Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1394

Why Fadeout is (Probably) Worse Than We Think: Adjusting for Correlated Sampling Error in Meta-Analyses of Behavioral Interventions

Joshua B. Gilbert
Harvard University

Zachary Himmelsbach
Harvard University

The extent to which intervention effects persist or fade over time is an important question in the behavioral sciences. In meta-analysis, persistence is typically assessed by meta-regressing effect sizes at followup on effect sizes at endline. While common, the standard meta-regression does not adjust for the shared sampling error between effect sizes across time points. We show that in general, estimated slopes from the standard meta-regression are inflated under mild assumptions about correlations between outcomes across time. We show how to adjust for correlated sampling error using a sensitivity analysis approach with meta-analytic data from a series of social-emotional learning interventions. Our results suggest that effect fadeout is likely more severe than current estimates suggest.

VERSION: January 2026

Suggested citation: Gilbert, Joshua B., and Zachary Himmelsbach. (2026). Why Fadeout is (Probably) Worse Than We Think: Adjusting for Correlated Sampling Error in Meta-Analyses of Behavioral Interventions. (EdWorkingPaper: 26-1394). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/87r9-qm15

Why Fadeout is (Probably) Worse Than We Think: Adjusting for Correlated Sampling Error in Meta-Analyses of Behavioral Interventions

Joshua B. Gilbert [1] and Zachary Himmelsbach [2]

[1]Harvard Graduate School of Education
[2]Harvard Medical School

January 29, 2026

Abstract

The extent to which intervention effects persist or fade over time is an important question in the behavioral sciences. In meta-analysis, persistence is typically assessed by meta-regressing effect sizes at followup on effect sizes at endline. While common, the standard meta-regression does not adjust for the shared sampling error between effect sizes across time points. We show that in general, estimated slopes from the standard meta-regression are inflated under mild assumptions about correlations between outcomes across time. We show how to adjust for correlated sampling error using a sensitivity analysis approach with meta-analytic data from a series of social-emotional learning interventions. Our results suggest that effect fadeout is likely more severe than current estimates suggest.

Keywords: meta-analysis, meta-regression, persistence, fade out, sensitivity analysis

Corresponding author: joshua_gilbert@g.harvard.edu

Author Contributions: Conceptualization: JG; Methodology: JG, ZH; Software: JG; Formal Analysis: ZH, JG; Writing—original draft preparation: JG; Writing—review and editing: JG, ZH.

Data and Code Availability: We include the code necessary to replicate our results at the following URL: [to be added upon publication].The original data from Hart et al. (2024) are available at the following URL: https://ldbase.org/projects/39e4d12c-17f2-41f4-819b-cc9e2f154929

Ethical Considerations: Not applicable as the data used in this study are drawn from public, de-identified secondary data sources.

Consent to Participate: Not applicable.

Consent for Publication: Not applicable.

Declaration of Conflicting Interest: The authors declare no conflicts of interest.

Funding Statement: The authors received no funding in support of this work.

1 Introduction

A critical question in the social and behavioral sciences is the extent to which intervention effects persist or fade over time (Abenavoli, 2019; Bailey, 2019; Bailey et al., 2017). Meta-analyses typically examine persistence by meta-regressing effect sizes (ES) at followup on ES at endline. The slope from this meta-regression provides an interpretable summary of effect persistence, providing the expected followup ES for a given endline ES (Hart et al., 2024).

One methodological issue with this meta-regression approach is that standard meta-regression models explicitly account for sampling error in the outcome (here, ES at followup), but not in the predictor (ES at baseline). When measurement error in the predictor is independent of the outcome, adjusting for predictor measurement error disattenuates regression slopes (Kline, 2023). However, the sampling error of two ES constructed from the same outcome variables and the same participant samples at two time points is likely to be correlated. That is, if sampling error yields an estimated ES larger than the true ES at endline, a similar sampling error is likely at followup to the extent that the outcomes are correlated across time. Thus, standard measurement error corrections are likely to be counterproductive and ignoring correlated sampling error will yield bias in estimated meta-regression slopes.

The issue of correlated sampling error in meta-regression is well known in the medical literature in the analysis of surrogate outcomes, and can be accounted for by modeling the dependence in sampling errors across time points, an approach denoted "bivariate meta-regression" (Daniels & Hughes, 1997; Gail, 2000; Van Houwelingen et al., 2002). As of yet, however, models that account for correlated sampling error are rarely applied in the social and behavioral sciences when examining questions of effect persistence. In a review of the literature, Riley (2009) suggests that such models are underused due to tradition, increased complexity, the need for specialized software, and a lack of understanding of the consequences of ignoring correlated sampling error. Furthermore, the sources cited above emphasize estimating pooled effects and improving precision rather than potential bias in

meta-regression slopes that forms the basis of this study.

As such, the purpose of this study is as follows. First, we provide an accessible overview of how correlated sampling error creates bias in meta-regression slopes. Second, we describe how to correct for the bias and apply the proposed correction in a sensitivity analysis style approach. Third, we apply the correction to a large meta-analysis of persistence in social-emotional learning (SEL) interventions (Hart et al., 2024) and demonstrate that, under mild assumptions about correlations between outcomes across time, true persistence is lower than current estimates suggest. Last, we include R code for researchers interested in applying our proposed method to their own datasets.

