purposes, the most important problem with this approach lies in the way that it examines the possibility that effects vary as a function of time lag.

As was mentioned earlier, one of the issues with which a meta-analyst of such studies must contend is that they are likely to contain different numbers of waves. To get an overview of the strategies previously used in the organizational sciences for meta-analyzing primary studies with varying numbers of waves and time lags, we identified 14 meta-analyses which had to deal with this issue (and one other, Nielsen et al., 2010, that almost certainly had to deal with the issue but did not provide enough information for us to be sure). The number of waves in primary studies included in these meta-analyses varied from 2 to 10, and time lags varied from a few weeks to 20 years! Clearly, some consideration must have been given by the authors of these papers to cope with this variability.

There are two issues with which one must cope when meta-analyzing effects from primary studies of this kind. First, one must decide how to deal with studies that have more than two waves, and therefore multiple, dependent effect sizes. We identified no fewer than six different strategies in these meta-analyses which were used to deal with this issue. First, all possible pairs of two-wave correlations were used (e.g., Mathieu et al., 2015). The meta-analysis by Rugulis et al. (2017) aggregated only 2-wave studies, but some of the primary studies already used aggregated relations across varying time lags between participants. Second, only non-overlapping correlations were used, for example, correlations between t0-t1 and t1-t2 were included, but correlations between t0-t2 were excluded (Jin & Rounds, 2012). Third, some meta-analyses accounted for the dependence of multiple pairs using multilevel modeling (Ford et al., 2014) or by averaging the correlations using Fisher's z-scores (Nohe et al., 2015). Fourth, only those waves that had the most common time lags were used (Mathieu et al.). Fifth, only the first two waves of any multi-wave study were used (Mathieu et al.; Riketta, 2008). Sixth, the correlations from the longest available time lag were used (Dormann & Zapf, 2001). Finally, five of the meta-analyses that included effects from multi-wave studies made no mention of their strategies for dealing with varying waves and lags (Crook, Todd, Combs, Woehr, & Ketchen, 2011; Judge, Piccolo, & Illies, 2004; Ketchen et al., 1997; Saridakis, Lai, & Cooper, 2017; Verkuil, Atasayi, & Molendijk, 2015). As we explain later, this makes it difficult and sometimes impossible for the reader to interpret mean effect size estimates from these papers. One great benefit of the CoTiMA approach introduced in the present paper is that it provides a coherent framework for using all available waves in primary studies simultaneously. There is no need to select or aggregate based on any of the aforementioned strategies. Therefore, CoTiMA prevents a loss of statistical power and the risk of overestimating effect sizes by aggregating dependent effect sizes (Dunlap, Cortina, Vaslow, & Burke, 1996).

The second issue relates to the problem identified by Mitchell and James (2001). Because lag length is often a matter of convenience, it varies widely across primary studies. In meta-analyses, this is often dealt with by categorizing panel studies according to length of time lag. This leads to several problems. First, although relationships between effect sizes and time are markedly nonlinear (Dormann & Griffin, 2015), this nonlinearity is unlikely to be detected if the number of categories is *too small*, for example, if only two categories are used (e.g., < 1 day versus > 1 day in Glasman & Albarracín, 2006). Second, even if there are more than two categories, nonlinear relations are also missed if the categories do not possess certain characteristics. Suppose effect sizes start to increase with time, then reach a maximum, and finally level off with a long tail such as is depicted in Figure 2. This shape corresponds to the expected distribution of effect sizes across time (Voelkle et al., 2012). When categories are chosen such that the category with the shortest time lags still contains studies whose lags are past the peak effect size, then one would conclude incorrectly that effect sizes are small and do not change much across time.

This problem is forced upon meta-analysts simply because of some of the realities of meta-analyzing longitudinal primary studies. If a meta-analyst were to adopt a rational or phenomenological time lag categorization scheme, it might result in categories that have very few effect sizes. As a simple example, suppose that a meta-analysis contains 50 primary studies whose lags vary from 1 week to 1 year. A rational categorization might be 1-17 weeks, 18-34 weeks, and 35-52 weeks. But what if most studies use shorter lags? Then there might be 40 effect sizes in the "short" category, 8 in the "medium" category, and only 2 in the "long category". Rather than have such disparity in *k*'s, meta-analysts choose a categorization scheme that balances out the *k*'s. So, maybe it is 17 effects from 1-4 weeks, 17 from 4-12 weeks, and 16 from >12 weeks. Here, we have solved the unbalanced *k*-problem, but we have replaced it with a categorization scheme that does not really allow us to answer the question, "Do effect sizes vary across the different time lags that are used to study this phenomenon?" A categorization scheme like this might prevent the meta-analyst from spotting the most important time trends in the data. For example, suppose effect sizes increase from 1 day to 1 week, then drop precipitously for a couple of weeks, and then level off. If the shortest category is 1-4-weeks as in the example above, the incorrect conclusion would be that effect sizes are modest over short periods of time (as suggested by the mean effect size for the studies in the 1-4 week category) and then drop off gradually (as suggested by the slightly smaller effects in the 4-12 week and the >12 week categories. Finally, as we demonstrate later, even if the time lags were *invariant* across studies and no categorization was necessary, simply aggregating longitudinal correlations or cross-lagged effects, for example, using the commonly applied procedures suggested by Borenstein et al. (2009) or by Hunter and Schmidt (2004), could lead to distorted results and incorrect conclusions.

Some longitudinal meta-analyses in various domains addressed differences in time lags among primary studies. In most cases, it was assumed that effect sizes decay as a gradual, linear function of time (e.g., Atkinson et al., 2000; Cohen, 1993; Griffeth, Hom, & Gaertner, 2000; Holden, Moncher, Schinke, & Barker, 1990; Hom, Caranikas-Walker, Prussia, & Griffeth, 1992; Hulin, Henry, & Noon, 1990; Steel, Hendrix, & Balogh, 1990; Steel & Ovalle, 1984). These meta-analyses used either of two approaches: They split primary studies into two groups using longer versus shorter time lags (e.g., Holden et al.), or they conducted linear meta-correlation or meta-regression analyses of effect sizes and time lags. For example, a meta-correlation approach was used by Steel et al., who correlated the time lags used in primary studies with the meta-analyzed (i.e., aggregated) predictor-outcome effect statistics. Meta-regression aims to relate the size of the aggregated effects to multiple primary study characteristics including time lags (cf. Thompson & Higgins, 2002). Meta-correlation can reveal linear trends only. Meta-regression could detect nonlinearity if polynomial terms were used, although they generally are not, but it too is limited by the time lags that happen to have been chosen at the primary study level.

Few meta-analyses investigate possible non-linear effects of time on effect size. Sowislo and Orth (2013) and Nohe et al. (2015) found neither linear nor nonlinear effects. Some others have found that lagged effects increase across time, then decrease, and eventually level out (e.g., Ford et al., 2014; Maricuțoiu et al., 017; Riketta, 2008). The latter corresponds to the shape that can be expected if the assumptions underlying continuous time modeling apply, and it also demonstrates the necessity of considering the non-linear relation of time and effect size in meta-analyses.

Taken together, some previous meta-analyses of panel studies revealed non-linear relations between effect size and time lag, mainly suggesting that effects are larger across shorter time lags. Several meta-analyses failed to show nonlinearities, which could be due to the chosen categorization or the type of statistical analyses. Effects are likely to be underestimated if possible non-linear relations are not appropriately modelled. Therefore, possible non-linear relations have to be considered carefully. However, our theories are generally too imprecise to suggest specific periods across which cross-lagged effects increase, when they become strongest, or when they start to decline. Hence, categorizing primary studies into meaningful groups using 'too short', 'optimal', and 'too long' intervals is not possible a priori. Even if it were possible, non-linear effects of time on effect sizes *within* these groups can still remain, and they must be taken into account in order to obtain a clear picture of the relationship in question.

An alternative to categorization of time lags involves *continuous time modeling.* In the next section, the conceptual and mathematical underpinnings of continuous time modeling are briefly reviewed. First, we describe how the unfolding of cross-lagged effect sizes across time takes place. Second, we demonstrate that even if the sizes of cross-lagged effects based even on identical time lags, they cannot be directly compared, which would be a prerequisite for aggregating them. Afterwards, we show how continuous time modeling solves these problems, and how continuous time modeling of CLPC matrices can be done with ctsem (Driver, Oud, & Voelkle, 2017), which is a package for R (R Development Core Team, 2011). Finally, the approach is then extended to multi-sample models, which allows for CoTiMA.

Continuous Time Modeling

Consideration of time is extremely important for comparing or aggregating findings of relations among variables measured at different measurement occasions. Consideration of time would even be important if studies used identical time lags, because the 'pace of change' could be different: A small effect in a very stable causal system may appear larger than it actually is. Therefore, we begin with a demonstration of how to compare cross-lagged effects of studies using identical time lags, which sets the stage for understanding how different time lags can be accounted for.

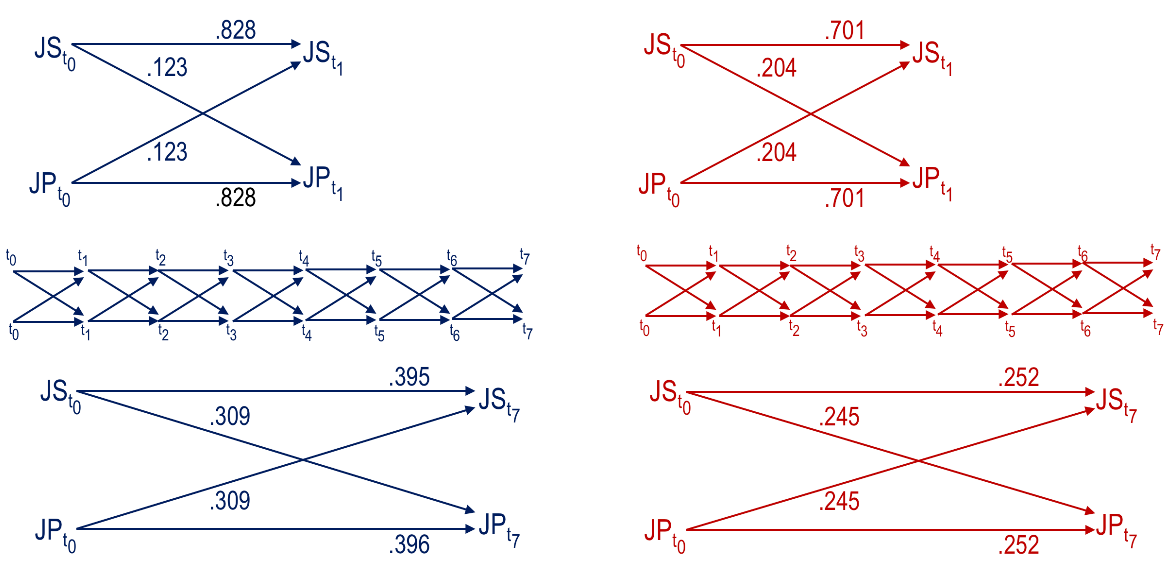

Figure 1. Cross-lagged panel SEM results of two hypothetical studies with identical designs.
Note: The path diagrams in the top part show the SEM parameter across one lag (e.g., 1 month) with larger cross-lagged effects in the red study, the middle part visualizes the stationary extension of the parameters of the top part across six further lags (seven altogether), and the path diagrams in the bottom part show the SEM parameters across seven lags with larger cross-lagged effects in the blue study.

Suppose two panel studies, which we will call 'red study' and 'blue study', of the relationship between job satisfaction and job performance used identical time lags, and our aim was to meta-analyze the cross-lagged effect (i.e., satisfaction at Time 0 predicting performance at Time 1). Further suppose that both studies were identical in terms of design, measurement tools, sample size etc. The only variables that might have varied were contextual variables like perceived organizational support, high performance work systems, etc. The measurement occasions were separated by one month in both studies. The data from each of the two studies were analyzed by standard (discrete time) SEM and yielded the unstandardized results shown in the top panel Figure 1. Note that for simplicity we used identical values for several effects, but this does not affect the subsequently derived conclusions.

Since everything except the context was identical, it seems to be well justified to simply aggregate the findings by computing the average of the two cross-lagged effects of job satisfaction at t0 on performance at t1, which yields (.123 + .204)/2 = .164 as the estimated population parameter. One could also test if the two cross-lagged effects vary between the two studies. If this test were significant, then it seems well justified to conclude that the effect of satisfaction on performance was smaller in the blue compared to the red study. Both seemingly reasonable conclusions are, in fact, incorrect. Indeed, small is sometimes large, or at least large for some times.

Small is sometimes large, or at least large for some times.

For the sake of simplicity, suppose that in the subsequent months the processes and the strength of effects remain invariant, that is, the effects from t0 to t1 are identical to the effects from t1 to t2 etc. Then suppose the process unfolds across 6 additional months, as sketched in the middle panel of Figure 1. Finally, suppose only t0 and t7 data are available for analysis, and a cross-lagged SEM yields the estimates shown in the bottom of Figure 1.

