Synthesis of non-overlap of all pairs using logistic transformation or binomial generalized linear mixed model

James E. Pustejovsky & Man Chen

University of Wisconsin - Madison

April 21, 2022

The problem

- Available methods for meta-analysis of findings from single-case designs include:
  - One-stage methods involving modeling of raw data from across multiple studies.
  - Two-stage methods involving calculation of effect sizes and subsequent meta-analysis.
- The two-stage approach works well for some effect size measures, such as log response ratios, but performs inadequately for the non-overlap of all pairs (NAP) index [CP21].
- NAP quantifies effect magnitude in terms of pairwise rank comparisons of outcomes under different treatment conditions, and is thus a useful metric for outcomes that are not normally distributed and not on a ratio metric [PV09].
- We examine two alternative approaches to meta-analysis of NAP:
  - transforming the effect size estimates
  - a binomial generalized linear mixed model.

Non-overlap of all pairs

- NAP is the proportion of all possible pairs of observations from a baseline phase and an intervention phase where the outcome from the intervention constitutes a therapeutic improvement over the outcome in baseline [PV09].
- For an outcome where increase is beneficial, the NAP parameter is

$$\theta = \Pr\left(Y^B > Y^A\right) + \frac{1}{2}\Pr\left(Y^B = Y^A\right),$$

- Unbiased estimator of θ:

$$\hat{\theta} = \frac{1}{n_A n_B}\sum_{s=1}^{n_A}\sum_{t=1}^{n_B} q_{st}$$

where

$$q_{st} = \begin{cases} 1 & \text{if } y_t^B \text{ improves over } y_s^A \\ \frac{1}{2} & \text{if } y_t^B = y_s^A \\ 0 & \text{if } y_t^B \text{ deteriorates from } y_s^A \end{cases}$$

Sampling distribution of NAP

- Sampling variance of NAP is known:

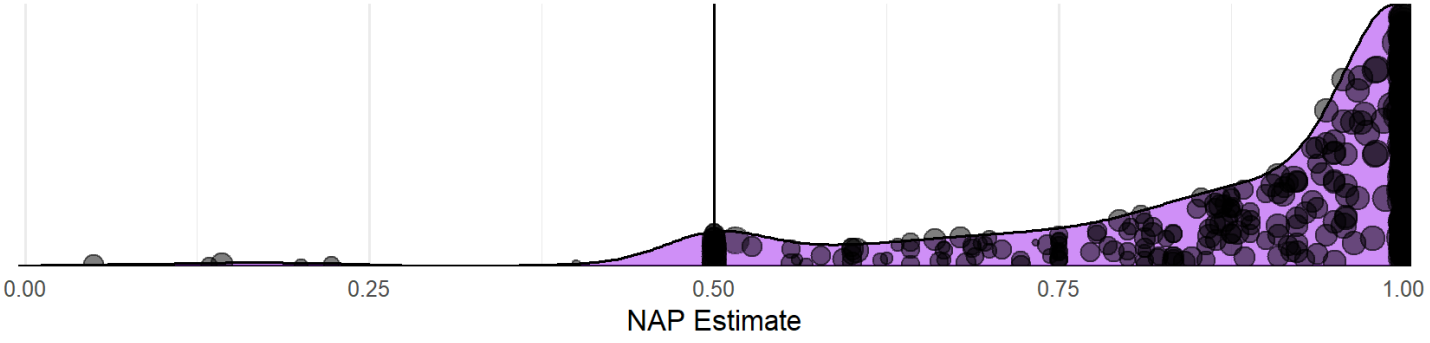
$$\mathrm{Var}\left(\hat{\theta}\right) = \frac{\theta(1-\theta)}{n_A n_B}\left[1 + (n_B - 1)\rho_1 + (n_A - 1)\rho_2\right],$$