2 Methods

2.1 The Meta-regression Model

Consider the standard meta-regression model:

$$\delta_k = \alpha + \beta X_k + u_k + e_k \tag{1}$$

$$u_k \sim \mathcal{N}(0, \tau^2) \tag{2}$$

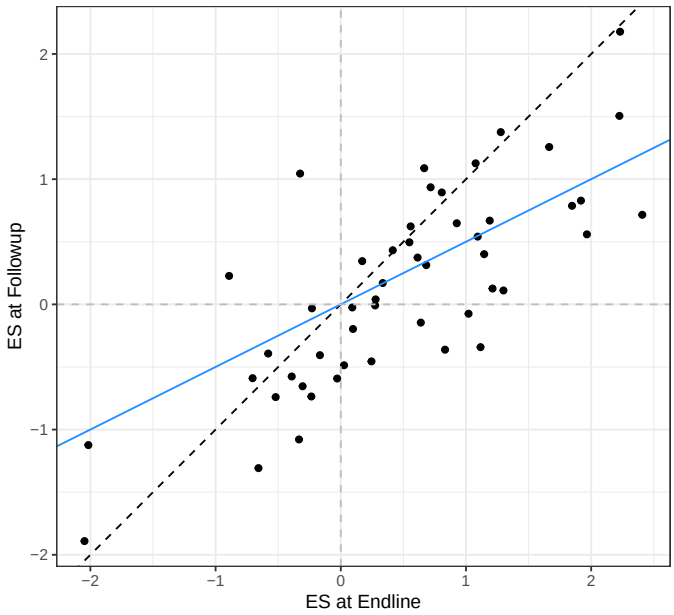
$$e_k \sim \mathcal{N}(0, \sigma^2_{e_k}), \tag{3}$$

where δ_k is the observed effect size in study k, α is the intercept, X_k is the covariate of interest, β is the meta-regression coefficient, u_k is a random effect for study with variance τ^2 and e_k is the sampling variation of the effect size, with (assumed known) variance $\sigma^2_{e_k}$. The standard meta-regression framework thus explicitly accounts for sampling variability in the outcome ES δ_k. For clarity of exposition, we consider the simple case of one ES per study (or one pair of ES when examining persistence), but note that our code and analytic models allow for multiple ES per study in an approach analogous to robust variance estimation

(RVE) (Pustejovsky & Tipton, 2022).[1]

The predictor X_k is assumed to be measured without error. Such assumptions are often reasonable, such as when X_k represents a low-inference study characteristic such as the reported sample size, whether the study is published in a peer-reviewed journal, or whether the study used an experimental or quasi-experimental design. Often however, X_k is measured with error. A common example is studies of effect persistence, where ES at followup is regressed on ES at endline. In this case, β represents the "conditional persistence" of effects, where a slope of 1 suggests that effects fully persist and a slope of 0 suggests that effects fully fade (Hart et al., 2024). We illustrate conditional persistence in Figure 1, which shows an observed conditional persistence of 50% compared to a theoretical conditional persistence of 100%.

Figure 1: Conditional Persistence of Effect Sizes Across Time



Notes: The y-axis shows the ES at followup and the x-axis shows the ES at endline. The dashed black line represents a theoretical conditional persistence of 100%. The blue line shows an observed conditional persistence of 50%.

However, when estimating meta-regressions for conditional persistence, the predictor X_k is

[1]Specifically, we assume a correlation of .80 for the study-level random effects across time in a multilevel meta-analysis framework.

itself another ES, and thus contains its own sampling error. Performing such meta-regressions without adjustments for predictor sampling error are common in meta-analysis of social and behavioral interventions (e.g., Hart et al., 2024; Mindy et al., 2024; Watts et al., 2025).

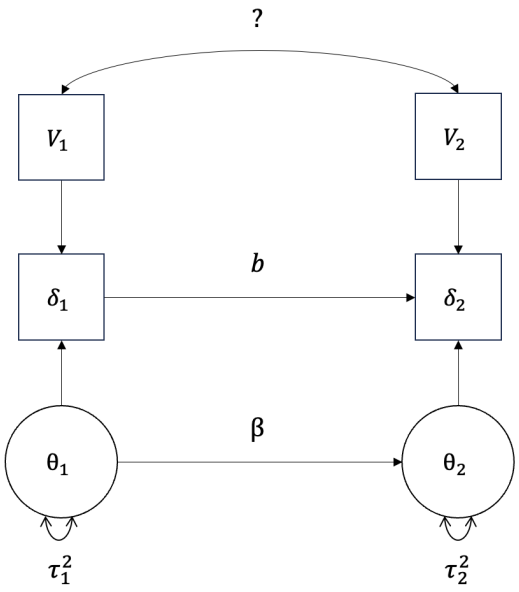
2.2 Adjusting for Correlated Sampling Error

In simple linear regression, measurement error in predictors attenuates regression coefficients towards zero, and this bias can be corrected with both errors-in-variables and structural equation modeling approaches (Gilbert, 2025; Kline, 2023; Lockwood & McCaffrey, 2020). However, standard measurement error adjustments are not applicable in the persistence meta-regression case because the measurement error in the outcome and predictor are not independent; rather, they represent some shared sampling variation. That is, if sampling error yields a point estimate larger than the true population ES for study k at endline, it is likely that this same sampling error would affect the point estimate for the ES at followup in a similar manner.

Figure 2 shows a directed acyclic graph to clarify the issue. β represents the true relationship between ES across time, but the estimated relationship b will be biased to the extent that the sampling errors are correlated. In effect, the shared sampling error functions identically to a confounder in other contexts (Pearl & Mackenzie, 2018).

However, when the correlation between outcomes across time points is known or can be estimated, we can correct for the resulting bias. We begin by deriving the sampling correlation in terms of potential outcomes and show that the sampling correlation is equal to the test-retest reliability of the outcome. We then describe how correlated sampling error creates bias in meta-regression slopes. We then use these results to propose a model that appropriately accounts for correlated sampling error (Riley, 2009).

Figure 2: Directed Acyclic Graph



Notes: Circles represent latent variables and squares represent observed variables. θ_t is the true ES, δ_t is the observed ES, and V_t is the (assumed known) sampling variance of the ES at time t. β is the true relationship between the ES across time, but we estimate b with the conventional meta-regression, potentially biased by the unknown sampling covariance across time. τ_t^2 are residual variances.