Although the causal system has not changed, the cross-lagged effect is now larger for the blue study then for the red study, the opposite of what it was before. Therefore, in fact, we could never say in which study/context job satisfaction has a larger causal effect. This is why directly comparing cross-lagged effects across studies is usually not meaningful. The reason for the change in relative magnitude of the effect of satisfaction is that any cross-lagged effect is influenced by the size of the reciprocal cross-lagged effects *and* the size of the autoregressive effects.

The autoregressive effects can be thought of as 'pace of change' values, which impact the interpretation of the cross-lagged effects, making it impossible to directly compare or aggregate them. Doing so would be a bit like saying that the average high temperature over the last two days was 50 degrees because yesterday it was 20 Celsius and today it was 80 Fahrenheit. Meta-analysts do it anyway because they feel the need to make more general claims, but it does not really make sense, just as it makes no sense to compare temperatures before they are re-scaled to a common scale. Linear transformation can do this for Celsius and Fahrenheit, and continuous time math can do it for cross-lagged effects.

The pace of change was larger in the red study than in the blue one, which is indicated by the lower autoregressive effects (i.e., less consistency over time). As a result of these differences, the effect of satisfaction on performance one month later is stronger in the red study early on but stronger in the blue study later. If the primary studies on which our meta-analysis was based had happened to collect data at t0 and t1, we would draw one set of conclusions from our meta-analysis. If the primary studies had collected data at t0 and t7, we would have drawn entirely different conclusions. However, it is possible to arrive at a single conclusion that is nevertheless consistent with the different observations through the use of continuous time math.

If the two autoregressive effects are put in the diagonal of a 2 x 2 matrix (job satisfaction in the top left) and the cross-lagged effects off the diagonal (job satisfaction t0 on job performance t1 in the bottom left), we have matrices A_B and A_R in Equations 1 and 2 below. Continuous time math involves the nonlinear transformation of these matrices into what are called ‘drift matrices’. Specifically, taking the logs (matrix logarithms, *logm*) of these matrices produces the drift matrices, which contain ‘drift coefficients’.

$$\mathbf{A}_B = logm\begin{pmatrix} .828 & .123 \\ .123 & .828 \end{pmatrix} = \begin{pmatrix} -.20 & .15 \\ .15 & -.20 \end{pmatrix} \tag{1}$$

$$\mathbf{A}_R = logm\begin{pmatrix} .701 & .204 \\ .204 & .701 \end{pmatrix} = \begin{pmatrix} -.40 & .30 \\ .30 & -.40 \end{pmatrix} \tag{2}$$

Because we are typically more interested in the cross effects rather than the auto effects, we focus our attention on the off-diagonal elements of the drift matrix. A continuous time drift coefficient can be heuristically interpreted as the ‘logarithmized’ magnitude of change across a given lag, one month in the case of the present example. The unfolding of these ‘logarithmized’ changes across any real and discrete time lags is nonlinear. This is analogous to logistic regression. Logistic regression copes with the intrinsic nonlinearity of the relationship between X and dichotomous Y through the use of a logistic weight that represents not magnitude of change in Y but magnitude of change in ln(odds), which does not lend itself to interpretation. Similarly, drift coefficients represent the derivatives of X and Y with regard to time (Voelkle, et al., 2012). Finally, just as ln(odds) become negative for odds < 1.0, the positive autoregressive effects turn into negative auto effects in continuous time modeling. The more negative they are, the less stable is the variable.

Whereas we usually (in discrete time) refer to autoregressive and cross-lagged effects, in continuous time the drift coefficients are labeled auto effects and cross effects (Voelkle et al., 2012). The cross effect of job satisfaction on performance is .15 in the blue study, which is smaller than its .30 counterpart in the red study. We now can say that, in fact, job satisfaction has a stronger effect on performance in the red study than in the blue study. We can even say that the cross effect of job satisfaction is twice as large in the red sample compared to the blue sample. Moreover, the average of these two effects, that is (.15 + .30)/2 = .225 is now meaningful. Note, that these drift coefficients do not lend themselves to interpretation and can only be interpreted as a ‘logarithmized’ change. In discrete time, the interpretation of a one-month, cross-lagged effect of, for example, .30 would be ‘one unit change in job satisfaction leads to a .30 unit change in performance across a time lag of one month, holding performance at t0 constant’. In continuous time, there is no intuitive counterpart to this discrete time interpretation. Therefore, in order to assign meaning to drift coefficients, we have to convert them back into discrete time values as we show later. This is again similar to logistic regression, in which the regression weight is used to compute p(Y=1|X) for specific values of X.

The *relative* sizes of the continuous time drift effects are independent of the actual discrete time lag used. To provide a brief demonstration, we use the values obtained for the seven month lagged analyses shown in the bottom part of Figure 1 for computing the matrix logarithm:

$$\mathbf{A}_B = logm\begin{pmatrix} .395 & .309 \\ .309 & .395 \end{pmatrix} = \begin{pmatrix} -1.40 & 1.05 \\ 1.05 & -1.40 \end{pmatrix} \tag{3}$$

$$\mathbf{A}_R = logm\begin{pmatrix} .252 & .245 \\ .245 & .252 \end{pmatrix} = \begin{pmatrix} -2.80 & 2.10 \\ 2.10 & -2.80 \end{pmatrix} \tag{4}$$

The resulting drift coefficients are different for the two studies. However, the *ratios* of the drift coefficients of the blue study to those of the red study are the same for the seven-month interval as they were for the one-month interval. As before, they are twice as high in the red compared to the blue study. The average continuous time cross effect is now (1.05 + 2.10)/2 = 1.575, which, obviously, is seven times larger than before (7 × .225).

The key point here is that it does not matter if one would use the drift coefficients of the red study across one month, or the drift coefficients of the red study across seven months divided by 7 as input for a meta-analysis. In either case, effect sizes are two times larger than in the blue study. It may seem that there is an 'apples and oranges' problem here, but continuous time drift coefficients could be simply multiplied/divided by time lag to, as it were, convert oranges into apples. This then allows one to compare effect sizes or to aggregate them (i.e., to compare apples with apples) even though they are based on different lags.

Time-scaling of cross effects.

Moreover, the average cross effect from *either* meta-analysis could be used to reach the same conclusion about the cross-lagged (i.e., discrete time) effect for any given time lag. For example, using the drift matrices obtained for one-month lags (Equations 1 and 2) to predict the discrete effects for four-month lags yields Equations 5 and 6:

$$\mathbf{A}^*_{\mathrm{B}.\Delta 7} = e^{\left(\begin{pmatrix} -.20 & .15 \\ .15 & -.20 \end{pmatrix} \times 4\right)} = \begin{pmatrix} .533 & .286 \\ .286 & .533 \end{pmatrix} \tag{5}$$

$$\mathbf{A}^*_{\mathrm{R}.\Delta 7} = e^{\left(\begin{pmatrix} -.40 & .30 \\ .30 & -.40 \end{pmatrix} \times 4\right)} = \begin{pmatrix} .366 & .305 \\ .305 & .366 \end{pmatrix} \tag{6}$$

The predicted four-month discrete time coefficients are identical to those that one would obtain if the primary studies had used seven-month lags, as evident from Equations 7 and 8:

$$\mathbf{A}^*_{\mathrm{B}.\Delta 7} = e^{\left(\begin{pmatrix} -1.40 & 1.05 \\ 1.05 & -1.40 \end{pmatrix} \times \frac{4}{7}\right)} = \begin{pmatrix} .533 & .286 \\ .286 & .533 \end{pmatrix} \tag{7}$$

$$\mathbf{A}^*_{\mathrm{R}.\Delta 7} = e^{\left(\begin{pmatrix} -2.80 & 2.10 \\ 2.10 & -2.80 \end{pmatrix} \times \frac{4}{7}\right)} = \begin{pmatrix} .366 & .305 \\ .305 & .366 \end{pmatrix} \tag{8}$$

Equations 5 to 8 show how drift coefficients can be converted to cross-lagged multiple regression coefficients that reflect the number of units by which Y1 changes, over a certain time lag, per unit increase in X0 holding Y0 constant . We can use this last point to demonstrate how to conduct meta-analyses for panel designs that differ in time lag. Suppose the blue study used one month and the red study used seven months. This can be properly dealt with by either first estimating and then multiplying the blue continuous time coefficients by 7 (e.g., .15 × 7 = 1.05), or by estimating and then multiplying the red continuous time coefficients by 1/7 (e.g., 2.10 × 1/7 = .30). Categorization of time lags is no longer necessary. This is what CoTiMA does: First, based on longitudinal correlations, continuous time coefficients are estimated for each primary study, then they are multiplied by the appropriate factors to re-scale them to an identical time scale, and finally they are aggregated.

As noted earlier, continuous time drift coefficients do not lend themselves to interpretation. Fortunately, this is no real drawback because we can directly transform our 'logarithmized' continuous time effects into discrete time regression coefficients. This could be done for any discrete time lag that one required. This is done by multiplying the right-hand continuous time matrices by the desired time lag (number of months in our example), and then applying the matrix exponential function, which just makes discrete time coefficients out of continuous time ones. For example, if we replace the factor 4 in Equations 5 and 6 and execute the computation for every integer lag from 1 to 120 months (fractions are also possible, such as 0.7 months) and plotted the discrete cross-lagged effects for the blue and red studies, we would have Figure 2.

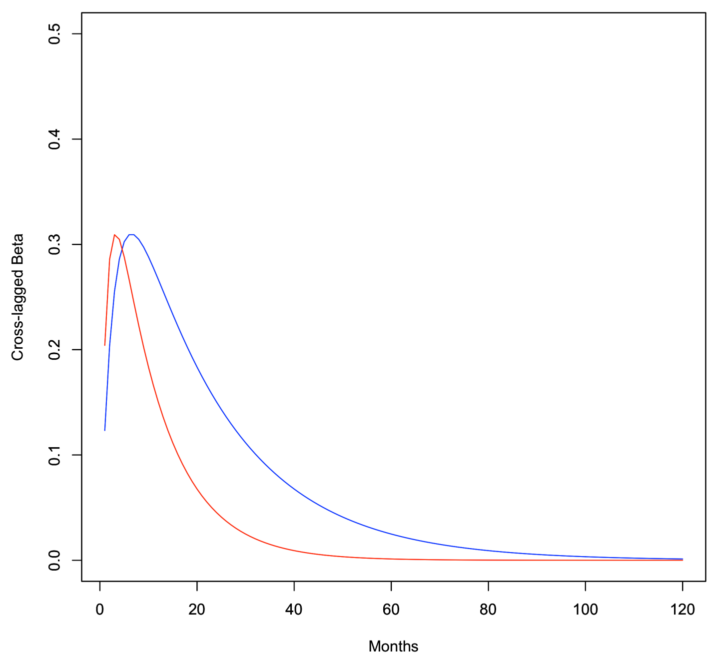

Figure 2. Sizes of the expected discrete cross-lagged effects of X on Y across different possible discrete time lags for the hypothetical Blue Study (blue line) and Red Study (red line).

Here, we see that the cross-lagged effect of X (satisfaction) on Y (performance) would peak at around .30 for both studies, but that this peak would be reached sooner in the red study. Moreover, the drop-off is steeper in the red study, with the effect dropping below .10 for time lags of about 1.5 years and vanishing completely after about 3.5 years. The peak for the blue study is a few months later, and the drop-off is more gradual, with the effect still being detectable even after the effect from the red study would have disappeared[2].

In summary, both drift coefficients and raw coefficients have their benefits. The major benefit of drift coefficients is that they, unlike raw coefficients, can be simply multiplied with time lag in order to estimate the drift coefficients that would be obtained for any time lag. These values can then be used to compare and aggregate effects even though the primary studies used different lags. The major benefit of raw coefficients is that they allow "standard" interpretation of effect magnitude

[2] We should note that when estimating continuous time parameters, one also has to take error terms (i.e., residuals, disturbances, or unexplained variance) into account. This is intuitive insofar as one can imagine that across very long time lags it becomes more difficult to predict outcomes and explain much of their variance. In other words, the unexplained variance is larger for longer lags. For a more detailed treatment of this topic, see Voelkle et al. (2012).

across the time lag on which they were based (e.g., a raw coefficient of .2 based on lag t means that an increase in X of 1.0 units results in a .2 unit increase in Y where Y is measured after lag *t* had elapsed). The problem is that this coefficient cannot be used to estimate what the coefficient would have been for 2*t* or .5*t*, but the drift coefficient can.

Continuous Time Meta-Analysis (CoTiMA)

Taking time into account in meta-analysis is extremely important, but it requires a different approach than typical, correlation-based meta-analyses. Correlations among variables *emerge* because structural relations (causal effects, reciprocal effects, random processes etc.) among variables *exist*, rather than the other way around.