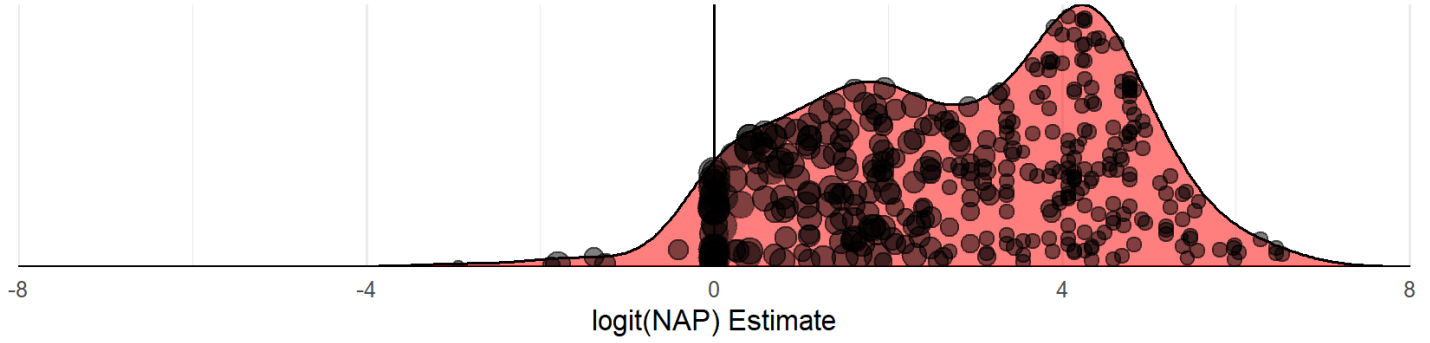
  where $\rho_1 = \frac{\mathrm{Cov}(q_{st}, q_{s't})}{\theta(1-\theta)}$ and $\rho_2 = \frac{\mathrm{Cov}(q_{st}, q_{st'})}{\theta(1-\theta)}$.

- Unbiased [Sen67; Mee90] and approximate [HM82] variance estimators have been described.

NAP estimates from AAC interventions

- [Gan+21] reported a synthesis of SCD studies on augmentative and alternative communication interventions for school-age participants with autism spectrum disorders.
- Highly skewed distribution of NAP estimates





Meta-analysis of NAP

- NAP parameter is bounded between zero and one.
- Distribution of the NAP estimator $\hat{\theta}$ is far from Gaussian and can be skewed.
- We consider multi-level meta-analytic models that describe the distribution of case-specific NAP parameters on the logistic scale:

$$\text{logit}(\theta) = \mu + u_k + v_{jk}$$

where μ is the overall average effect size (on the logistic scale), $u_k \sim N(0, \tau^2)$ is a study-level random effect, and $v_{jk} \sim N(0, \omega^2)$ is a case-level random effect.

Strategy 1: Meta-analysis of transformed NAP

- Conventional meta-analytic approach involves meta-analysis of a *transformed effect-size estimator*:

$$\text{logit}(\tilde{\theta}_{jk}) = \mu + u_k + v_{jk} + e_{jk},$$

where we assume that $\text{E}(e_{jk}) = 0$ and $\text{Var}(e_{jk}) = V_{jk} / \left[\tilde{\theta}_{jk}^2 (1 - \tilde{\theta}_{jk})^2\right]$.

  - This requires ***truncating*** the NAP estimator as

$$\tilde{\theta} = \max\left\{\frac{1}{2n_A n_B}, \min\left\{\hat{\theta}, \frac{2n_A n_B - 1}{2n_A n_B}\right\}\right\}$$

  - Transformed effect size estimator can be biased.
  - Variance estimator might not work well with small sample size.

Strategy 2: Binomial generalized linear mixed model

- An alternative is to approximate the distribution of $\hat{\theta}_{jk}$ as a binomial with probability θ_{jk} and $\tilde{N}_{jk}$ pseudo-trials,

$$\tilde{N}_{jk} = \frac{\theta_{jk}(1-\theta_{jk})}{V_{jk}}$$

- This leads to a binomial-family generalized linear mixed model with a logistic link:

$$n_{Ajk} n_{Bjk} \hat{\theta}_{jk} \sim Binom\left(\theta_{jk}, \tilde{N}_{jk}\right)$$

  where

$$\text{logit}(\theta) = \mu + u_k + v_{jk}$$

- Binomial likelihood captures the skew of the sampling distribution, avoids having to truncate $\hat{\theta}_{jk}$.

- But the pseudo-trials are *estimated* rather than known, could be estimated poorly with small sample size.