2.2.1 Derivation of the Sampling Covariance Between Effect Sizes Across Time

Consider an RCT of a binary treatment condition measuring outcomes Y_{tj} at two timepoints, $t \in \{1, 2\}$, for participants j, where Y_{1j} is the immediate endline outcome and Y_{2j} is the long term followup outcome. Under the potential outcomes framework (Rubin, 2005), let $Y(0)_{tj}$ denote the potential outcome under the control counterfactual and $Y(1)_{tj}$ denote the potential outcome under treatment. We assume that (a) the number of participants in each arm, n, is equal and (b) outcomes of different participants are independent. That is, $Y_{tj} \perp\!\!\!\perp Y_{t'k}$ for all $j \neq k$.

In the absence of treatment, we observe $Y(0)_{1j}$ and $Y(0)_{2j}$. Let $\text{Var}(Y(0)_{1j}) = \sigma_1^2$ and $\text{Var}(Y(0)_{2j}) = \sigma_2^2$. Now assume at each time an additive, constant treatment effect, θ_t, so that $Y(1)_{tj} = Y(0)_{tj} + \theta_t$. Because the treatment effect is an additive constant, the variances within the treatment condition equal those in the control condition, i.e., $\text{Var}(Y(1)_{1j}) = \sigma_1^2$

and $\text{Var}(Y(1)_{2j}) = \sigma_2^2$. (We explore the consequences of heterogeneous treatment effects in Appendix A.)

Now let $\text{Cov}(Y(0)_{1j}, Y(0)_{2k}) = \text{Cov}(Y(1)_{1j}, Y(1)_{2k}) = \rho\sigma_1\sigma_2$ if $j = k$, and 0 otherwise. Equivalently, the test-retest reliability of the outcome measure is ρ, i.e. $\text{Corr}(Y_{1j}, Y_{2j}) = \rho$, within both treatment conditions.

To estimate treatment effects in each time period, we use a difference-in-means estimator,

$$\delta_1 = \overline{Y}_{1T} - \overline{Y}_{1C}, \tag{4}$$

where $\overline{Y}_{1T}$ and $\overline{Y}_{1C}$ are, respectively, the treatment and control group sample means at endline. Similarly,

$$\delta_2 = \overline{Y}_{2T} - \overline{Y}_{2C}. \tag{5}$$

What is consequential here for the meta-regression is $\text{Corr}(\delta_1, \delta_2) = \frac{\text{Cov}(\delta_1, \delta_2)}{\sqrt{\text{Var}(\delta_1)\text{Var}(\delta_2)}}$, which we now derive.

We first decompose the estimators into the true ES θ_t and sampling error:

$$\delta_1 = \theta_1 + \{\overline{Y(0)}_{1T} - \overline{Y(0)}_{1C}\} \tag{6}$$

$$\delta_2 = \theta_2 + \{\overline{Y(0)}_{2T} - \overline{Y(0)}_{2C}\}. \tag{7}$$

Because constants do not affect variance or covariances,

$$\text{Var}(\delta_1) = \text{Var}(\overline{Y(0)}_{1T} - \overline{Y(0)}_{1C}) = \frac{2\sigma_1^2}{n} \tag{8}$$

$$\text{Var}(\delta_2) = \text{Var}(\overline{Y(0)}_{2T} - \overline{Y(0)}_{2C}) = \frac{2\sigma_2^2}{n} \tag{9}$$

$$\text{Cov}(\delta_1, \delta_2) = \text{Cov}(\overline{Y(0)}_{1T} - \overline{Y(0)}_{1C}, \overline{Y(0)}_{2T} - \overline{Y(0)}_{2C}) = \frac{2\rho\sigma_1\sigma_2}{n}. \tag{10}$$

Plugging Equation 10 into the definition of correlation, we have,

$$\text{Corr}(\delta_1, \delta_2) = \frac{\frac{2\rho\sigma_1\sigma_2}{n}}{\sqrt{\frac{2\sigma_1^2}{n}\frac{2\sigma_2^2}{n}}} = \frac{\frac{2\rho\sigma_1\sigma_2}{n}}{\frac{2\sigma_1\sigma_2}{n}} = \rho. \tag{11}$$

That is, the correlation between the ES estimates (i.e., the correlated sampling error) is equal to the test-retest reliability of the outcome.

2.2.2 Bias in Meta-regression Slopes

Now consider the consequences of Equation 11 for a "naive" conditional persistence meta-regression, where δ_{tk} indicates an observed ES at time t for study k, fit by ordinary least squares:

$$\delta_{2k} = \alpha + \beta\delta_{1k} + \varepsilon_k. \tag{12}$$

Let studies have true ES related by

$$\theta_{2k} = \alpha + \beta\theta_{1k} + \eta_k \tag{13}$$

$$\text{Cov}(\theta_{1k}, \eta_k) = 0. \tag{14}$$

The observed ES estimates δ_{tk} are functions of the true ES θ_{tk} plus sampling error e_{tk}:

$$\delta_{1k} = \theta_{1k} + e_{1k} \tag{15}$$

$$\delta_{2k} = \theta_{2k} + e_{2k}, \tag{16}$$

where e_{tk} are independent of (θ_{1k}, η_k).