Table 1. Differences Between Correlation-Based and Parameter-Based MASEM in Estimating a Regression Parameter (β_2) Using the Random Effects Aggregation Procedure described in Borenstein, Hedges and Rothstein (2007)

Data		*Study 1* JS0 P0 JS1 JS0 1.00, 0.11, 0.64 P0 0.11, 1.00, 0.20 JS1 0.64, 0.20, 1.00 *Study 2* JS0 P0 JS1 JS0 1.00, 0.15, 0.53 P0 0.15, 1.00, 0.19 JS1 0.53, 0.19, 1.00	
		Correlation-based MASEM	*Parameter-based MASEM*
Step 1		Aggregate correlations	Estimate regression slopes: JS1 = $\beta_0 + \beta_1$ JS0 + β_2 P0 + ε
	Results	1.00, 0.12, 0.62, 0.12, 1.00, 0.20, 0.62, 0.20, 1.00	For β_2 (SE): Study 1: β_2 = .131 (.051) Study 2: β_2 = .113 (.106)
Step 2		Estimate regression: JS1 = $\beta_0 + \beta_1$ JS0 + β_2 P0 + ε ε	Aggregate regression slopes
	Results	β_2’ = .126 (.046)	β_2’ = .128 (.046)

This is contrary to correlation-based MASEM, which represent an *aggregated correlations first-structural relations later* approach. Consideration of time requires a *structural relations first-aggregated effects later* approach. This is called parameter-based MASEM. According to Cheung and Cheung (2016) *"Correlation-based MASEM basically follows the random effects model on the correlation matrices [...] in the first stage of the analysis, whereas the estimated average correlation matrix is used to fit structural models in the second stage of the analysis [...]. Because the meta-analysis is applied to the correlation matrices, it is called correlation-based MASEM."* (p. 143) A parameter-based MASEM considers *"the parameter estimates or functions of the parameter estimates as the effect sizes [...]. Here, this approach is called parameter-based MASEM because meta-analysis is applied to the parameter estimates or functions of the parameter estimates rather than to the correlation coefficients."* (p. 144). Table 1 illustrates these differences between correlation-based and parameter-based MASEM using two correlations matrices, which show the correlations among job performance t0, job satisfaction t0, and job satisfaction t1 reported by Ashforth and Saks (1996) and Bechtold, Sims, and Szilagyi (1981).

In CoTiMA, based on the correlations reported in primary studies, the structural relations (i.e., cross effects and auto effects) are estimated first using continuous time SEM (CTSEM). It is these structural coefficients, and not correlations, that are then aggregated meta-analytically. Parameter-based meta-analysis is necessary for CoTiMA because each drift coefficient is scaled by the particular time lag in that study and would lose its meaning if it were computed from aggregated correlations, where the particular time lags of the primary studies are not taken into account. In the case of two-variable panel studies, CTSEM would generate 10 values for each primary study: 2 t0 variances and 1 t0 covariance, 2 error variances and 1 error covariance for later waves, and 2 auto and 2 cross effects. CoTiMA then aggregates the 2 auto and 2 cross effects, by constraining them to be invariant across primary studies. CoTiMA can also be extended to more than two waves, a topic that we take up in the Discussion section.

In CTSEM, discrete and continuous time parameters are estimated simultaneously. Hence, requirements (e.g., sample size) and most assumptions (e.g., multivariate normality) are identical for discrete and continuous time analyses. One additional assumption of continuous time analysis, however, is, *stationarity*. The process parameters are assumed to be invariant over the course of a study. This means that autoregressive and cross-lagged effects (and error variances and covariances) over a given lag should be identical regardless of whether that time lag occurred earlier or later (e.g., January 1 – January 31 versus March 1 – March 31). Further assumptions which have to be made for continuous time analysis and CoTiMA are typically made in panel studies anyway. For example, we assume that a cause (e.g., job satisfaction) precedes its effect (e.g., performance).

As noted earlier, most previous meta-analyses of panel studies aggregate correlations first, and then use them to estimate process parameters (e.g., cross-lagged effects) in a regression or SEM model. CoTiMA estimates the parameters of interest (e.g., cross effects) using a multi-sample continuous time SEM, with each study representing a different “sample”. The continuous time SEM part handles time related-information, and the multi-study part allows constraining of parameters to be invariant across studies as a means to aggregate effects. It also allows one to test this invariance assumption. The aggregated (invariant) parameters are estimated by minimizing the maximum likelihood fit function across all primary studies simultaneously, that is, the aggregated parameter is the one for which the sum of the log-likelihoods of all primary studies becomes smallest (e.g., Asparouhov & Muthén, 2012, Eq. 6).

In the next section, we briefly outline how estimating continuous time parameters could be accomplished using ctsem. Continuous time parameters are used afterwards as a basis for the CoTiMA approach.

Continuous Time Modeling and CoTiMA with “ctsem”

Equations 1 to 6 show how discrete time and continuous time parameters are related, and similar equations exist for error variances and intercepts (cf. Voelkle et al., 2012). To put it differently, discrete time parameters are a function (i.e., matrix exponentiation) of continuous time parameters multiplied by the respective time lag. Whereas simple functions could be easily used as constraints in many SEM programs (e.g., $c = a \times b$), the matrix exponential function is not yet implemented in common SEM packages such as MPLUS or AMOS, but the R package ctsem (Driver et al., 2017) implements all the necessary math. This involves equations for the drift parameters, which are typically of most importance, but also the necessary equations for continuous time intercepts and disturbances (for equations see Voelkle et al., 2012). Since ctsem implements all the necessary equations, it is sufficient to know that the variances and covariances of the disturbances are represented in a symmetric matrix called ‘diffusion’ (cf. Voelkle et al., 2012).

We developed two R-Scripts that facilitate conducting CoTiMA using ctsem, and which are available as online supplements to this article. Users just have to open the first (“edit”) script (CoTiMA V1.1 edit.R), which is shown in Appendix A and available as online supplement S-1, and enter the correlations, sample sizes, and time lags of the primary studies. Then the script is run in an R environment. Further options could be selected or de-selected such as “fixed and random effect analysis”, plots, and more. The second of the two R-scripts (CoTiMA V1.1 noedit.R) is a “noedit” script, and it is available as online supplement S-2. It has to be saved in the same directory as the first script. When running the edit script, the noedit script is automatically executed and performs the model fitting, presents model statistics and estimated parameters in a results file, and creates the figures to illustrate the results.

A possible shortcoming of the user-friendly “edit” script is the complexity of the accompanying “noedit” script. To facilitate understanding of the full workflow, Appendix B presents a flowchart illustrating the different steps performed in conducting a CoTiMA, their purposes, and the respective findings. The online supplement S-3 follows this flowchart and illustrates the workings of each step using two empirical correlation matrices.

Having described CoTiMA, we next illustrate its use and its differences with traditional MASEM by conducting both sorts of analyses on studies from the Riketta (2008) meta-analysis.

CoTiMA of the Job Satisfaction-In-Role Performance Relation

Previous meta-analyses of panel studies are limited because they do not deal appropriately with variability in time lags. In many cases, lengths of time lags are ignored altogether. Where they are considered, only linear trends are analyzed. In the rare instances in which more than two categories of time lags (e.g., short, medium, long) are used to detect possible nonlinearities, the categorization is based on convenience, intuition, or some notion of balance (i.e., equal numbers of effects in each category). CoTiMA incorporates time lags without the need to categorize them, and it can result in very different conclusions regarding relationships of interest. We demonstrate this by comparing results of a CoTiMA with the common correlation-based meta-analysis of longitudinal studies.

To begin with, we used correlations of job satisfaction and in-role performance from primary studies included in Riketta (2008, Table 3), as a starting point. These correlations are shown in the first panel of Table 2. In the second panel, the correlation of these effect sizes with the length of the time lag, the squared time lag, and time lag first centered and then squared are shown. All correlations with time failed to reach significance, but they suggest that there could be a small linear decline of correlations with time, or that there could be a slight non-linear relation between time and correlation. The third panel of Table 2 shows the random effects correlations across studies, which we aggregated using the random effect approach of Borenstein et al. (2009).

As outlined earlier, previous meta-analyses sometimes fully ignore time lags while others categorized primary studies into, for example, short, medium, and long. Like Riketta (2008), we used three categories covering 1 to 6 month time lags, 7 to 12 month time lags, and time lags with 13 or more months. This categorization was a compromise between conflicting aims. Not unreasonably, Riketta (p. 747) aimed at having a fair number of studies in each category. Unfortunately, this meant that a maximum of three categories was possible. Since he also aimed at analyzing possible nonlinear effect of time lags, the lower limit of categories was also three. Such compromises are unnecessary in CoTiMA.

Results of correlation-based structural equation meta-analyses.

Discrete cross-lagged panel SEM were fitted to four sets of aggregated random effect correlations (i.e., those based on all studies, on studies with 1-6 month time lags, on studies with 7-12 month time lags, and on studies with 13 or more month time lags). Online supplement S-4 provides an example using two correlation matrices of how we conducted this MASEM. The results are shown in the last panel of Table 2. The first rows show that across all time lags there was no evidence for an effect of job satisfaction on in-role performance (β_{P2JS1}), whereas the reversed effect of performance on later satisfaction (β_{JS2P1}) was significant.

Table 2. Correlations of Two-Wave Primary Studies of Job Satisfaction and In-role Performance (adapted from Riketta, 2008, Table 2), Their Correlation with Time, Aggregated Random Effect Correlations, and Parameter Estimates based on Aggregated Correlations in Different Time Lag Categories (Category-Based MASEM)

			Correlations (from Riketta; 2008, Table 2)					
Study	*N*	Lag	r_{JS1JS2}	r_{P2JS1}	r_{JS2P1}	r_{P1P2}	r_{JS1P1}	r_{JS2P2}
Ashforth & Saks (1996)	222	6	.64	.14	.20	.69	.11	.21
Bechtold et al. (1981)	64	18	.53	.17	.19	.57	.15	.21
Bond & Bunce (2003)	412	12	.66	.05	.66	.21	.26	.17
Griffin (1991a)	526	18	.61	-.02	.06	.53	.04	.06
Nathan et al. (1991)	300	3.5	.56	.14	-.02	.23	.06	.17
Sheridan & Slocum (1975a)	59	11	.45	-.08	-.06	.50	-.03	.15
Sheridan & Slocum (1975b)	35	12	.68	.21	.24	.49	.20	.21
Szilagyi (1980)	128	3	.62	.09	.09	.65	.09	.05
Tharenou (1993)	200	12	.48	.11	.08	.64	.19	.08
Wanous (1974)	80	2	.73	.18	.24	.44	.09	.15
			Correlations of Time with Correlations Above					
		Lag	-.35	-.29	.10	.10	.20	.03
		Lag^2	-.29	-.25	.05	.11	.12	-.03
		Lag^2cnt	.21	.15	-.02	.07	-.30	-.24
			Aggregated Random Effects Correlations (own computations, not disattenuated)					
	2.026	all lags	.61	.08	.17	.50	.12	.13
	730	lag 1-6	.63	.15	.12	.51	.10	.17
	706	lag 7-12	.58	.06	.24	.46	.20	.03
	590	lag 13	.62	.04	.08	.47	.04	.09
			Discrete Autoregressive and Cross-Lagged Regression Coefficients Based on Aggregated Correlations from the Panel Above					
			β_{JS2JS1}	β_{P2JS1}	β_{JS2P1}	β_{P2P1}		
	2.026	all lags	.60 (.02)	.02 (.02)	**.10 (.02)**	.49 (.02)		
	730	lag 1-6	.63 (.03)	**.10 (.03)**	**.06 (.03)**	.50 (.03)		
	706	lag 7-12	.55 (.03)	-.03 (.03)	**.13 (.03)**	.46 (.03)		
	590	lag 13+	.62 (.03)	.02 (.04)	.05 (.03)	.47 (.04)		

Note: Lag in months. Significant parameter estimates (p < .05, two-tailed) in bold face. Lag = time lag between waves in months. Lag^2cnt = centered time lag squared. JS = job satisfaction; P = in-role performance.

Across all analyses, only a single effect of job satisfaction on later performance was significant: this was obtained in studies using time lags of six or fewer months. The effect was positive, suggesting that job satisfaction increased performance. Across longer time lags, the cross-lagged effects became smaller and nonsignificant. On the other hand, for effects of prior performance on later satisfaction, all analyses except across time lags exceeding 12 months yielded significant and positive effects. The results suggest that the reversed cross-lagged effect first increased with time, then reached its maximum, and then started to become smaller again. Thus, based on the findings of this category-based MASEM, one could conclude that job satisfaction might impact subsequent in-role performance and vice versa, but these effects are unlikely to be observed across time lags exceeding 6 or 12 months, respectively.

Results of the parameter-based CoTiMA of satisfaction and in-role performance.

Next, a CoTiMA was conducted. First, we tested a heterogeneity model of the ten primary studies, in which all parameters were allowed to vary freely across the studies. The four drift coefficients and their standard errors for all ten samples are presented in the top panel of Table 3. The parameter estimates of a heterogeneity model are identical to those obtained when estimating the ten models separately, and the overall fit (LL) is identical to the sum of the ten fit values.