Simulation Design

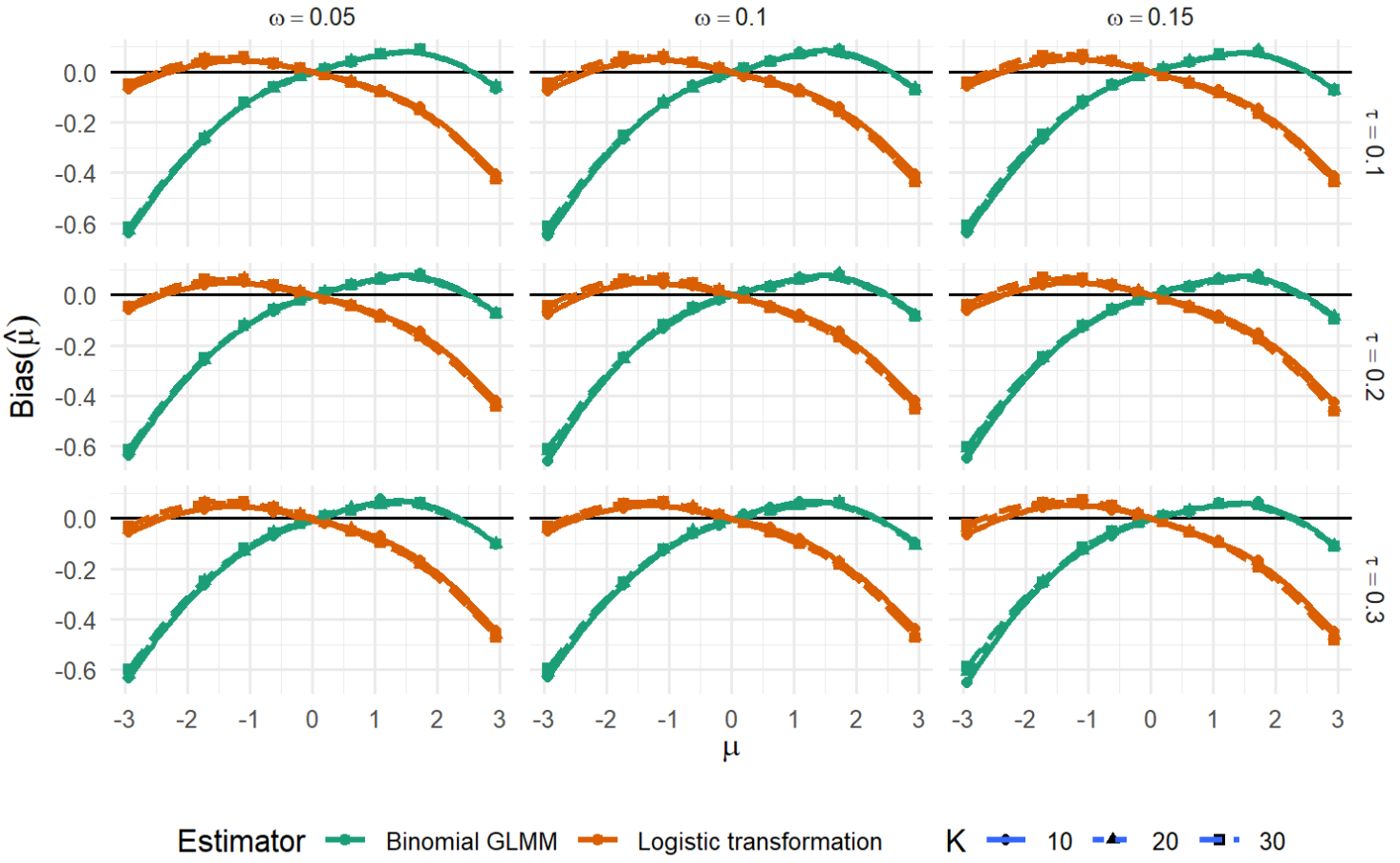
- Our simulation design generally followed [CP21].
- $K = 10, 20, 30$ primary studies.
- Average effect sizes corresponding to $\theta = .05, .15, .25, \ldots, .95$.
- Between-study heterogeneity of $\tau = 0.0, 0.1, 0.2, 0.3$.
- Within-study heterogeneity of $\omega = 0.00, 0.05, 0.10, 0.15$.
- Cases per study J_k sampled from $\{1, 2, 3, 4, 5\}$
- Phase lengths sampled from shifted Poisson distributions:

$$n_A \sim 3 + Poisson(4), \qquad n_B \sim 3 + Poisson(4)$$

- Outcome data either:
  - normally distributed with unit variance
  - Poisson-distributed