The large-sample bias of the persistence slope from Equation 12 is (Pischke, 2007):

$$\text{Bias}(\widehat{\beta}) = \frac{\text{Cov}(e_1, e_2) - \beta\text{Var}(e_1)}{\text{Var}(\theta_1) + \text{Var}(e_1)}. \tag{17}$$

Plugging in Equation 10, this yields

$$\text{Bias}(\widehat{\beta}) = \frac{\rho\sqrt{\text{Var}(e_1)\text{Var}(e_2)} - \beta\text{Var}(e_1)}{\text{Var}(\theta_1) + \text{Var}(e_1)} \tag{18}$$

$$= \underbrace{\frac{\rho\sqrt{\text{Var}(e_1)\text{Var}(e_2)}}{\text{Var}(\theta_1) + \text{Var}(e_1)}}_{\text{correlation-driven bias}} - \underbrace{\frac{\beta\text{Var}(e_1)}{\text{Var}(\theta_1) + \text{Var}(e_1)}}_{\text{attenuation bias}}. \tag{19}$$

Setting $\rho = 0$, we recover the attenuation bias under classical measurement error (Gilbert, 2025; Kline, 2023). When $\rho > 0$, the bias is more positive.

In our setting, where meta-regression is used to estimate persistence, we assume that $\beta > 0$. Additionally, because within-person measures are correlated across time in behavioral settings, we assume $\rho > 0$. Typically these test-retest reliabilities are high, with one individual participant data meta-analysis of behavioral RCTs showing a mean $\rho = .51$, SD $= .23$ across 97 outcomes (Veltri & Gilbert, 2026). Attenuation bias will therefore shrink estimates toward zero, but the correlation bias will likely overcome it, leading to overestimates of persistence.

2.2.3 A Multivariate Meta-Regression Accounting for Correlated Sampling Error

We now describe a meta-regression model to account for correlated sampling error. We denote δ_{kt} and σ^2_{kt} as the ES and variance of the ES for study k at time t, respectively. We first specify the sampling covariance matrix for study k, where ρ is the correlation between outcomes at $t = 1$ and $t = 2$:

$$\Sigma_k = \begin{bmatrix} \sigma^2_{1k} & \rho\sigma_{1k}\sigma_{2k} \\ \rho\sigma_{1k}\sigma_{2k} & \sigma^2_{2k} \end{bmatrix}. \tag{20}$$

Now consider the true ES, θ_{tk}. These determine the joint distribution of observed ES δ_{tk} as follows:

$$\begin{bmatrix} \delta_{1k} \\ \delta_{2k} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \theta_{1k} \\ \theta_{2k} \end{bmatrix}, \Sigma_k \right) \tag{21}$$

We then specify a persistence model for the true ES:

$$\theta_{1k} \sim \mathcal{N}(\mu, \tau_1^2) \tag{22}$$

$$\theta_{2k}|\theta_{1k} = \mathcal{N}(\alpha + \beta\theta_{1k}, \tau_2^2), \tag{23}$$

where μ is the grand mean of true ES at time 1, α is the mean true ES at time 2 when $\theta_{1k} = 0$, β is the conditional persistence corrected for correlated sampling error at both time points, τ_1^2 is the overall variance in true ES at time 1, and τ_2^2 is the residual variance in true ES at time 2 (Riley, 2009).

Because ρ may be inconsistently reported across studies or unknown, we propose a sensitivity analysis style approach, in which we fix ρ to a range of values (e.g., 0, .1, .2, ..., .9) to determine to extent to which model results are sensitive to alternative assumptions. We use `Stan` to fit the model (Carpenter et al., 2017). We use the following priors and include example code in Appendix B:

$$\mu \sim \mathcal{N}(0, 1) \tag{24}$$

$$\alpha \sim \mathcal{N}(0, 1) \tag{25}$$

$$\beta \sim \mathcal{N}(0, 1) \tag{26}$$

$$\tau_1 \sim \text{Half-Cauchy}(0, .5) \tag{27}$$

$$\tau_2 \sim \text{Half-Cauchy}(0, .5). \tag{28}$$

2.3 Simulation Study

Given the derivation above, we conduct a targeted simulation study. The simulation study is not intended to be exhaustive, but rather to concretely demonstrate the issues at play in realistic sample sizes. We first simulate $\theta_1 \sim \mathcal{N}(.5, .25^2)$ and $\theta_2 = .5\theta_1$, that is, a conditional persistence of 50% based on endline ES mostly in the $(0, 1)$ interval. We then simulate RCTs with 100 participants randomly assigned to treatment or control conditions to represent the small to moderate sample sizes common in behavioral meta-analyses. Within each RCT, we generate the potential outcomes under control as

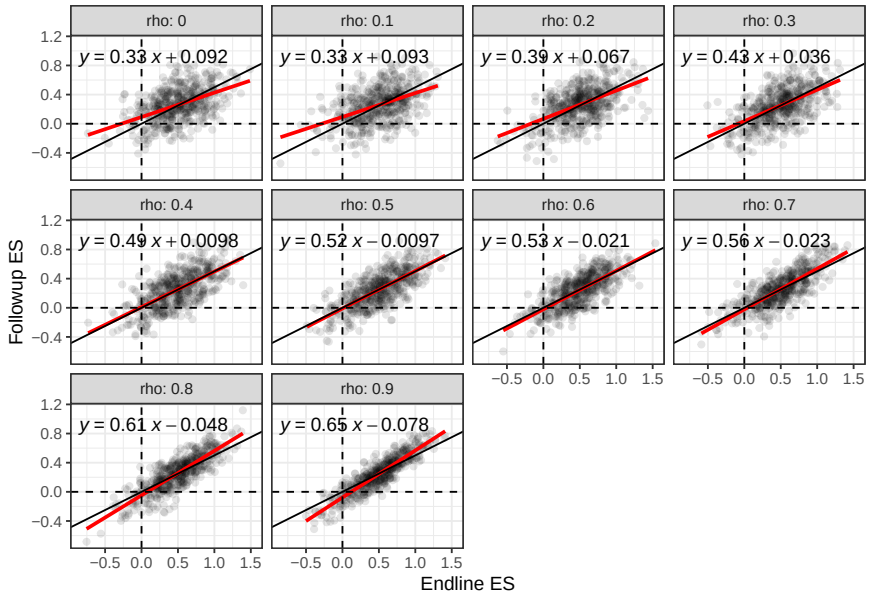
$$\begin{bmatrix} Y(0)_{1j} \\ Y(0)_{2j} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \tag{29}$$

where ρ is the test-retest reliability. We then generate the potential outcomes under treatment as $Y(1)_{1j} = Y(0)_{1j} + \theta_1$ and $Y(1)_{2j} = Y(0)_{2j} + \theta_2$. We then create the observed outcomes by selecting $Y(1)$ for units randomly assigned to treatment and $Y(0)$ for units randomly assigned to control.