Regarding the auto effects, the heterogeneity models yielded, as expected, negative auto effects for job satisfaction and for performance. The negative sign is just a result of the continuous time math; it indicates a positive relation over time. As explained earlier, the more negative a continuous time auto effect is, the more volatile a variable is, and the closer an auto effect is to zero, the more stable a variable is. There was no clear evidence which variable is more stable; some studies (e.g., Wanous, 1974) revealed that job satisfaction is more stable, whereas others (e.g., Tharenou, 1993) revealed that job performance is more stable.

Table 3. CoTiMA Estimates of ctsem-based Drift Parameters of the Relation of Job Satisfaction and In-role Performance Over Time (Explanations in the Text)

Study	Auto JS	*SE*	JS -> P	*SE*	P -> JS	*SE*	Auto P	*SE*
				Heterogeneity Model				
Primary Studies								
Ashforth & Saks (1996)	**-.080**	.014	.017	.013	**.034**	.013	**-.065**	.012
Bechtold et al. (1981)	**-.038**	.012	.009	.011	.012	.012	**-.033**	.011
Bond & Bunce (2003)	**-.053**	.008	-.001	.012	**.126**	.019	**-.128**	.026
Griffin (1991a)	**-.027**	.003	-.004	.004	.003	.003	**-.035**	.004
Nathan et al. (1991)	**-.159**	.025	**.097**	.043	-.041	.037	**-.420**	.068
Sheridan & Slocum (1975a)	**-.074**	.024	-.013	.022	-.009	.023	**-.064**	.021
Sheridan & Slocum (1975b)	**-.036**	.017	.018	.024	.016	.020	**-.065**	.028
Szilagyi (1980)	**-.161**	.038	.017	.036	.018	.037	**-.146**	.035
Tharenou (1993)	**-.061**	.011	-.002	.008	-.002	.009	**-.037**	.007
Wanous (1974)	**-.186**	.059	.130	.098	**.162**	.076	**-.450**	.129
LL (#params = 100; df = 0)				21079.50				
Heterogeneity (I^2)	87.403		21.836		82.657		87.293	
95% CI up	92.500		61.748		90.155		92.444	
95% CI low	78.843		-59.721		69.448		78.630	
Random Effects Results								
	-.070	.011	.002	.004	**.022**	.011	**-.077**	.013
95% CI up	-.048		.011		.044		-.052	
95% CI low	-.092		-.006		.001		-.102	
p-value	$p < .001$		$p < .295$		$p < .023$		$p < .001$	
				Multi-Sample Homogeneity Model				
	-.047	.003	.005	.004	**.036**	.004	**-.062**	.005
LL (#params = 64; df = 36)				21.533.12				
ΔLL (df = 36); *p-value*				453.626; $p < .001$				

Note: #params = number of estimated parameters. Significant parameter estimates ($p < .05$, two-tailed) in bold face. LL = minus 2 loglikelihood value. JS = job satisfaction; P = in-role performance.

The signs of the cross effects of satisfaction on performance and vice versa were mixed; some were positive and others were negative. Contrary to the signs of auto effects, the signs of continuous time cross effects can be interpreted in the same way as in discrete time. Thus, some studies seem to suggest that satisfaction and performance mutually enhance each other, while others imply mutually inhibiting effects. However, only four out of 20 cross effects were significant. The positive cross effect of satisfaction on performance was significant in the study of Nathan, Mohrman, and Milliman (1991, 3.5month time lag) only. The reversed cross effect of performance on satisfaction was significant and positive in the studies of Ashforth and Saks (1996, 6month time lag), Bond and Bunce (2003, 12month time lag), and Wanous (1974, 2month time lag).

Before applying multi-sample analysis to aggregate the effects, we first used traditional aggregation methods as suggested by Borenstein et al. (2009) for reasons of comparison. First, we assessed heterogeneity of the four drift coefficients. The second panel of Table 3 shows I^2 (e.g., Borenstein et al., 2010) values and their confidence intervals. In virtually all meta-analyses, there is some true heterogeneity of effects among primary studies, and I^2 quantifies the proportion of the observed variance that can be attributed to this heterogeneity. The 95% confidence intervals around I^2 are also shown; as in most meta-analyses (cf. Borenstein et al., 2010) they are relatively wide. In general, the results show that the heterogeneity of drift effects is very large except the cross effect of job satisfaction on job performance, for which I^2 was not significant.

The next panel of Table 3 shows the results from random effects analysis of the four single drift coefficients, that is, the weighted means of the drift coefficients ($\overline{T}$) as suggested by Higgins and Thompson (2002, Borenstein et al., 2009). Both aggregated cross effects were very small with the cross effect of job satisfaction on performance being .002 and not significant ($p < .295$). The reversed effect was .022 and significant ($p < .023$). Thus, in line with Riketta (2008), the random effects analysis confirmed cross effects of performance on satisfaction.

Although the previous random effects analyses provide meaningful aggregated results, we advocate the use of multi-sample modeling with particular constraints set across studies for aggregation. Compared to single, univariate effect sizes, with more complex models, *"conventional univariate meta-analytic techniques (e.g., Hedges & Olkin, 1985; Schmidt & Hunter, 2015) may not be sufficient to handle the dependence of the effect sizes. More advanced techniques, such as multivariate meta-analysis, three-level meta-analysis, or meta-analytic structural equation modeling (MASEM), may be required to handle the dependent nature of the effect sizes or to address the research questions (M. W.-L. Cheung, 2015b)"* (Cheung & Cheung, 2016, p. 140). Therefore, CoTiMA applies multi-sample MASEM, which offers the great advantage that the structure of the entire process and the dependencies among the effects is taken into account when minimizing the fit function. For example, the correlation of X with Y, their reciprocal effects, and the residuals are not taken into account if traditional meta-analysis is used to aggregate the auto effect of X. On the other hand, if all four drift coefficients are simultaneously constrained to be invariant across primary studies, the multi-sample fit function yields the optimal estimate for each of them conditional on the optimal estimates the respective other three effects.

The next panel of Table 3 shows the CoTiMA results. The results confirm a non-significant effect of satisfaction on performance and a significant effect of performance on satisfaction. As expected based on the I^2 values, which already revealed that heterogeneity exists, this model fitted significantly worse than the heterogeneity model. However, this model provided the aggregated overall drift matrix required for plotting the discrete time effects across different time lags. These plots are shown in Figure 3. The dashed black lines represent the discrete time cross-lagged effects based on the aggregated drift matrix. The solid grey lines represent the discrete time cross-lagged effects of different primary studies.

The top part (A) of Figure 3 shows the discrete time cross-lagged effects of satisfaction on performance across possible time lags. Recall that I^2 of the continuous time cross effects was small and that they were not significantly heterogeneous, and the discrete time cross-lagged effects indeed appear quite homogeneous. Two studies (Nathan et al., 1991; Wanous, 1974) implied increasing discrete effects until 3.5 months, with decreasing effects thereafter. These two studies had the weakest (most negative) auto effects and the strongest cross effect. Nevertheless, analyses revealed that in general, there is no (significant) effect of job satisfaction on in-role performance. This is mirrored in the very flat dashed black line.

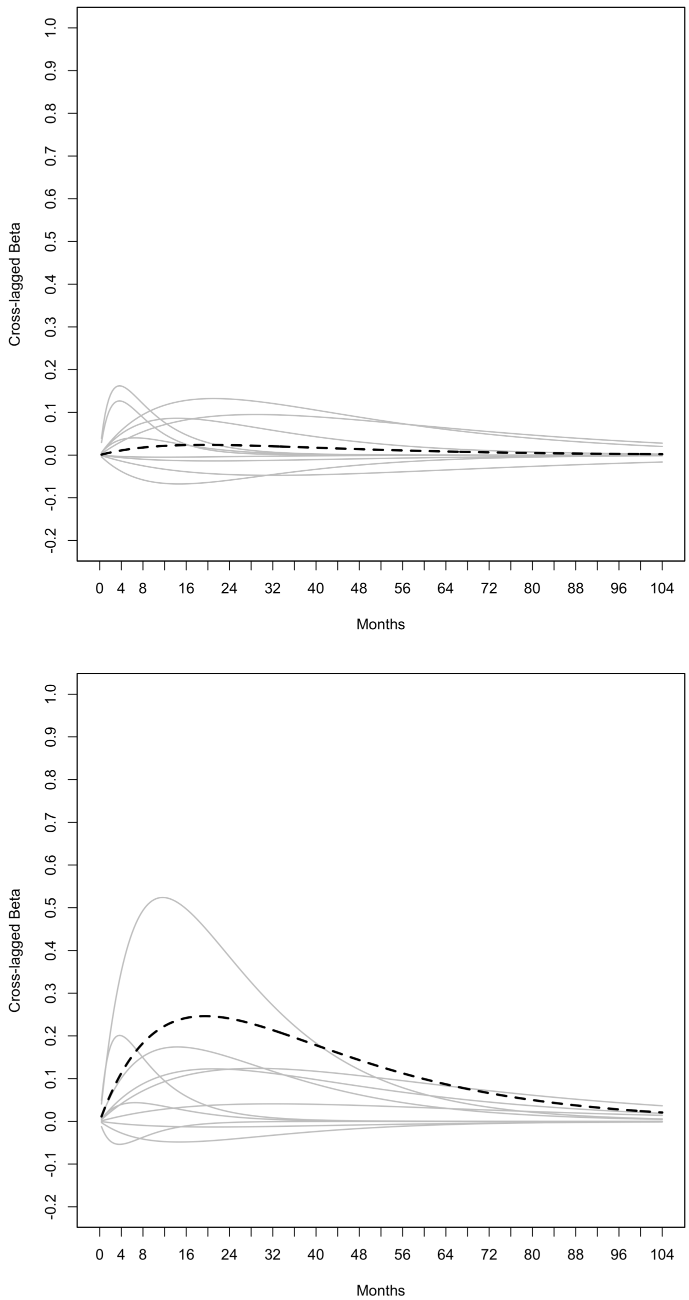

Figure 3. Sizes of the expected discrete cross-lagged effects of satisfaction on performance (Part A) and of performance on satisfaction (Part B) across different possible time lags for the ten primary studies listed in Table 2 (solid grey lines) and based on a CoTiMA of all ten studies (dashed black lines).

The reversed discrete time cross-lagged effects of performance on job satisfaction, which are shown in the bottom part (B) of Figure 3, appear more diverse. This was already apparent in large I^2 value. On average, the strongest discrete time cross-lagged effects occur across time lags below around 19 months (black dashed line), but the studies that yielded particularly strong effects typically suggest that they occur across time lags shorter than 12 months. In particular, these were the studies by Bond and Bunce (2003) and Wanous (1974).

The results thus far revealed that the effect of job satisfaction on performance was very small and close to zero, whereas the reversed effect of performance on satisfaction was significant. We also statistically compared the magnitudes of the two cross effects in order to determine if the effect of satisfaction on performance differs from that of performance on satisfaction. To do so, we first tested a model in which each of the two cross-effects were invariant across primary studies but different from each other. In the second model, the two cross effects were then constrained to be equal. The fit of the second model was significantly worse than the fit of the first model ($\Delta\chi^2 = 31.293$, $\Delta df = 1$, $p < .01$). Thus, the effect of performance on job satisfaction is significantly larger than the reversed effect of job satisfaction on performance.

Table 4: Summary and Interpretations of the Job Satisfaction-Performance Relations Using Different Meta-Analytic Approaches

Analysis	Pattern of Findings		Interpretation
	Satisfaction -> Performance	Performance -> Satisfaction	
Correlation-based MASEM without categorization	• no effect	• significant	• a 'random or fixed' effect of performance on satisfaction across the average time lag used in primary studies exists • the effect of satisfaction on performance across the average time lag used in primary studies is near zero
Correlation-based MASEM with categorization	• decreasing over time • significant across short lags	• inversely u-shaped over time • significant across short and medium lags	• a 'random or fixed' effect of performance on satisfaction exists across the average time lags used in primary studies with time lags below 6 months • a 'random or fixed' effect of satisfaction on performance exists across the average time lags used in primary studies with time lags below 6 months • a 'random or fixed' effect of performance on satisfaction exists across the average time lags used in primary studies with time lags below 12 months • effects among satisfaction and performance across the average time lags used in primary studies with time lags above 12 months is near zero
Parameter-based CTSEM with random effects aggregation	• no effect	• significant	• the effect of satisfaction on performance is near zero • a random effect of performance may on satisfaction across time exists
Parameter-based homogeneity CoTiMA with multi-sample constraints	• no effect	• significant	• a fixed effect of performance on job satisfaction exists across time, but the effect of satisfaction on performance across time is near zero

Note: For further explanations for the different interpretations refer to the body of text. Further note that the word 'affect' may be more valid in terms of causality compared to cross-sectional studies, but inferring causality is not fully valid.