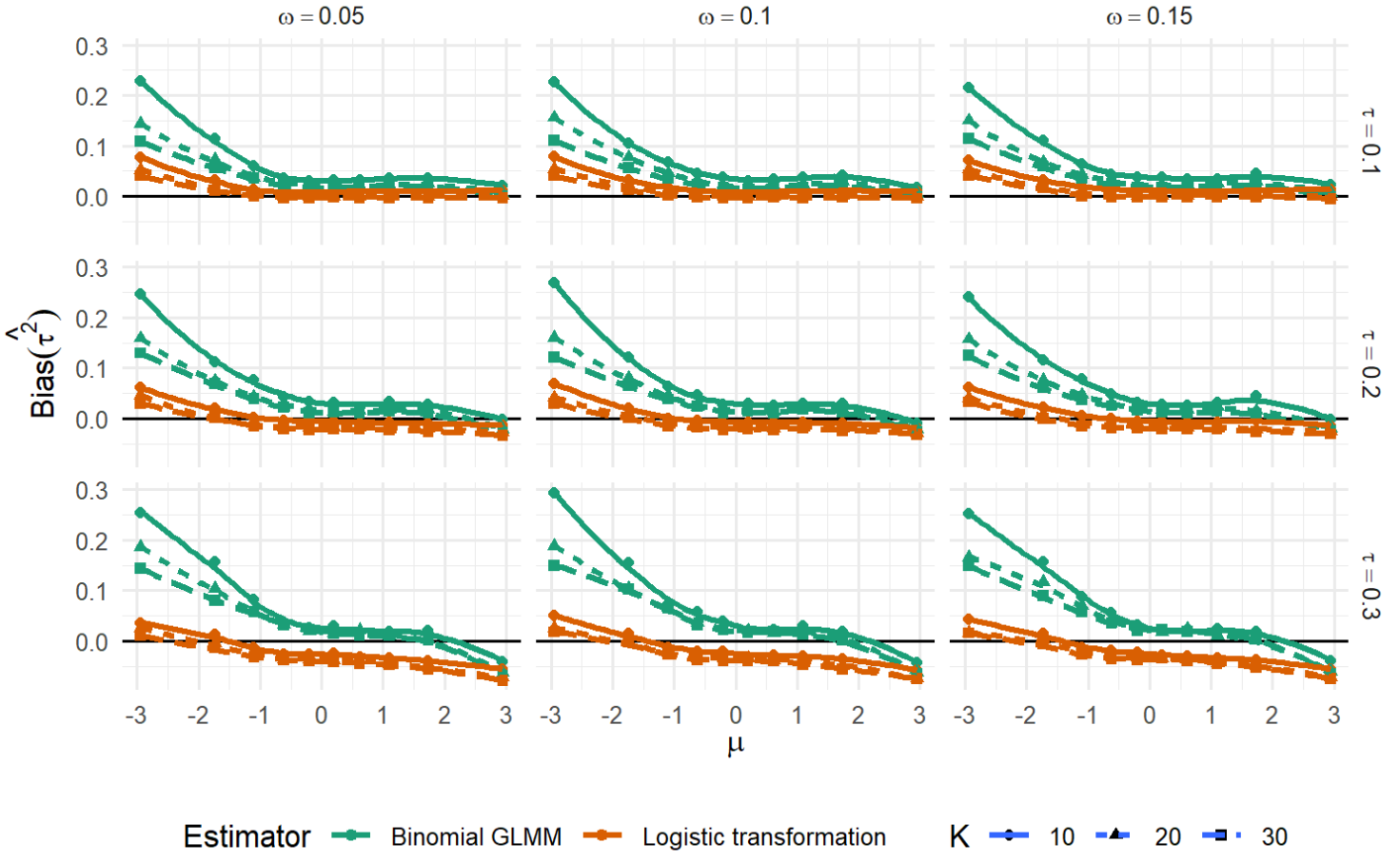
Bias of overall average effect size estimator

- Both strategies have systematically biased estimators for non-null μ.