We then generate 500 RCTs with true ES θ_1 and θ_2, varying ρ from 0 to .9 in increments of .1. Figure 3 shows the scatterplots of the meta-regression of observed ES (i.e., δ_2 on δ_1), faceted by ρ, with the true meta-regression in black and the OLS fit in red. In line with the results above, when $\rho = 0$, measurement error attenuates the slope, and as ρ increases, the slope inflates. A metamodel of these simulation results (Gilbert & Miratrix, 2025) confirms the pattern, as each .1 increase in ρ causes a .036 increase in the estimated conditional persistence slope (95% CI = [.031, .041]).

We next fit a model that accounts for the correlated sampling error (here, assuming ρ is known) to the data (i.e., Equation 22). Figure 4 shows the results. We find that the 95% CIs for the proposed Bayesian estimator capture the true persistence of 50% in all cases whereas the OLS estimator systematically under-estimates persistence when ρ is low and

Figure 3: Simulated Meta-regressions Across Test-Retest Reliabilities



Notes: The y-axis shows the ES at followup and the x-axis shows the ES at endline. Each dot is a simulated RCT of 100 participants randomly assigned to treatment or control conditions. The black line is the true persistence of 50% and the red line shows OLS fits. The plot is faceted by test-retest reliability ρ.

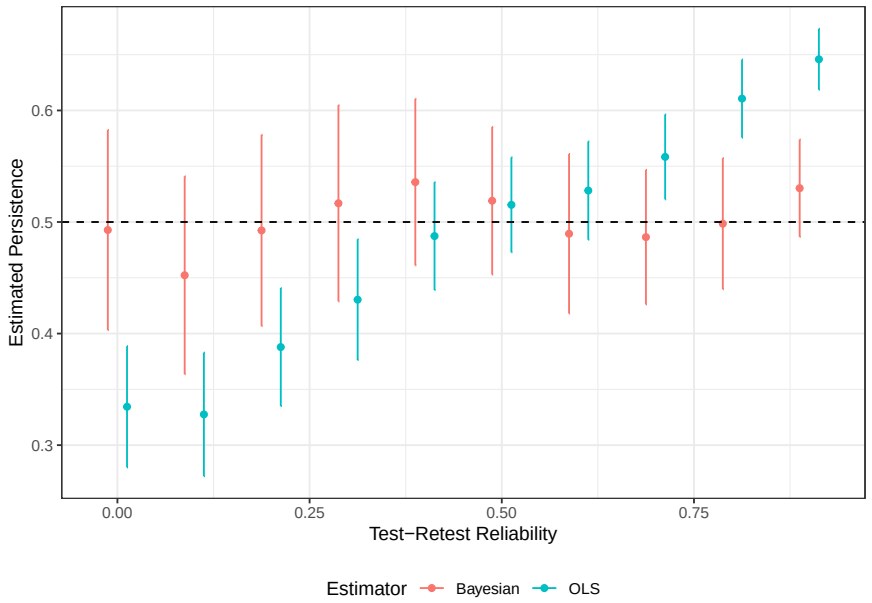
systematically over-estimates persistence when ρ is high. However, the CIs of the Bayesian estimator are much wider than the OLS estimator. Given that the true value of ρ is either unknown or inconsistently reported in primary studies, we use a sensitivity analysis approach in our empirical application to determine how the estimated conditional persistence varies across a range of plausible values for ρ.

3 Results

3.1 Data Source

We use data from Hart et al. (2024), who examine conditional persistence of social-emotional learning (SEL) interventions across a wide range of studies. In total, they include 420 ES estimates (Glass's Δ based on descriptive statistics) from 60 studies at endline and at 6-12 month followup. In the original analysis, the authors use a robust variance estimation (RVE)

Figure 4: Estimated Conditional Persistence Accounting for and Ignoring Correlated Sampling Error



Notes: The y-axis shows the estimated conditional persistence and the x-axis shows the test-retest reliability ρ. Points and lines represent point estimates and 95% CIs (credible intervals for the Bayesian estimator and confidence intervals for the OLS estimator).

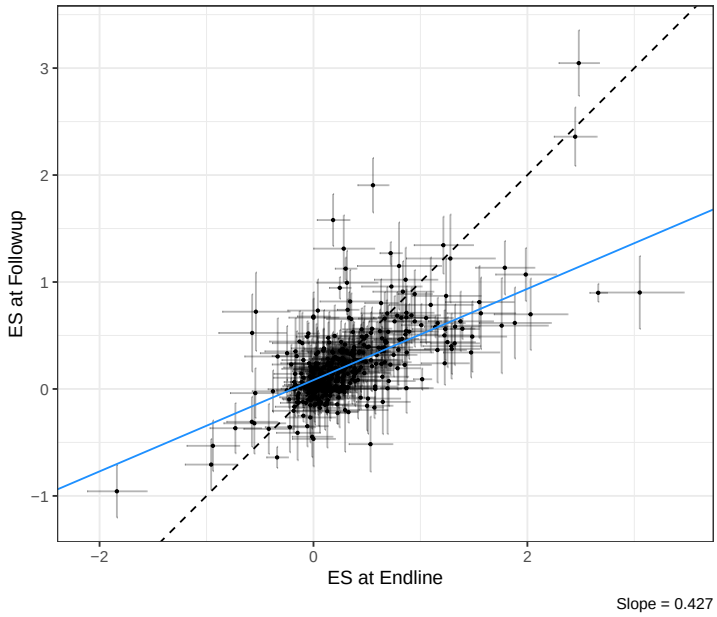
approach to account for the multiple ES within studies (Pustejovsky & Tipton, 2022), but do not account for potential correlated sampling error across time.