Table 4 summarizes the results of the CoTiMA, the previously presented correlation-based MASEM, and the random effect aggregation of the CTSEM-based single drift coefficients. It also contains language interpreting those results. Recall that general differences in interpretation of results exist. First, in the correlation-based MASEM, the random effect is lost in the SEM part of the analysis, making unclear the degree to which one should refer to random effects rather than fixed effects (cf. Cheung & Cheung, 2016). Second, interpretations differ regarding time. Whereas the CoTiMA results apply across *all* time lags, the correlation-based MASEM results without categorization do not. The differences in interpretation arise because CoTiMA has accounted for differences in time lags across primary studies without the loss of information associated with categorizing them. The correlation-based MASEM results apply to the respective 'average time lag' only. Third, whereas CoTiMA accounts for the entire structure of effects among variables by a *set* of constraints, the correlation-based MASEM and the random effects aggregation of the separately obtained CTSEM parameter estimates do not. Thus, strictly speaking, any CoTiMA effect can be interpreted without potentially violating the interpretation of another effect (e.g., X affects Y *and* Y affects X; indicated by using a common bullet point in Table 4), whereas such an interpretation has to be made with caution when random effects of CTSEM and correlation-based MASEM effects are considered (e.g., X affects Y *and/or* Y affects X; indicated by using two common bullet points in Table 4). However, neither CoTiMA nor CTSEM-based interpretations are limited to 'average time lags' as are correlation-based MASEM interpretations.

Even though the mathematical foundations of continuous time modeling are widely known and have been accepted for decades, their application to meta-analyses is less clear. For instance, one could ask if the computationally simpler category-based MASEM approach yields aggregated estimates that are less biased and more efficient than estimates delivered by CoTiMA, or if CoTiMA performs well only if the stationarity assumption is met. In order to address these and related concerns, we conducted a Monte Carlo simulation study.

Monte Carlo Simulation of Category-Based MASEM and CoTiMA

In order to compare CoTiMA with the time lag-categorization approach, we conducted a series of Monte Carlo simulations. Recall that a categorization-based MASEM yields aggregated path coefficients, whereas a CoTiMA yields aggregated drift coefficients. In order to avoid comparing apples with oranges, we transformed all subsequent CoTiMA results into path coefficients using matrix exponentiation as explained earlier. In the following we thus deal no longer with auto effects and cross effects and focus on auto-regressive effects and cross-lagged effects.

In our series of MC simulations, we varied the number of primary studies analyzed (k=9, 24, & 42) because most meta-analyses have a great deal of variability in k and because the time-lag categorization approach can produce odd results when there are small numbers of effect sizes in each category. We did not vary effect size because there is no reason why the approaches should differ depending on magnitude of effect. For the same reason, we did not examine Type I error rates.

In each cell of the simulation design, cross-lagged effects of X on Y and vice versa were estimated based on simulated data at the meta analytic level. Then their deviations from the expected parameters were examined. Each simulation was replicated 1000 times and checked against independent results obtained with 500 repetitions for cross validation purposes (e.g., Nye, Bradburn, Olenick, Bialko, & Drasgow, 2018). There were only minor differences between 500 and 1000 replications, none of which would affect conclusions, so we report results for the 1000 repetitions.

In order to select a realistic range of simulated waves, we contacted the authors of reviews and meta-analyses of stressor-strain relations in order to discover the range of time lags that they had found (Ford et al., 2014; Zapf, Dormann, & Frese, 1996). We also conducted our own search for studies in this domain. Among those studies that applied some sort of lagged analysis, the shortest lag was used by Sonnentag (2001), who used 5 days. The longest lag was used by Dalgard et al. (2009) which was 132 months. Thus, the longest lag was roughly 800 times longer than the shortest one. We decided to use a smaller ratio and to simulate a time range varying from a 1 week lag to a 52 week lag because expected cross-lagged effect sizes already started to fade out after 12 weeks, and were essentially zero after 80 weeks, which would eliminate any differences between simulated and re-analyzed effects.

After selecting simulation parameters, which we describe below, for each simulated primary study we generated 53 waves (i.e., Time 0 to Time 52, representing 52 time lags) of data using the process shown in the middle panel of Figure 1. The sample size for each simulated primary study was N = 273, which was the average sample size in the meta-analysis of Ford et al. (2014). Then we randomly generated a value i for each primary study in the range from 1 to 52, representing the lag for this study. Thus, out of the generated 53 waves only 2 were used for subsequent meta-analyses. The two variables simulated at Time 0 and Time i were used to compute the six correlations for each primary study, which served as input for reanalysis using CoTiMA and the categorization-based MASEM.

Choice of modeling parameters and estimators.

To select modeling parameters and estimators, we used the study by Demerouti, Bakker, and Bulters (2004). They investigated the reciprocal relations among work pressure and exhaustion in a 3-wave study. Time lags between measurement occasions were both 6 weeks. The correlations from Demerouti et al. are contained in online supplement S-5 as Table S-5.

We obtained discrete time parameter estimates for our simulation, with a stationary (i.e., constrained) and non-stationary (i.e., unconstrained) SEM, using lavaan (Rossel, 2012). The exact procedure and the results are presented in online supplement as Table S-6. In brief, this yielded the expected parameters (shown in Table 5 as *'M(expected)'*), which were compared with the empirical results of the Monte Carlo simulations.

Analyses.

Each CoTiMA yielded four drift coefficients, that is, two auto and two cross effects (as in the last panel of Table 3). In order to compute estimation biases for the CoTiMA, we used matrix exponentiation (see Eq. 3 & 4) to convert them into one week lagged discrete time path coefficients. Their deviations from the expected values were then used to estimate the biases and their standard deviations that occurred in each of the 1000 repetitions.

For the categorization-based MASEM it is important to note that, because time lags were sampled, the 33rd and 66th percentiles of the time lags varied across the 1000 repetitions (i.e., what was 'short', 'medium' and 'long'). Therefore, expected values had to be computed separately in each repetition. They were then used to estimate the biases (deviations) that occurred in each of the 1000 repetitions in each of the three categories. The expected values across the full range of 52 weeks are shown in Figure 4. As noted earlier, cross-lagged effects increased until the lag was 12 weeks and then leveled off, which was the reason why we decided to limit our simulation study to 52 weeks.

Results.

Results are presented in Table 5. For each cell of the simulation design, we computed the mean time lag of the primary studies generated, the expected cross-lagged effects, the mean of the empirical aggregated cross-lagged effect, the mean of the absolute difference between expected and aggregated effect and their standard deviations across repetitions, the *mean percentage bias,* which is the mean of the absolute difference between expected and aggregated effect divided by the expected effect, and its standard deviation across repetitions. A minus sign indicates the empirical effect overestimated the expected, a plus sign indicates underestimation. Bias in estimation is probably best represented by the *mean percentage bias* (cf. Aguirre-Urreta, Ellis, & Sun, 2012), which are shown in the second but last row in each panel of Table 5, and which we focus on in the next subsections.

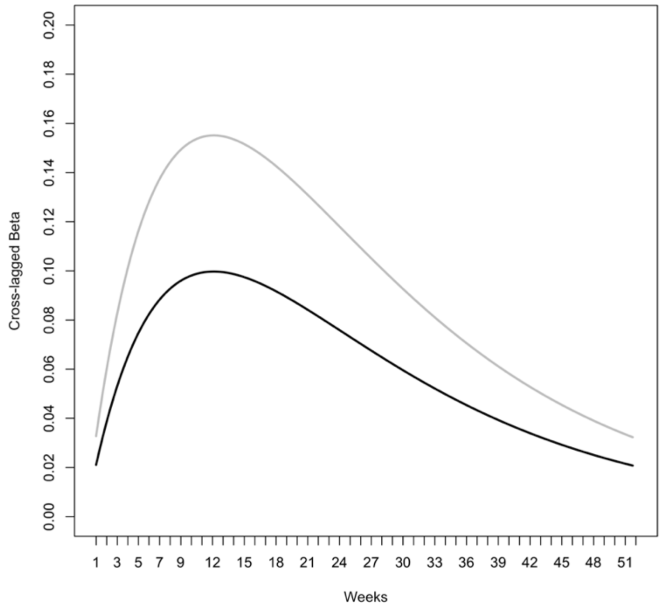

Figure 4. Expected Cross-Lagged Effects of X on Y (black) and Y on X (grey) Across 52 Simulated Lags.

Of particular note is the fact that the means of the empirical aggregated cross-lagged effects in all categorization-based MASEM decline with increasing mean time lags. If we compare this finding to Figure 4, we see that categorization-based MASEM produces an inaccurate picture of cross-lagged effects over time. This is the result of there being too many studies with relatively long time lags in the “medium” category. For example, the first panel in Table 5 shows that the average lag in the short category was close to 11 weeks, it was close to 27 weeks for the medium category, and close to 42 weeks in the long category. Figure 4 shows that expected effect sizes were larger for 11 weeks than for 27 weeks, which were larger than for 42 weeks. Thus, even though true effects sizes vary in an inversely u-shaped fashion across time lags, the results of the category-based MASEM suggest they may monotonously decrease with increasing time lags. Note that this is not an artifact of any choice that we made. Instead, it is the result of the need in categorization-based MASEM to have similar numbers of effect sizes in each category. This inaccurate picture of differences in cross-lagged effects over time is likely to be present in previous reviews and meta-analyses and has led to the conventional wisdom that *“cross-lagged effects decline as a function of time”* (Cohen et al., 2003, p. 571).

Results for stationary models show that, as expected, CoTiMAs yielded unbiased estimates. On average, all cross effects were estimated accurately within the limits of 3 digits, and bias never exceeded 1.4%. There was more bias in category-based MASEM of stationary data. For short time lags, it varied between 10.0% and 13.3%, for medium time lags between 0.7% and 1.2%, and for long time lags between 0.2% and 12.7%. Moreover, the direction of bias was consistently different for short versus long lags. For short time lags, effects were underestimated, whereas for long time lags they were overestimated. Thus, only medium time lags yield relatively unbiased results, but since researchers never know what “medium” actually is (the lags that make up the medium category depend on the lags that happen to have been represented in the primary studies), even this result is not especially encouraging. For example, we conducted an additional Monte Carlo simulation (r = 100, N = 1.000, k = 42, non-stationarity) in order to demonstrate that the smallest bias could be observed

Table 5. Results of the Monte Carlo Simulation Comparing Aggregated Cross-Lagged effects by categorization-based MASEM and by CoTiMA.

	Categorization-Based MASEM						CoTiMA	
	Short Lags		Medium Lags		Long Lags		All Lags	
Cross-lagged effect	X->Y	Y->X	X->Y	Y->X	X->Y	Y->X	X->Y	Y->X
				Stationary				
				***k* = 9**				
M(Interval)	11.009		26.759		42.441		26.508	
M(Expected)	0.088	0.137	0.069	0.107	0.034	0.053	0.023	0.036
M(Empirical)	0.079	0.122	0.068	0.106	0.038	0.053	0.023	0.036
M(Abs. Bias)	-0.009	-0.015	0.000	-0.001	0.004	0.000	0.000	0.001
SD(Abs.Bia)	0.033	0.035	0.037	0.038	0.037	0.039	0.007	0.008
M(%Bias)	-0.100	-0.107	0.007	-0.008	0.127	-0.002	-0.009	0.014
SD(% Bias)	0.380	0.251	0.605	0.394	1.191	0.829	0.316	0.217
				***k* = 24**				
M(Interval)	9.991		26.835		43.260		26.409	
M(Expected)	0.094	0.146	0.068	0.106	0.032	0.050	0.023	0.036
M(Empirical)	0.081	0.126	0.069	0.107	0.033	0.053	0.023	0.036
M(Abs. Bias)	-0.013	-0.020	0.000	0.001	0.001	0.003	0.000	0.000
SD(Abs.Bia)	0.021	0.021	0.023	0.023	0.025	0.024	0.004	0.004
M(%Bias)	-0.134	-0.136	0.008	0.012	0.025	0.044	0.006	0.002
SD(% Bias)	0.230	0.144	0.351	0.229	0.801	0.489	0.176	0.125
				***k* = 42**				
M(Interval)	9.782		26.973		43.687		26.502	
M(Expected)	0.095	0.148	0.068	0.105	0.031	0.049	0.023	0.036
M(Empirical)	0.081	0.127	0.069	0.106	0.033	0.051	0.023	0.036
M(Abs. Bias)	-0.014	-0.021	0.001	0.001	0.001	0.002	0.000	0.000
SD(Abs. Bias)	0.015	0.016	0.017	0.017	0.018	0.017	0.003	0.003
M(%Bias)	-0.143	-0.143	0.010	0.010	0.040	0.037	0.003	0.002
SD(% Bias)	0.164	0.105	0.264	0.164	0.592	0.358	0.132	0.094
				Non-Stationary				
				***k* = 9**				
M(Interval)	10.996		26.858		42.236		26.467	
M(Expected)	0.089	0.138	0.068	0.106	0.035	0.054	0.023	0.036
M(Empirical)	0.078	0.119	0.066	0.108	0.040	0.057	0.023	0.036
M(Abs. Bias)	-0.011	-0.019	-0.002	0.002	0.006	0.004	0.000	0.000
SD(Abs.Bia)	0.071	0.058	0.070	0.060	0.055	0.049	0.014	0.018
M(%Bias)	-0.130	-0.130	-0.007	0.029	0.174	0.073	0.019	0.013
SD(% Bias)	0.809	0.423	1.068	0.609	1.693	0.966	0.611	0.506
				***k* = 24**				
M(Interval)	10.193		27.195		43.497		26.666	
M(Expected)	0.094	0.146	0.067	0.105	0.032	0.050	0.023	0.036
M(Empirical)	0.074	0.124	0.066	0.105	0.040	0.057	0.023	0.035
M(Abs. Bias)	-0.02	-0.022	-0.001	0.000	0.008	0.008	0.000	-0.001
SD(Abs.Bia)	0.043	0.037	0.040	0.037	0.033	0.031	0.006	0.008
M(%Bias)	-0.214	-0.153	-0.014	0.009	0.244	0.155	0.001	-0.020
SD(% Bias)	0.456	0.254	0.608	0.360	1.044	0.634	0.282	0.236
				***k* = 42**				
M(Interval)	9.912		27.205		43.738		26.642	
M(Expected)	0.095	0.148	0.067	0.105	0.031	0.049	0.023	0.036
M(Empirical)	0.075	0.123	0.07	0.105	0.038	0.056	0.023	0.035
M(Abs. Bias)	-0.020	-0.025	0.002	0.001	0.007	0.008	0.000	-0.001
SD(Abs.Bia)	0.032	0.028	0.031	0.028	0.024	0.023	0.005	0.006
M(%Bias)	-0.215	-0.167	0.039	0.012	0.213	0.155	-0.012	-0.022
SD(% Bias)	0.343	0.188	0.461	0.268	0.768	0.483	0.206	0.166