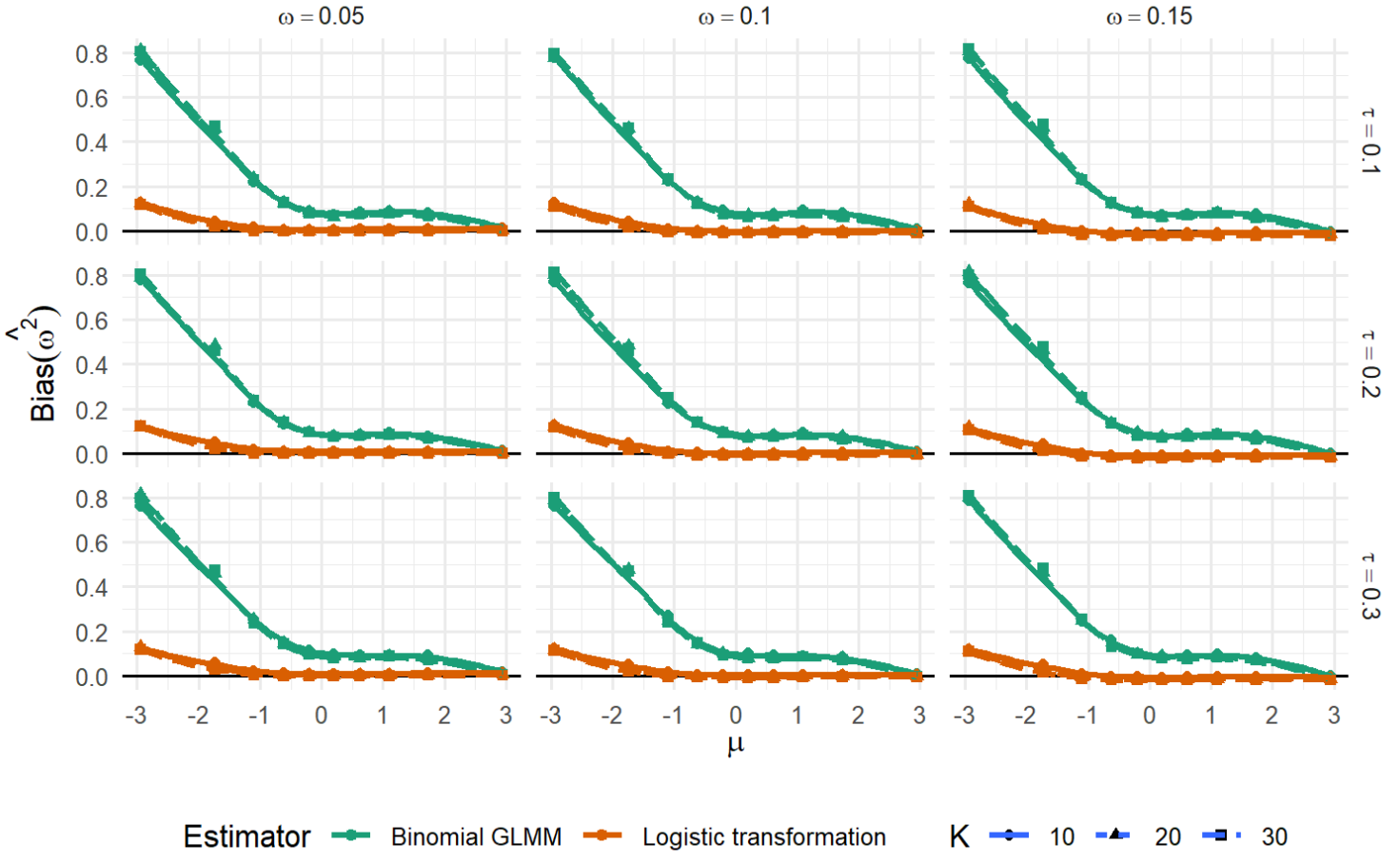
Bias of between-study heterogeneity estimator

- Both strategies have systematically biased estimators for τ^2.