3.2 Standard Meta-regression Model Results

Figure 5 essentially replicates the original analysis (Hart et al., 2024, Figure 5), showing the followup ES on the y-axis and the endline ES on the x-axis.[2] We find a conditional persistence of 43% (SE = .026). Substantively this means that interventions that show an ES of 1SD at endline are predicted to show an ES of .43 at followup. The I^2 (proportion of variance in observed ES that reflects true underlying heterogeneity) is 89% at endline and 81% at followup, suggesting that the observed ES are quite reliable at both time points.

[2]To maintain comparability between our models, we use a three-level meta-regression model for our baseline analysis. The original study uses RVE.

Figure 5: Conditional Persistence based on RVE Meta-regression Model



Notes: The y-axis shows the ES at followup and the x-axis shows the ES at endline. Thin lines represent ±1 standard error from the point estimate. The dashed black line represents a theoretical conditional persistence of 100%. The blue line shows the estimated slope using the standard meta-regression approach.

3.3 Sensitivity Analysis

Table 1 shows the adjusted model results, where ρ indicates the assumed outcome correlation across time points. When $\rho = 0$, we find an estimated conditional persistence of 45%, which is slightly larger in magnitude than the initial result of 43% due to disattenuation. As ρ increases, we see the estimated conditional persistence decrease, in line with the arguments presented in Section 2. For example, assuming $\rho = .5$, we estimate 33% conditional persistence. Assuming $\rho = .9$, we estimate 20% conditional persistence. We illustrate the results graphically in Figure 6, in which each of the colored lines represents a different assumed ρ value. Thus, the initial conditional persistence estimate of 43% is likely to be too high under even mild assumptions about outcome correlation across time.

Table 1: Results of Meta-regression Models

ρ	Term	Coef. (SE)	95% CI
0.00	Intercept	0.097 (0.026)	[0.04, 0.15]
0.00	Slope	0.447 (0.03)	[0.39, 0.51]
0.10	Intercept	0.102 (0.028)	[0.05, 0.16]
0.10	Slope	0.429 (0.032)	[0.37, 0.49]
0.20	Intercept	0.107 (0.029)	[0.05, 0.16]
0.20	Slope	0.409 (0.032)	[0.35, 0.47]
0.30	Intercept	0.112 (0.029)	[0.05, 0.17]
0.30	Slope	0.384 (0.032)	[0.32, 0.45]
0.40	Intercept	0.12 (0.031)	[0.06, 0.18]
0.40	Slope	0.36 (0.032)	[0.3, 0.42]
0.50	Intercept	0.126 (0.032)	[0.06, 0.19]
0.50	Slope	0.334 (0.033)	[0.27, 0.4]
0.60	Intercept	0.131 (0.033)	[0.07, 0.2]
0.60	Slope	0.306 (0.033)	[0.24, 0.37]
0.70	Intercept	0.139 (0.035)	[0.07, 0.21]
0.70	Slope	0.273 (0.033)	[0.21, 0.34]
0.80	Intercept	0.154 (0.036)	[0.08, 0.22]
0.80	Slope	0.239 (0.033)	[0.17, 0.3]
0.90	Intercept	0.159 (0.036)	[0.09, 0.23]
0.90	Slope	0.2 (0.035)	[0.13, 0.27]

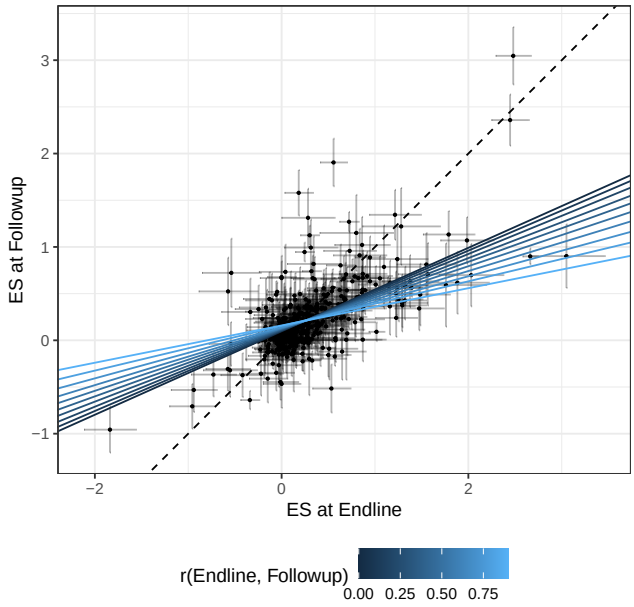
Notes: ρ indicates the assumed correlation between outcomes across time points. 95% credible intervals are calculated as the point estimate ± 1.96 times the posterior SD.

4 Discussion

Fadeout and persistence of intervention effects are critical issues in the social and behavioral sciences. The standard approach to estimating persistence is to fit a meta-regression model of followup ES on endline ES. However, this common practice ignores the correlated sampling error across time points, resulting in biased conditional persistence estimates. While models to correct for this issue have a long history in fields such as medicine (Daniels & Hughes, 1997; Riley, 2009), extant treatments emphasize precision rather than bias and applications in the social sciences are rare.