Note: M(Interval) = mean time lag of primary studies. M(Expected) = expected cross-lagged effect. M(Empirical) = aggregated cross-lagged effect. M(Abs. Bias) = mean of the absolute difference between expected and aggregated effect. SD(Abs. Bias) = Standard deviation of M(Abs. Bias) across repetitions. *M*(% Bias) mean of the absolute difference between expected and aggregated effect divided by the expected effect (= percentage deviation from expected effect). SD(% Bias) = Standard deviation of M(% Bias) across repetitions. A minus sign indicates the empirical effect overestimated the expected, a plus sign indicates underestimation.

in the short rather than the medium category. The mean time lags were 29.23, 79.86, and 129.89, for the short, medium, and long category, respectively. In the short category, categorization-based MASEM only slightly overestimated the X->Y effects by 5.7% and only slightly underestimated the Y->X effects by -3.8%. Effects were heavily underestimated in the medium and long category (on average 137.35% and 424.55%, respectively; note that the bias using CoTiMA was 0.7% and -2.9%). Thus, because bias depends on the actual time lags that have been used in the primary studies, one could never be sure if one's category-based MASEM estimator is biased or not.

The pattern of results was very similar for the non-stationary conditions. CoTiMA still produced very low mean percentages biases. Biases were larger for category-based MASEM. The pattern of mean percentages biases for category-based MASEM was as before, with overestimation in the long category and underestimation in the short category.

As indicated per the rows showing the standard deviations of the mean percentage bias in Table 5, CoTiMA estimates were generally more efficient than were categorization based MASEM estimates.

Discussion

Panel designs allow for more valid inferences than cross-sectional designs (e.g., Kenny, 1979), and meta-analysis of correlations obtained in panel studies therefore have the potential to provide the best evidence in areas where experiments cannot be conducted. However, for a valid meta-analysis one should be able to aggregate an effect observed in Study 1 at, for example, time lag of one week and an effect observed in Study 2 with time 1 year. Hence, meta-analysis of correlations obtained in panel studies has to take the time lags applied in primary studies into account. To this end, the present study introduced parameter-based CoTiMA, which applies multi-sample analyses of continuous time SEM (Voelkle et al., 2012) using the R-package ctsem (Driver et al., 2017).

Recent advancements in continuous time modeling (e.g., Voelkle et al., 2012) provide the foundations to relate discrete time correlations to the underlying continuous time process, and multi-sample SEM with constraints across primary studies allows for meta-analysis of continuous time parameters (CoTiMA). The CoTiMA approach presented in this manuscript represents a parameter-based meta-analysis of continuous time structural equation models (cf. Cheung & Cheung, 2016).

CoTiMA solves several problems associated with more traditional approaches. First, no a prior theory is required for predicting the time lags across which effects sizes are likely to increase or decline. The continuous time math underlying CoTiMA extracts this information empirically from the effects of the primary studies and their differences in time lags. Even in the unlikely event that a researcher has chosen 'correct' categories of time lags, traditional meta-analyses suffer from two remaining problems. First, within these categories, non-linear relations still exist, which will lead to bias. Second, and much more problematic, aggregating cross-lagged effects (and cross-lagged correlations) of primary studies is usually not meaningful even if all primary studies used identical time lags. The 'pace of change' could be different, and aggregation requires taking both cross effects and both auto effects into account. Let us return to the satisfaction-performance example.

The previously mentioned problems in interpreting a meta-analysis of the satisfaction-performance relationship are solved with CoTiMA. For ease of interpretation, the summary in Table 4, first, shows that the results of the CoTiMA and the (not recommend) correlation-based discrete time MASEM without categorization of time lags (first panel of Table 4) were similar in some ways. As noted, however, interpretations differ. Even though the CoTiMA results cannot be interpreted as random effects, the interpretation generalizes to all time lags. The MASEM results may reflect a mixture of the random and fixed effects (Cheung & Cheung, 2016), but the interpretation is limited to the average time lag that happened to be used in previous studies (the sample size-weighted time lag was 10.84 years).

Second, the results of the correlation-based MASEM with categorization of different time lags differ from the CoTiMA results in important ways. Correlation-based MASEM yielded an effect of satisfaction on performance across (average) short time lags, whereas CoTiMA results imply that satisfaction does not affect performance across any time lag, except by chance. Further, correlation-based MASEM did not find an effect of performance on satisfaction across (average) long lags, whereas CoTiMA results imply that this effect exists across all time lags

Third, the interpretations of CoTiMA and the correlation-based MASEM differ. According to CoTiMA, there is an effect of performance on satisfaction and, *simultaneously,* no such effect of satisfaction on performance. Because aggregation considered the entire system including both effects among performance and satisfaction, we can conclude that these two results – one effect is present and the other is not – apply simultaneously. Therefore, we can *exclude reciprocal effects* between performance and satisfaction over time. Conversely, from the series of correlation-based MASEM, one model yielded an effect of performance on satisfaction. Another model of this series showed that no such effect of satisfaction on performance exists. Even though this is similar to as the CoTiMA result, interpretations differ. The series of correlation-based MASEM *cannot exclude reciprocal effects* between performance and satisfaction over time because the effects resulted from examining a series of different systems rather than a single one.

Taking these results together, our conclusions from the CoTiMA are (a) that job satisfaction does not cause performance and (b) that performance does affect satisfaction. Converting the CoTiMA results into discrete time SEM, we see that one *could not expect* a significant cross-lagged effect of satisfaction on performance across any time lag except by chance. In addition, one *could expect* a significant effect of performance on satisfaction if the time lag is optimally chosen and neither too short nor too long. Figure 3 suggest an optimal time lag of about 19 months (18.92 months according to Eq. 12 for continuous time coefficients in Dormann & Griffin, 2015). This seems to contradict the MASEM results at the bottom of Table 2, which shows that only cross-lagged effects for lags shorter than 12 months were significant. However, as Dormann and Griffin (2015) noted, "*many studies reported in the literature use time lags that are too long, and using these parameter estimates to compute the optimal time interval for one's own study will usually lead to further exaggerated time lags"* (p. 496). This is entirely consistent with our findings. Indeed, calculating the optimal time lag based on CoTiMA of the four studies with time lags shorter than seven months yields 7.49 months. This time lag applies to cross-lagged effects in either direction, and, following from the quote above, may still be biased upwards. Therefore, CoTiMA suggests that future studies use time lags of less than six months in order to find significant cross-lagged effects among satisfaction and performance.

The causal process generating the correlations among job satisfaction and performance across time might be more complex than assumed in the present paper. For instance, mediating variables or personality traits could be involved, and even though causal interpretations of effects could be more valid compared to cross-sectional studies they still have to be made with caution. More complex mechanisms could be modelled in CoTiMA if the required variables were reported in the primary studies. A general problem, however, is that primary studies rarely share a common set of such variables.

Having explained CoTiMA and how it is conducted, we then sought to evaluate its performance in Monte Carlo simulations. Results showed that CoTiMA estimates are relatively unbiased even if a basic assumption, stationarity, is violated. In contrast, the category-based MASEM approach produced biased estimates, with underestimation of true effects in short categories and overestimation in long categories.

In addition to this pattern of underestimation in short categories and overestimation in long categories, the results of the category-based MASEM also led to the incorrect conclusion that the cross-lagged effects decline with increasing time lags. In fact, they first increase and then decrease as per Figure 5, but this pattern is masked by the fact that many of the lags in the "medium" category are already beyond the optimal lags. The true shape would be even more hidden if the time lags in most primary studies were much too long, for example, if we had simulated data across two years instead of one year. Given the widespread belief that cross-lagged effects decline with time (Cohen et al., 2003), our conclusion is that most previous panel studies use time lags that are too long for the respective phenomena under study. This underscores the importance of the recommendation of Dormann and Griffin (2015), who called for more "shortitudinal" research. In other words, even though longitudinal studies with short lags are often criticized for failing to allow enough time to elapse, they are crucial to the meta-analytic estimation of changes in cross-lagged effects over time.

Although our results show that underestimation of true effects in short categories and overestimation in long categories occurred, this cannot be generalized to all categorization-based meta-analyses. As we noted earlier, one of the realities of meta-analyzing longitudinal data is that one must juggle the unbalanced k-problem and the problem of having an inappropriate categorization scheme because categorization-based meta-analyses must rely on the actual lags available for meta-analysis. Obviously, one cannot use a "short" category of 1 to 12 weeks if all of the primary studies used lags greater than 3 months. Therefore, in categorization-based meta-analyses researchers never know whether a time lag category is really short or long. Hence, one cannot exactly say if overestimation or underestimation is present, which in turn means that there is no magic bullet solution in categorization-based MASEM. CoTiMA circumvents this problem by utilizing information about the length of the time lags during parameter estimation. As demonstrated, this substantially reduces bias.

It should also be noted that, although CoTiMA reduces bias compared to categorization-based meta-analyses, CoTiMA results still depend on the time lags used in primary studies. If all primary studies had used time lags longer than, for example, three months, even CoTiMA would underestimate true effects. Similarly, overestimation would occur if all primary studies used lags shorter than three months.

Future Questions

Because CoTiMA is a new approach, and because continuous time math is relatively new to our field, there are a variety of questions that should be explored in the future. One has to do with data requirements. In the 2-variable 3-wave case, the number of estimated parameters in continuous time SEM is identical to a discrete SEM. In principle, in this case one could even compute the continuous time parameters from the discrete coefficients obtained by standard, that is, discrete SEM. Hence, data requirements are identical in continuous and in discrete 2-variable 3-wave models.

Where there are more than two waves, the number of coefficients required for continuous SEM is identical to or less than that required by discrete models. It could be less because of the stationarity assumption in ctsem. However, this assumption is often made in discrete SEM, for example, when in a 3-wave model, effects during the first time lag are constrained to be equal to the effects during the second (e.g., Marcelissen, Winnubst, Buunk, & de Wolff, 1988). Also note that stationarity is also assumed in other meta-analyses if one pair of time points out of three or more available time points is arbitrarily selected.

Whether in the context of CoTiMA or traditional approaches, the assumption of stationarity could be challenged. For example, causal processes may change over different developmental stages of individuals or teams. In principle, time-invariance of the drift matrix can be tested statistically by fitting a second model allowing for different drift matrices for different time periods within a study, or for different subgroups of studies that cover different developmental stages. This would leave the question of how to integrate, for example, two cross effects of X on Y, of which one describes effects at an earlier time and the other at a later time. Because one is dealing with qualitatively different stages at that point, categorizations of time periods that rest on some sort of rational basis could be reasonable for such analyses; time lags within these periods, however, could still be aggregated using CoTiMA.

Possible Extensions

There are other possible advantages of CoTiMA, which avoid limitations of extant approaches and which we mention briefly. First, although most panel studies contain only two waves, all previous longitudinal meta-analyses in the organizational literature included at least one study with three or more waves. In these cases, CoTiMA assumes stationarity and estimates a single drift matrix that operates across all time lags. Thus, CoTiMA makes use of all correlations across all waves available in primary studies. By contrast, in previous meta-analyses, researchers typically had to select two out of the many possible waves. This had to be done because (discrete time) SEM does not allow for different numbers of waves between participants or between studies. For example, Mathieu et al. (2015) selected time points that were separated by a time lag that was similar to those used in other primary studies, and Riketta (2008) always selected the first two out of three or more waves. Like CoTiMA, such procedures also rest on the stationarity assumption because arbitrarily selecting two time points is valid only if the same processes were operating across the full time range of the primary studies. The advantage of CoTiMA is that no such choice needs to be made, and all time points can be used to estimate the auto and cross effects. Differences in time lags within a study are correctly accounted for, and more information is actually used, which results in more efficient and more precise estimates.