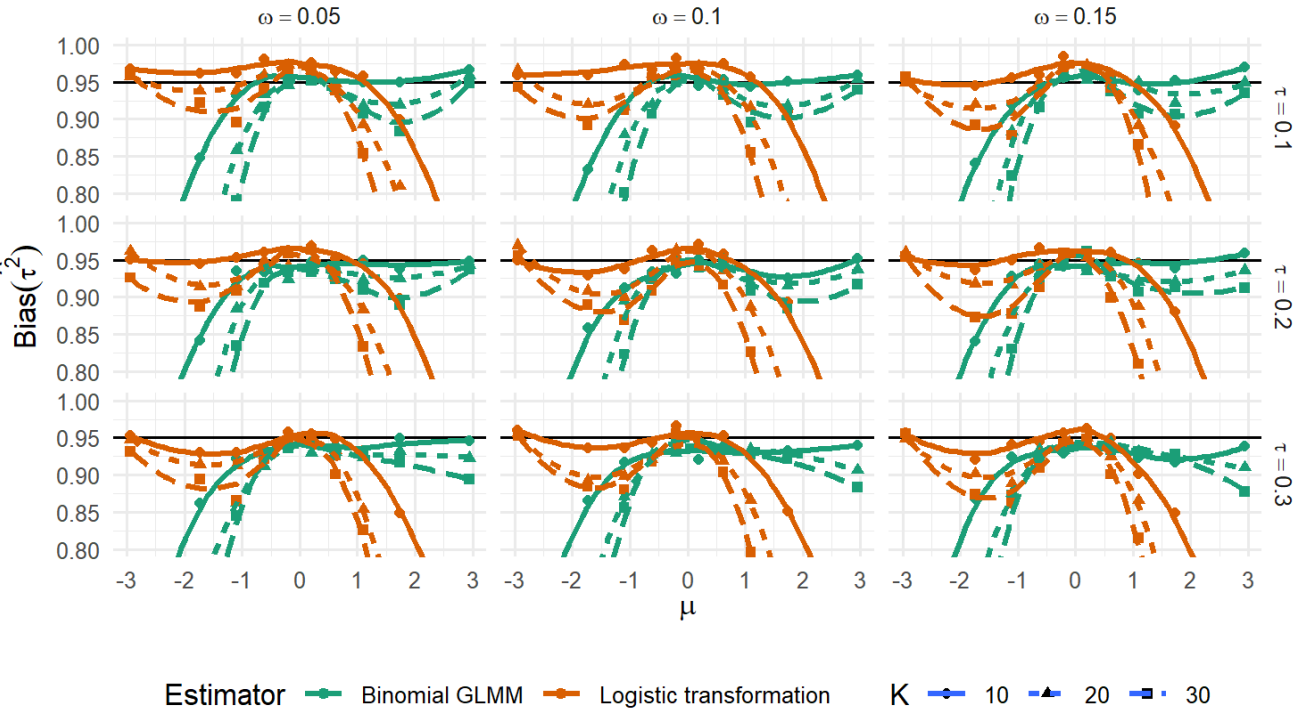
Bias of within-study heterogeneity estimator

- Both strategies have systematically biased estimators for ω^2.