In this study, we propose a simple sensitivity check to address this issue, illustrated with empirical meta-analysis data from SEL interventions (Hart et al., 2024). We find that while the standard approach provides an estimated conditional persistence of 43%, even a mild

Figure 6: Conditional Persistence Corrected for Correlated Sampling Error



Notes: The y-axis shows the ES at followup and the x-axis shows the ES at endline. Thin lines represent ±1 standard error from the point estimate. The dashed black line represents a theoretical conditional persistence of 100%. The multicolored lines represent estimated conditional persistence assuming varying correlations between outcomes across time points.

assumption of an outcome correlation of $\rho = .5$ attenuates this estimate to 33%. Given that outcomes across time are almost certainly positively correlated, current estimates of persistence are likely too optimistic. In educational interventions, for example, correlations exceeding $\rho = .7$ are likely to obtain for standardized tests of math or reading.

We highlight two extensions to our approach. First, using covariate-adjusted ES rather than standardized mean differences such as Cohen's d may be desirable when examining persistence, because covariate adjustment may reduce the correlated sampling error and therefore lessen the bias. Second, rather than assuming a constant ρ in a sensitivity analysis, we could set a prior for the correlation itself, for example, from the Beta distribution (given that test-retest correlations are almost certain to be positive), in cases where prior evidence suggests plausible values for ρ (e.g., Veltri and Gilbert, 2026).

References

Abenavoli, R. M. (2019). The mechanisms and moderators of "fade-out": Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin*, *145*(12), 1103–1127.

Bailey, D. (2019). Explanations and implications of diminishing intervention impacts across time. In *Cognitive foundations for improving mathematical learning* (pp. 321–346). Elsevier.

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan*: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Daniels, M. J., & Hughes, M. D. (1997). Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine*, *16*(17), 1965–1982. https://doi.org/10.1002/(SICI)1097-0258(19970915)16:17⟨1965::AID-SIM630⟩3.0.CO;2-M

Gail, M. H. (2000). On meta-analytic assessment of surrogate outcomes. *Biostatistics*, *1*(3), 231–246. https://doi.org/10.1093/biostatistics/1.3.231

Gilbert, J. B. (2025). How measurement affects causal inference: Attenuation bias is (usually) more important than outcome scoring weights. *Methodology*, *21*(2), 91–122. https://doi.org/10.5964/meth.15773

Gilbert, J. B., & Miratrix, L. W. (2025). Multilevel metamodels: Enhancing inference, interpretability, and generalizability in monte carlo simulation studies. *Multivariate Behavioral Research*, 1–24. https://doi.org/10.1080/00273171.2025.2586631

Hart, E. R., Bailey, D. H., Luo, S., Sengupta, P., & Watts, T. W. (2024). Fadeout and persistence of intervention impacts on social–emotional and cognitive skills in children

and adolescents: A meta-analytic review of randomized controlled trials. *Psychological Bulletin*, *150*(10), 1207–1236. https://doi.org/10.1037/bul0000450

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Publications.

Lockwood, J. R., & McCaffrey, D. F. (2020). Recommendations about estimating errors-in-variables regression in Stata. *The Stata Journal: Promoting communications on statistics and Stata*, *20*(1), 116–130. https://doi.org/10.1177/1536867X20909692

Mindy, R., Hart, E., Bailey, D., McCormick, M., Lovett, B., & Watts, T. (2024). Using meta-analytic data to examine fadeout and persistence of intervention impacts on constrained and unconstrained skills. https://doi.org/10.26300/QN5Q-MS33

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic books.

Pischke, S. (2007). Lecture notes on measurement error. https://econ.lse.ac.uk/staff/spischke/ec524/Merr_new.pdf

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Riley, R. D. (2009). Multivariate meta-analysis: The effect of ignoring within-study correlation. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *172*(4), 789–811. https://doi.org/10.1111/j.1467-985X.2008.00593.x

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331.

Van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine*, *21*(4), 589–624. https://doi.org/10.1002/sim.1040

Veltri, G. A., & Gilbert, J. (2026). Results from randomized controlled trials are highly sensitive to data preprocessing decisions: A multiverse analysis of 97 outcomes. https://doi.org/10.31222/osf.io/kbgc2_v1

Watts, T., Hart, E., & Bailey, D. (2025). How general is educational intervention fadeout? A meta-analysis of educational RCTs with follow-up. https://doi.org/10.26300/HXHD-CV70

Appendices

A Heterogeneous Treatment Effects

Suppose that, instead of a constant additive effect at time t, θ_t, treatment effects vary also across individuals j:

$$Y(1)_{tj} = Y(0)_{tj} + \tau_{tj} \tag{30}$$

$$\theta_t = \mathbb{E}(\tau_{tj}). \tag{31}$$

Here, τ_{tj} denotes the treatment effect for individual j at time t. Continuing with the previous assumptions of equal group sizes and independence across participants, the difference-in-means estimators satisfy

$$\delta_t = \theta_t + \{\overline{Y(0)}_{tT} - \overline{Y(0)}_{tC}\} + \{\overline{\tau}_{tT} - \mathbb{E}(\tau_{tj})\}. \tag{32}$$

The covariance of the two estimators is then

$$\begin{aligned}
\mathrm{Cov}(\delta_1, \delta_2) = \frac{1}{n}\Big[&\mathrm{Cov}\big(Y(0)_{1j}, Y(0)_{2j}\big) && (33)\\
&+ \mathrm{Cov}\big(Y(0)_{1j}, \tau_{2j}\big) && (34)\\
&+ \mathrm{Cov}\big(\tau_{1j}, Y(0)_{2j}\big) && (35)\\
&+ \mathrm{Cov}\big(\tau_{1j}, \tau_{2j}\big)\Big]. && (36)
\end{aligned}$$