A further advantage of simultaneously modeling more than two waves is the possibility of accounting for unobserved heterogeneity (e.g., Bollen & Brand, 2010; Hanges, & Wang, 2012), which otherwise could bias estimates. This is mathematically impossible for two-wave studies because a cross-lagged panel model is already fully saturated[3]. Without making further assumptions and adding further constraints (cf. Driver et al., 2017), a fully saturated model does not allow for estimation of additional parameters that could account for unobserved heterogeneity; three or more waves are required for such models to be mathematically identified (e.g., Dormann, 2001).

Another question to be explored has to do with multiple operationalizations of constructs. It is perfectly common in meta-analysis to use different operationalizations of key constructs for primary studies. It is possible that this would have different effects on drift coefficients than on traditional coefficients. One possibility is to treat multiple operationalizations of constructs in primary studies as multiple manifest variables. Instead of aggregating correlations within studies (e.g., Riketta, 2008), they could be used as manifest indicators of latent variables in CTSEM. On the one hand, this could provide latent factors that are free of measurement error, which, as with discrete SEM, leads to more efficient and unbiased estimates. On the other hand, using latent factors also circumvents the problem of aggregating correlations within primary studies. Of course, one could also use reliability estimates (e.g., Cronbach's alpha) of primary studies to constrain the error variance of single or multiple indicators (operationalizations) to 1-alpha. In so doing, artifactual variance that it attributable to differences in reliability is accounted for as well.

Range restriction may also cause problems that must be dealt with. For example, low performing employees may have a higher probability of getting fired, and employees suffering from high levels of strain may be more frequently on sick leave. In both cases, these employees will be underrepresented in a survey of employees. Therefore, the variance in performance or strain will be restricted which attenuates relations with other variables. Correcting for range restriction would be particularly challenging for parameter-based MASEM approaches including CoTiMA because it involves multivariate data. Hunter, Schmidt and Lee (2006, p. 594) noted "*Corrections for multivariate range restriction are also available* [...]. *However, to use this procedure one must know the intercorrelations of the independent variable measures* [in the case of CoTiMA: the t0 measures and possible third variables] *in both the restricted and unrestricted populations and must know both what tests are used in the selection composite and what the selection ratio is. This information is rarely available*

---

3 A 2-wave 2-variable study provides 10 information: 2 cross-sectional covariances, 2 cross-lagged covariances, 2 test-retest covariances, and 4 variances. In a cross-lagged panel SEM, this information is used to estimate 10 parameters: 2 variances and 1 covariance at T0, 2 autoregressive effects, 2 cross-lagged effects, and 2 error variances and 1 error covariance at T1. Thus, further effects cannot be estimated.

outside the military testing context.' Because of this, correction for range restriction is extremely complicated for longitudinal studies, and applicable procedures do not yet exist.

A random effects extension of the CoTiMA approach represents another challenge for future research. Technically, this could be accomplished by changing the OpenMx code generated by CTSEM and then making use of so-called definition variables (Mehta & Neale, 2005). However, even in the case of only two variables such as job satisfaction and performance, a model has eight parameters that may vary across primary studies (two auto effects, two cross effects, the correlation at t0, the two variances of the error variances at t1 and their covariance; only the two t0 variances are invariantly 1.0 if correlations are used). Thus, there are 36 random effects (8 variances and 28 covariances) and large samples of primary studies will be required to estimate them reliably.

Conclusion

The increasing number of panel studies and other studies using repeated measures has led to accumulating evidence on possible relations among a huge range of variables in the literature. Difficulties in aggregating this body of evidence arise because of the use of different time lags and different number of waves. CoTiMA provides an elegant method to successfully overcome these problems. Application of CoTiMA results in a more nuanced (and in most cases, accurate) understanding of longitudinal phenomena.

References

References marked with one asterisk (*) indicate studies included in the meta-analysis.

Aguirre-Urreta, M. I., Ellis, M. E., & Sun, W. (2012). Performance of a proportion-based approach to meta-analytic moderator estimation: results from Monte Carlo simulations. *Research Synthesis Methods, 3,* 11-29. doi: 10.1002/jrsm.1083

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., . . . Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, *136*, 151–173. doi:10.1037/a0018251

*Ashforth, B. E., & Saks, A. M. (1996). Socialization tactics: Longitudinal effects on newcomer adjustment. *Academy of Management Journal*, *39*, 149–178. doi:10.2307/256634

Asparouhov, T., & Muthén, B. (2012). *Multiple Group Multilevel Analysis*. Mplus Web Notes: No. 16, November 15, 2012. URL: https://www.statmodel.com/examples/webnotes/webnote16.pdf.

Atkinson, L., Niccols, A., Paglia, A., Coolbear, J., Parker, K. C. H., Poulton, L., Guger, S., & Sitarenios, G. (2000). A meta-analysis of time between maternal sensitivity and attachment assessments: Implications for internal working models in infancy/toddlerhood. *Journal of Social and Personal Relationships*, *17*, 791–810. doi:10.1177/0265407500176005

*Bechtold, S. E., Sims, H. P., & Szilagyi, A. D. (1981). Job scope relationships: A three-wave longitudinal analysis. *Journal of Occupational Behaviour*, *2*, 189–202. doi:10.1002/job.4030020304

Bollen K. A., & Brand, J. E. (2010). A general panel model with random and fixed effects: A structural equations approach. *Social Forces*, *89*, *1*–34. doi:10.1353/sof.2010.0072.

* Bond, F. W., & Bunce, D. (2003). The role of acceptance and job control in mental health, job satisfaction, and work performance. *Journal of Applied Psychology*, *88*, 1057–1067. doi:10.1037/0021-9010.88.6.1057

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

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1,* 97-111.

Cheung., M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. New York: Wiley. doi:10.1111/biom.12673

Cheung, M. W.-L., & Cheung, S.-F. (2016). Random-effects models for meta-analytic structural equation modeling: Review, issues, and illustrations. *Research Synthesis Methods*, *7*, 140–55. doi:10.1002/jrsm.1166

Cohen, A. (1993). Organizational commitment and turnover: A meta-analysis. *Academy of Management Journal*, *36*, 1140–1157. doi:10.2307/256650

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

Cole, D. A. & Maxwell, S. E. (2003). Testing meditational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*, 558–577. doi:/10.1037/0021-843X.112.4.558

Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen Jr, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology*, *96*, 443–456. doi:10.1037/a0022147.

Dalgard, O., Sørensen, T., Sandanger, I., Nygård, J. F., Svensson, E., & Reas, D. L. (2009). Job demands, job control, and mental health in an 11-year follow-up study: Normal and reversed relationships. *Work & Stress, 23,* 284–296. doi: 10.1080/07399332.2014.980889

Demerouti, E., Bakker, A. B., & Bulters, A. J. (2004). The loss spiral of work pressure, work-home interference and exhaustion: reciprocal relations in a three-wave study. *Journal of Vocational Behavior, 64*, 131–149. doi: 10.1016/S0001-8791(03)00030-7

Dormann, C. (2001). Modeling unmeasured third variables in longitudinal studies. *Structural Equation Modeling*, *8*, 575–598. doi:10.1207/S15328007SEM0804_04

Dormann, C., & Griffin, M. A. (2015). Optimal time lags in panel studies. *Psychological Methods*, *20*, 489–505. doi:10.1037/met0000041

Dormann, C. & Zapf, D. (2001). Job satisfaction – A meta-analysis of stabilities. *Journal of Organizational Behavior, 22*, 1–22. doi:10.1002/job.98

Driver, C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*, 1–35. doi:10.18637/jss.v077.i05

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-Analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170-177.

Finkel, S. E. (1995). *Causal analysis with panel data.* London: Sage.

Ford, M. T., Matthews, R. A., Woolridge, J. D., Mishra, V., Kakar, U. M., & Strahan, S. R. (2014). How do occupational stressor-strain effects vary with time? A review and meta-analysis of the relevance of time lags in longitudinal studies. *Work & Stress*, *28*, 9–30. doi:10.1080/02678373.2013.877096

Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, *132*, 778–822. doi:10.1037/0033-2909.132.5.778

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, *26*, 463–488. doi:10.1177/014920630002600305

*Griffin, R. W. (1991). Effects of work redesign on employee perceptions, attitudes, and behaviors: A long-term investigation. *Academy of Management Journal*, *34*, 425–435. doi:10.2307/256449

Hanges, P. J., & Wang, M. (2012). Seeking the holy grail in organizational science: Uncovering causality through research design. In S. W. J. Kozlowski (Ed.), Oxford library of psychology. The Oxford handbook of organizational psychology, Vol. 1, pp. 79-116). New York: Oxford University Press. doi:10.1093/oxfordhb/9780199928309.013.0003

Hedges L.V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.

Holden, G., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990). Self-efficacy of children and adolescents: A meta-analysis. *Psychological Reports*, *66*, 1044–1046. doi:10.2466/PR0.66.3.1044-1046

Hom, P. W., Caranikas-Walker, F., Prussia, G. E., & Griffeth, R. W. (1992). A meta-analytical structural equations analysis of a model of employee turnover. *Journal of Applied Psychology*, *77*, 890–909. doi:10.1037/0021-9010.77.6.890

Hulin, C. L., Henry, R.A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin*, *107*, 328–340. doi:10.1037/0033-2909.107.3.328

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

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91,* 594–612.

Jin, J., & Rounds, J. (2012). Stability and change in work values: A meta-analysis of longitudinal studies. *Journal of Vocational Behavior*, *80*, 326–339. doi: 10.1016/j.jvb.2011.10.007

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology*, *89*, 36–51. doi: 10.1037/0021-9010.89.1.36

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kessler, R. C., & Greenberg, D. F. (1981). *Linear panel analysis.* New York: Academic Press.

Ketchen Jr, D. J., Combs, J. G., Russell, C. J., Shook, C., Dean, M. A., Runge, J., ... & Beckstein, B. A. (1997). Organizational configurations and performance: A meta-analysis. *Academy of Management Journal*, *40*, 223–240. doi:10.5465/257028

Marcelissen, F. H., Winnubst, J. A., Buunk, B. & de Wolff, C. J. (1988). Social support and occupational stress: A causal analysis. *Social Science and Medicine*, *26*, 365–373. doi:10.1016/0277-9536(88)90402-9

Maricuțoiu, L. P., Sulea, C., & Iancu, A. (2017). Work engagement or burnout: Which comes first? A meta-analysis of longitudinal evidence. *Burnout Research*, *5*, 35–43. doi:10.1016/j.burn.2017.05.001

Mathieu, J. E., Kukenberger, M. R., D'Innocenzo, L., & Reilly, G. (2015). Modeling reciprocal team cohesion–performance relationships, as impacted by shared leadership and members' competence. *Journal of Applied Psychology*, *100*, 713–734. doi:10.1037/a0038898

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10,* 259-284. doi:10.1037/1082-989X.10.3.259

Mitchell, T. E., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, *26*, 530–547. doi:10.5465/AMR.2001.5393889

*Nathan, B. R., Mohrman, A. M., Jr., & Milliman, J. (1991). Interpersonal relations as a context for the effects of appraisal interviews on performance and satisfaction: A longitudinal study. *Academy of Management Journal*, *34*, 352–369. doi:10.2307/256446

Nielsen, K., Nielsen, M. B., Ogbonnaya, C., Känsälä, M., Saari, E., & Isaksson, K. (2017). Workplace resources to improve both employee well-being and performance: A systematic review and meta-analysis. *Work & Stress*, *31*, 101–120. doi:10.1080/02678373.2017.1304463

Nohe, C., Meier, L. L., Sonntag, K., & Michel, A. (2015). The chicken or the egg? A meta-analysis of panel studies of the relationship between work–family conflict and strain. *Journal of Applied Psychology*, *100*, 522–536. doi:10.1037/a0038012

Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2018). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. Organizational Research Methods, 1-32. doi: 10.1177/1094428118761122.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Riketta, M. (2008). The causal relation between job attitudes and performance: A Meta-analysis of panel studies. *Journal of Applied Psychology*, *93*, 472–481. doi:10.1037/0021-9010.93.2.472

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48 (2),* 1–36. doi: 10.18637/jss.v048.i02

Rugulies, R., Aust, B., & Madsen, I. E. H. (2017). Effort–reward imbalance at work and risk of depressive disorders. A systematic review and meta-analysis of prospective cohort studies. *Scandinavian Journal of Work, Environment & Health*, *43*, 294–306. doi:10.5271/sjweh.3632

Saridakis, G., Lai, Y., & Cooper, C. L. (2017). Exploring the relationship between HRM and firm performance: A meta-analysis of longitudinal studies. *Human Resource Management Review*, *27*, 87–96. doi: 10.1016/j.hrmr.2016.09.005

Schmidt, F. L. & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

*Sheridan, J. E., & Slocum, J. W. (1975). The direction of the causal relationship between job satisfaction and work performance. *Organizational Behavior and Human Performance*, *14*, 159–172. doi:10.1016/0030-5073(75)90021-5