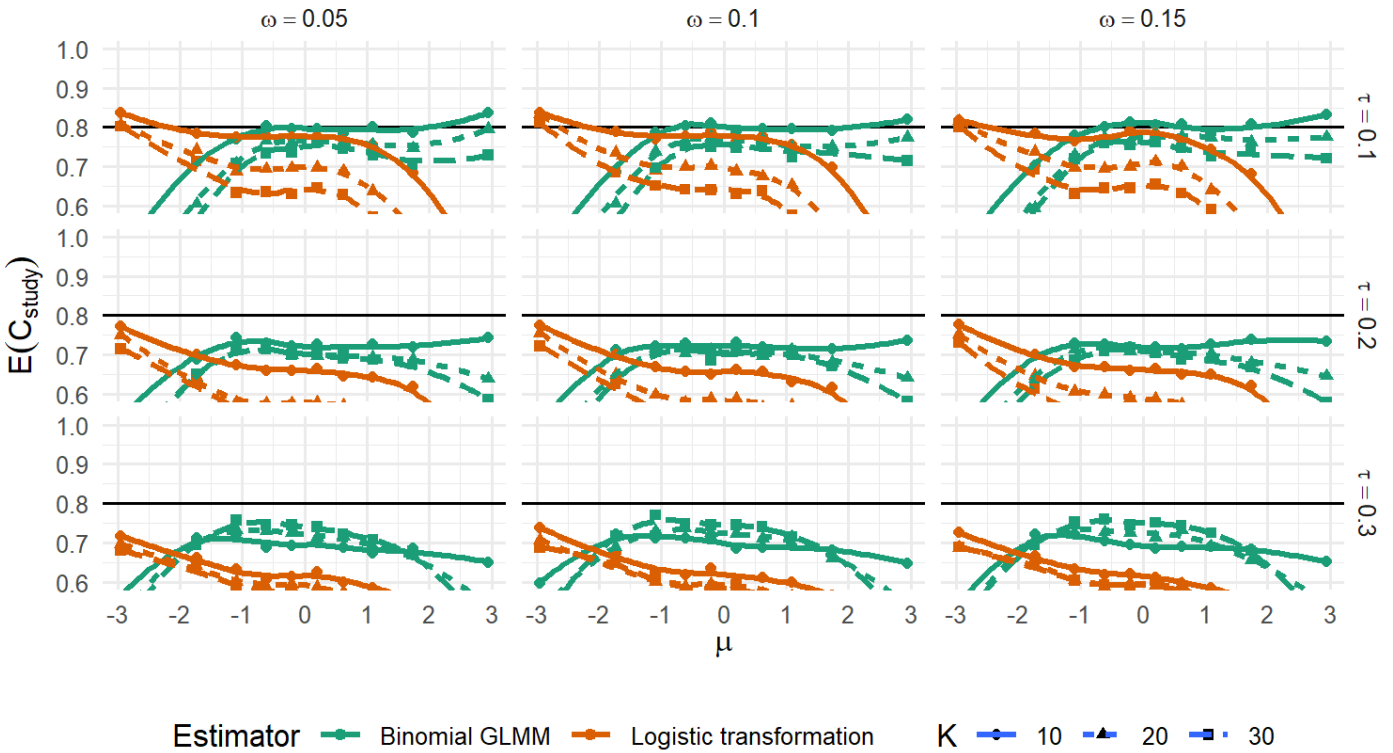
Confidence interval coverage

- Neither strategy has properly calibrated confidence intervals.
  - Mis-calibration tends to occur when $\hat{\mu}$ is more biased.
  - Coverage levels get worse with larger number of studies.



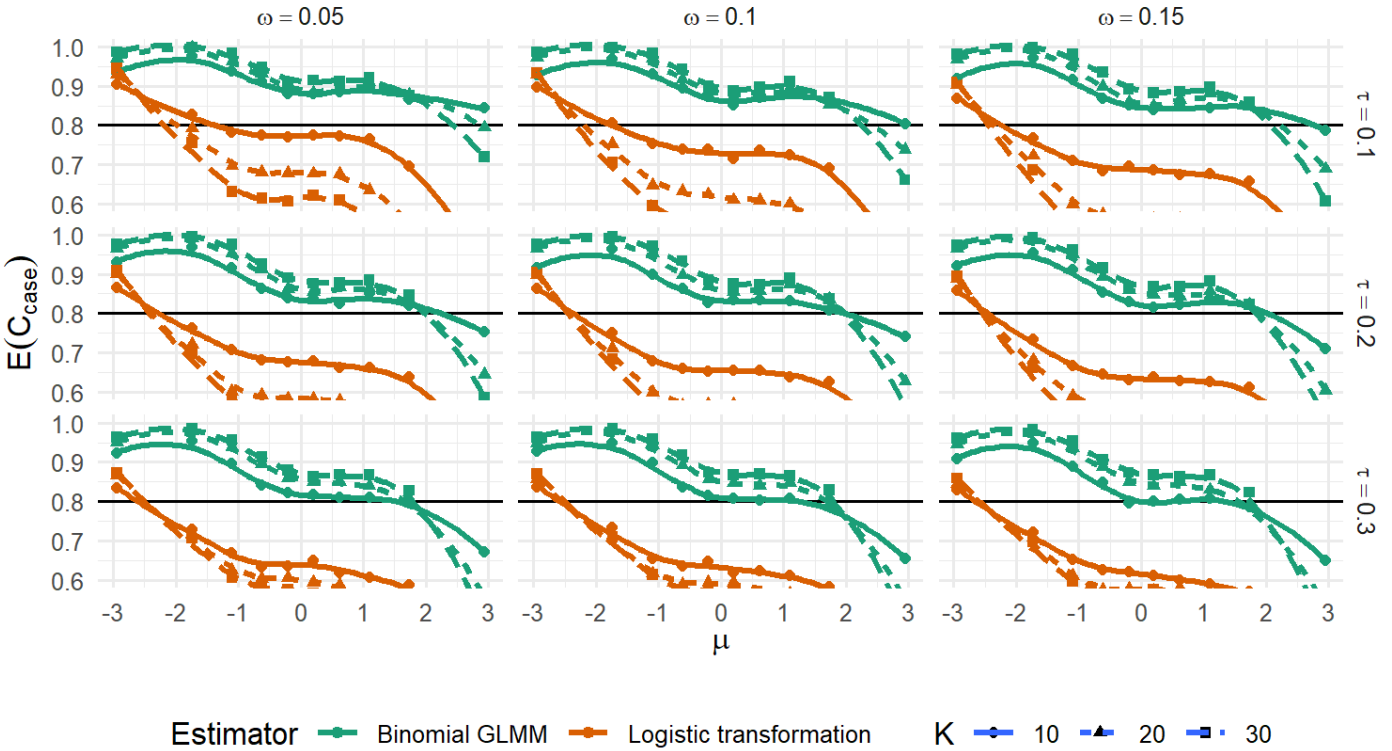
Expected coverage of study-level prediction intervals