Consequently, the correlation between the two estimated effects becomes

$$\mathrm{Corr}(\delta_1, \delta_2) = \frac{\mathrm{Cov}\big(Y(0)_{1j}, Y(0)_{2j}\big) + \mathrm{Cov}\big(Y(0)_{1j}, \tau_{2j}\big) + \mathrm{Cov}\big(\tau_{1j}, Y(0)_{2j}\big) + \mathrm{Cov}\big(\tau_{1j}, \tau_{2j}\big)}{\frac{\sqrt{\mathrm{Var}(\delta_1)\mathrm{Var}(\delta_2)}}{n}}. \tag{37}$$

In this case, the correlation of the estimators is no longer determined solely by the test-retest reliability of the outcome ρ. It now also depends on the joint distribution of individual treatment effects across time and on their covariance with baseline potential outcomes. If individual treatment effects vary *and* those effects persist across time, so that $\text{Cov}(\tau_{1j}, \tau_{2j}) > 0$, the correlation between δ_1 and δ_2 may exceed ρ. Conversely, if treatment effects tend to reverse over time, $\text{Cov}(\tau_{1j}, \tau_{2j}) < 0$, the correlation may be attenuated or even negative.

B Example `Stan` Code

```
## ------------------------------------------------------------------------------------------
stan_code <- "
data {
  int<lower=1> K;                // Total number of ES pairs (rows)
  int<lower=1> J;                // Number of unique studies (clusters)
  array[K] int<lower=1, upper=J> study_id;

  array[K] vector[2] y;          // [es_endline, es_followup]
  vector<lower=0>[K] se1;        // SE at endline
  vector<lower=0>[K] se2;        // SE at followup

  vector<lower=-1, upper=1>[K] r; // Sampling correlation (from sensitivity analysis)
  real<lower=-1, upper=1> rho;    // RVE Correlation (study-level dependency, e.g., 0.8)
}

transformed data {
  matrix[2,2] L_S[K];
  matrix[2,2] L_Rho;

  // 1. Build the RVE Study-level Correlation Matrix (fixed rho)
  {
    matrix[2,2] Rho;
    Rho[1,1] = 1.0;
    Rho[2,2] = 1.0;
    Rho[1,2] = rho;
    Rho[2,1] = rho;
    L_Rho = cholesky_decompose(Rho);
  }

  // 2. Build the Sampling Error Cholesky factors (observation level)
  for (i in 1:K) {
    matrix[2,2] S;
    S[1,1] = square(se1[i]);
    S[2,2] = square(se2[i]);
    S[1,2] = r[i] * se1[i] * se2[i];
    S[2,1] = S[1,2];
    L_S[i] = cholesky_decompose(S);
  }
```

```
}

parameters {
  // Global Intercepts and Persistence Slope
  real mu1;        // Average endline ES
  real alpha;      // Persistence intercept
  real beta;       // Persistence slope (The corrected parameter)

  // Hierarchical Variances
  real<lower=0> sigma_u1; // Study-level SD (Endline)
  real<lower=0> sigma_u2; // Study-level SD (Followup)
  real<lower=0> tau1;     // Residual SD (Endline)
  real<lower=0> tau2;     // Residual SD (Followup)

  // Standardized random effects for non-centered parameterization
  matrix[2, J] u_raw;     // Study-level shifts
  vector[K] eps1_raw;     // Row-level residual for endline
  vector[K] eps2_raw;     // Row-level residual for followup
}

transformed parameters {
  vector[K] theta1;
  vector[K] theta2;
  matrix[2, J] u;

  // Apply the RVE correlation (rho) to the study-level random effects
  for (j in 1:J) {
    u[, j] = diag_pre_multiply([sigma_u1, sigma_u2]', L_Rho) * u_raw[, j];
  }

  for (i in 1:K) {
    // 1. Latent True ES at Endline (Study shift + residual)
    theta1[i] = mu1 + u[1, study_id[i]] + (eps1_raw[i] * tau1);

    // 2. Latent True ES at Followup (Regression on theta1 + study shift + residual)
    theta2[i] = alpha + (beta * theta1[i]) + u[2, study_id[i]] + (eps2_raw[i] * tau2);
  }
}

model {
  // Priors
  mu1 ~ normal(0, 1);
  alpha ~ normal(0, 1);
  beta ~ normal(0, 1);
  [sigma_u1, sigma_u2, tau1, tau2] ~ cauchy(0, 0.5);

  to_vector(u_raw) ~ std_normal();
  eps1_raw ~ std_normal();
  eps2_raw ~ std_normal();

  // Likelihood: Observed ES given the Latent True ES and Correlated Sampling Error
  for (i in 1:K) {
    y[i] ~ multi_normal_cholesky([theta1[i], theta2[i]]', L_S[i]);
  }
}
```

```
"


## ------------------------------------------------------------------------------------------
K <- nrow(sel)
J <- sel |> distinct(study_id) |> nrow()
study_counts <- as.numeric(table(sel$study_id))

# create a function to do this for a specified correlation
fit_mod <- function(r){

  # get the data for stan in a list format
  stan_data <- list(
      K           = K,
      J           = J,
      study_id    = as.numeric(as.factor(sel$study_id)),
      study_sizes = study_counts,
      y           = as.matrix(sel[, c("es1", "es2")]),
      se1         = sel$se1,
      se2         = sel$se2,
      # test retest
      r           = rep(r, nrow(sel)),
      # RVE rho
      rho = .8
    )

  # fit model - this takes about 90 seconds each
  fit <- stan(
    model_code = stan_code,
    data = stan_data,
    iter = 1000,
    warmup = 500,
    chains = 4,
    seed = 2026
  )

}


## ------------------------------------------------------------------------------------------
# get a sequence of correlations
rs <- seq(0, .9, .1)

# map across the inputs
if(fit_mods){

  # map across the inputs
  stan_mods <- map(rs, fit_mod, .progress = TRUE)

  # save the fitted models
  save(stan_mods, file = "data/mods/stan_mods.Rdata")

}

# load the fitted models
```

```r
load("data/mods/stan_mods.Rdata")
```