Sonnentag, S. (2001). Work, recovery activities, and individual well-being: A diary study. *Journal of Occupational Health Psychology, 6,* 196-210. doi: 10.1037/1076-8998.6.3.196

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin*, *139*, 213–240. doi:10.1037/a0028931

Steel, R. P, Hendrix, W. H., & Balogh, S. P. (1990). Confounding effects of the turnover base rate on relations between time lag and turnover study outcomes: An extension of me-ta-analysis findings and conclusions. *Journal of Organizational Behavior*, *11*, 237–242. doi:10.1002/job.4030110306

Steel, R. P., & Ovalle, N. K. (1984). A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. *Journal of Applied Psychology*, *69*, 673–686. doi:10.1037/0021-9010.69.4.673

*Szilagyi, A. D. (1980). Causal inferences between leader reward behavior and subordinate performance, absenteeism, and work satisfaction. *Journal of Occupational Psychology*, *53*, 195–204. doi:10.1111/j.2044-8325.1980.tb00025.x

*Tharenou, P. (1993). A test of reciprocal causality for absenteeism. *Journal of Organizational Behavior*, *14*, 269–290. doi:10.1002/job.4030140306

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21,* 1559-1573. doi: 10.1002/sim.1187

Verkuil, B., Atasayi, S., & Molendijk, M. L. (2015). Workplace bullying and mental health: a meta-analysis on cross-sectional and longitudinal data. *PloS one*, *10*, e0135225. doi:10.1371/journal.pone.0135225

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). A SEM approach to continuous time modeling of panel data: Relating authoritarism and anomia. *Psychological Methods*, *17*, 176–192. doi:10.1037/a0027543

*Wanous, J. P. (1974). A causal– correlational analysis of the job satisfaction and performance relationship. *Journal of Applied Psychology*, *59*, 139–144. doi:10.1037/h0036532

Zapf, D., Dormann, C. & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology, 1,* 145-169. http://dx.doi.org/10.1037/1076-8998.1.2.14

Appendix A

CoTiMA of Job Satisfaction and Performance as Presented in the Present Study. This Script is also Provided as “Online Supplement S-1 CoTiMA V1.1 edit.R”. In Addition, the script “Online Supplement S-2 CoTiMA V1.1 noedit.R” is also Required and Provided as Online Supplement S-2

```
#######################################################################################
## Continuous Time Meta-Analysis (CoTiMA) - R script Version 1.1 (cf. Dormann, Guthier, & Cortina, 2019) ##
#######################################################################################

# contact: cdormann@uni-mainz.de

# This R script performs a CoTiMA based on correlation matrices of primary studies.
# Necessary Steps:
#   1. Download and install R (https://cran.r-project.org).
#      You may also want to install RStudio (https://www.rstudio.com), a desktop program facilitating using R.
#   2. Open this file in RStudio.
#   3. Change the 3 "Presets that SHOULD be adapted" in the Subsection "Presets" below:
#      3a) Enter the working directory where this file is stored. Any output will be directed to this directory, too.
#      3b) Enter a file name for the results (e.g., "results.txt")
#      3c) Enter a file "prefix" (e.g., "result") that will be used for all saved figures (e.g., "result_CrossXtoY.png")
#   4. In the Subsection "Primary Study Information",  for each primary study, enter ...
#      4a) ... the time delta_t(s) used in the primary studies (e.g., delta_t1 <- 3 or delta_t2 <- c(1, 2, 5) for ...
#            ... three different delta_ts). Note that "t1", "t2" refer to the study numbers.
#      4b) ... the sample sizes (sampleSize1 <- 668, sampleSize2 <- 22 etc.).
#      4c) ... the correlation matrix of the primary study variables X and Y at all time points (X0, Y0, X1, Y1, etc.)
#   5. Mark the entire script and select "Run"

######################################### Presets ##########################################

### Presets that SHOULD be adapted

# Working directory to save results, & prefix for figure file names. Use "/" instead of "\".
workingDirectory <- "/Users/"
resultsFileName <- "CoTiMA SatPerf.txt"
figureFilePrefix <- "CoTiMA SatPerf"

### Primary Study Information

delta_t1 <-6;
sampleSize1 <- 222
empcov1 <- matrix(c(
 1,      0.11,   0.64,   0.14,
 0.11,   1,      0.2,    0.69,
 0.64,   0.2,    1,      0.21,
 0.14,   0.69,   0.21,   1), nrow=4, ncol=4)

delta_t2 <-18
sampleSize2 <- 64
empcov2 <- matrix(c(
 1,      0.15,   0.53,   0.17,
 0.15,   1,      0.19,   0.57,
 0.53,   0.19,   1,      0.21,
 0.17,   0.57,   0.21,   1), nrow=4, ncol=4)

delta_t3 <-12
sampleSize3 <- 412
empcov3 <- matrix(c(
 1,      0.26,   0.66,   0.05,
 0.26,   1,      0.66,   0.21,
 0.66,   0.66,   1,      0.17,
 0.05,   0.21,   0.17,   1), nrow=4, ncol=4)
```

```
delta_t4 <-18
sampleSize4 <- 526
empcov4 <- matrix(c(
  1,      0.04,   0.61,   -0.02,
  0.04,   1,      0.06,   0.53,
  0.61,   0.06,   1,      0.06,
 -0.02,   0.53,   0.06,   1), nrow=4, ncol=4)

delta_t5 <-3.5
sampleSize5 <- 300
empcov5 <- matrix(c(
  1,      0.06,   0.56,   0.14,
  0.06,   1,      -0.02,  0.23,
  0.56,   -0.02,  1,      0.17,
  0.14,   0.23,   0.17,   1), nrow=4, ncol=4)

delta_t6 <- 11
sampleSize6 <- 59
empcov6 <- matrix(c(
  1,      -0.03,  0.45,   -0.08,
  -0.03,  1,      -0.06,  0.5,
  0.45,   -0.06,  1,      0.15,
  -0.08,  0.5,    0.15,   1), nrow=4, ncol=4)

delta_t7 <-12
sampleSize7 <- 35
empcov7 <- matrix(c(
  1,      0.2,    0.68,   0.21,
  0.2,    1,      0.24,   0.49,
  0.68,   0.24,   1,      0.21,
  0.21,   0.49,   0.21,   1), nrow=4, ncol=4)

delta_t8 <-3
sampleSize8 <- 128
empcov8 <- matrix(c(
  1,      0.09,   0.62,   0.09,
  0.09,   1,      0.09,   0.65,
  0.62,   0.09,   1,      0.05,
  0.09,   0.65,   0.05,   1), nrow=4, ncol=4)

delta_t9 <-12
sampleSize9 <- 200
empcov9 <- matrix(c(
  1,      0.19,   0.48,   0.11,
  0.19,   1,      0.08,   0.64,
  0.48,   0.08,   1,      0.08,
  0.11,   0.64,   0.08,   1),  nrow=4, ncol=4)

delta_t10 <-2
sampleSize10 <- 80
empcov10 <- matrix(c(
  1,      0.09,   0.73,   0.18,
  0.09,   1,      0.24,   0.44,
  0.73,   0.24,   1,      0.15,
  0.18,   0.44,   0.15,   1), nrow=4, ncol=4)
```

```
############################### Presets that COULD be adapted ##############################
######### (adjust your expectations: If all set "TRUE" this example script clould take up to hours) #########

nlatents <- 2                           # select 1 for 1 variable case, 2 for 2 variable case

retryattempts <- 50                     # number of attempts to fit a model (similar to "number of iterations",…
#                                       # ... e.g., in factor analysis )

refits <- 20                            # how many times all retryattempts should be re-done. We …
#                                       # ... recommend 10-20 if the script runs smoothly. Start with 1.

digits <- 3                             # number of digits displayed in results file

confidenceIntervals <- TRUE             # Computing confidence intervals is time consuming. We …
#                                       # ... recommend to set to TRUE after CoTiMA runs without problems.

testDRIFTallModel <- TRUE               # test a homogeneity model with all drift effects invariant across ...
#                                       # ... primary studies (= recommended CoTiMA)
testDRIFTSingleModel <- FALSE           # test a series of homogeneity models with single drift effects ...
#                                       # ... invariant (across primary studies).

testDRIFTCrossModel <- TRUE             # test a homogeneity model with both cross effects to be identical …
#                                       # … and invariant.
testDRIFTAutoModel <- FALSE             # test a homogeneity model with both auto effects to be identical …
#                                       # … and invariant.

fixedAndRandomEffects <- TRUE           # use estimated cross and auto effects of primary studies to perform …
#                                       # ... fixed and random effects analysis (Borenstein et al., 2009)

plotCrossEffects <- TRUE                # plot discrete time cross-lagged effects of CoTiMA results …
#                                       # … (recommended).
plotAutoEffects <- FALSE                # Plot discrete time autoregressive effects of CoTiMA results.

NPSOL <- TRUE                           # In fitting the model, various so-called optimizers could be applied….
#                                       #... NPSOL is such an optimizer but usually requires manual …
#                                       # … installation. If set to "TRUE" this is done (by replacing OpenMx …
#                                       # … from CRAN by OpenMx from ssri.psu.edu)

testHeterogeneityModel <- FALSE         # A heterogeneity model could be tested, but fitting may take long (!) …
#                                       # … time and results are redundant (it is identical to all studies fitted …
#                                       # … separately, which is always done); (not recommended, except …
#                                       # … for possible diagnosis of error in fitting homogeneity model).
#                                       # … Included for the sake of completeness only.

################################# Do not edit the lines below ################################
# Set the working directory & run the 2nd CoTiMA script that performs all analyses and saves results & figures
setwd(workingDirectory)
source(paste0(workingDirectory, "CoTiMA V1.1 noedit.R"))
```

Appendix B

CoTiMA Flowchart Comprising Seven Steps and the Respective Data Treatment in R, the Purposes of the Steps, and Suggestions of how to Present the Findings

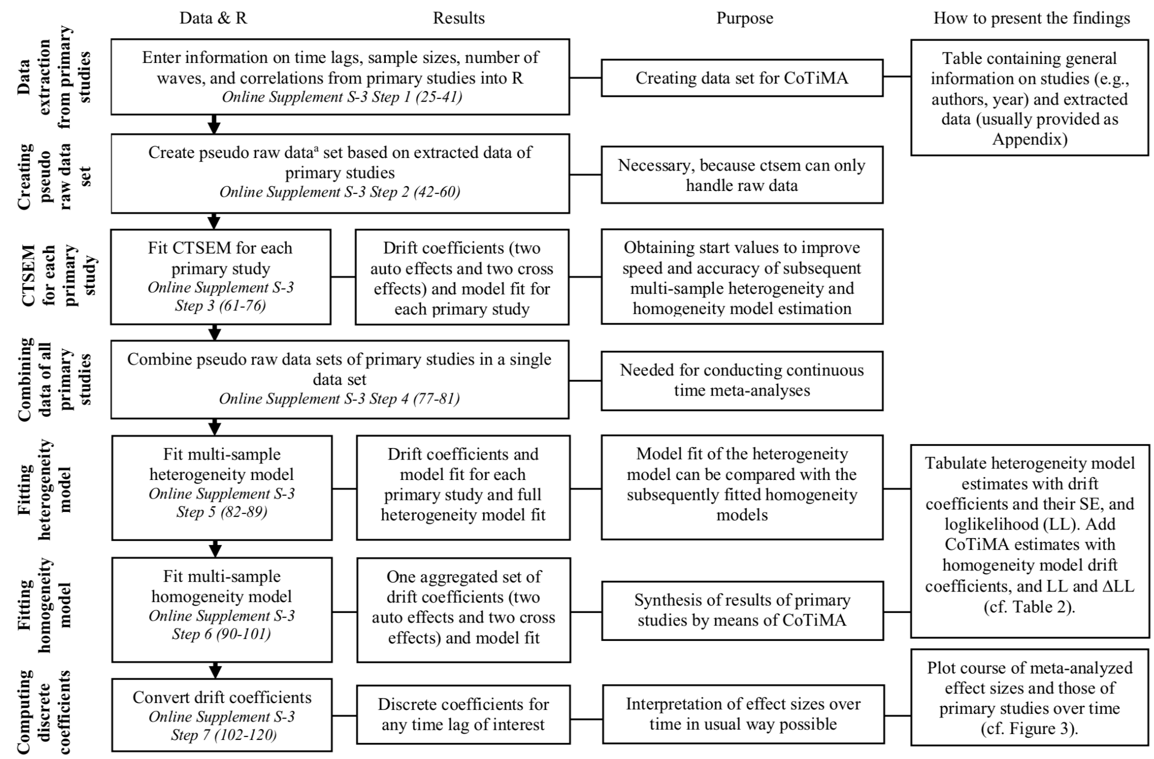

Figure B1. Flowchart for CoTiMA

[a] Required for ctsem (Driver et al., 2017). Pseudo raw data exactly re-produce the correlation matrices and their use does not affect results. With additional lines of code it is possible to use correlations instead of pseudo raw data. The code is available from the fist author upon request.