- Neither strategy has properly calibrated prediction intervals.
  - Expected coverage levels get worse with larger number of studies.



Expected coverage of case-level prediction intervals

- Neither strategy has properly calibrated prediction intervals.
  - Binomial GLMM intervals over-cover because of overestimation of ω^2.



Discussion

- Two strategies investigated, involving different distributional approximations.
  - Logistic transformation relies on delta-method variance, approximate normality of NAP estimator, requires truncation of extreme estimates.
  - Binomial GLMM avoid truncation but requires estimation of pseudo-trials $\tilde{N}_{jk}$ for each case.
- Monte Carlo simulations under conditions typical of real single-case designs, including relatively short phase lengths, small number of cases per study.
- Neither approach worked adequately across the full parameter space.
  - Neither strategy is ready for use in practice.
- Further investigation needed.
  - Can variance estimators or pseudo-trials be stabilized by pooling across cases?
  - Correlation between effect size estimator and sampling variance estimator?

References

Chen, M. and J. E. Pustejovsky (2021). *Multi-Level Meta-Analysis of Single-Case Experimental Designs Using Robust Variance Estimation*. Preprint. PsyArXiv. DOI: 10.31234/osf.io/59h32.

Ganz, J., J. E. Pustejovsky, J. Reichle, et al. (2021). *Participant Characteristics Predicting Communication Outcomes in AAC Implementation for Individuals with ASD and IDD: A Systematic Review and Meta-analysis*. preprint. EdArXiv. DOI: 10.35542/osf.io/6sgba. URL: https://osf.io/6sgba (visited on Apr. 07, 2022).

Hanley, J. A. and B. J. McNeil (1982). "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." In: *Radiology* 143.1, pp. 29-36. DOI: 10.1148/radiology.143.1.7063747.

Mee, R. W. (1990). "Confidence intervals for probabilities and tolerance Regions based on a generalization of the Mann-Whitney statistic". In: *Journal of the American Statistical Association* 85.411, p. 793. DOI: 10.2307/2290017.

Parker, R. I. and K. Vannest (2009). "An improved Effect Size for Single-Case Research: Nonoverlap of All Pairs". In: *Behavior Therapy* 40.4, pp. 357-367. DOI: 10.1016/j.beth.2008.10.006.

Sen, P. K. (1967). "A note on asymptotically Distribution-Free Confidence Bounds for P{X $<$ Y}, Based on Two Independent Samples". In: *Sankhy\=a: The Indian Journal of Statistics, Series A (1961-2002)* 29.1, pp. 95-102.