Random-effects meta-analysis via generalized linear mixed models (GLMMs) for few studies

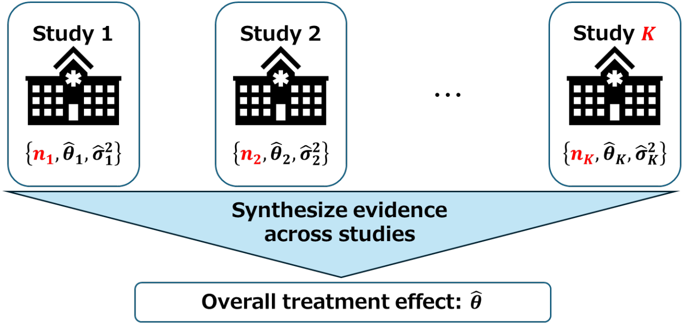
Keisuke Hanada   Tomoyuki Sugimoto

Department of Biostatistics, Faculty of Medicine, Wakayama Medical University

Graduate School of Engineering Science, The University of Osaka

The 13th Conference of the IASC-ARS

Aim of this talk



- Small-sample issues in meta-analysis:
  - Few studies (K is small)
  - GLMM meta-analysis often suffers from bias when K is small

Aim

Provide a small-sample correction for GLMM meta-analysis using only aggregate data.

Meta-analyses often include few studies

- Most meta-analyses include very few studies.
- Small-sample correction is essential in practice.

Table: Number of studies in empirical meta-analyses [Davey et al., 2011]

	All	50%	75%	90%	99%	Max
All meta-analyses	22,453	**3**	6	**10**	28	294
Binary outcomes	14,886	**3**	6	**10**	28	294
Continuous outcomes	6,672	**3**	5	**8**	24	98
Binary & continuous	895	**4**	7	**12**	46	133

- About 90% of meta-analyses include 10 studies or fewer.
  $\Rightarrow$ Inference usually occurs under small K.

Existing small-sample corrections

✓ Several small-sample corrections have been proposed:
- Profile likelihood + Bartlett correction [Noma, 2011]
- Exact confidence intervals [Michael et al., 2019]

✓ These methods improve coverage when K is small.

✗ All rely on the Normal–Normal model and its strong assumptions.

Normal–Normal model assumptions [Jackson et al., 2018]

(A1) Unbiased study-level estimates: $E[\hat{\theta}_k \mid V_k] = \theta_k$.

(A2) Known within-study variance: $\mathrm{Var}(\hat{\theta}_k \mid V_k) = \sigma_k^2$.

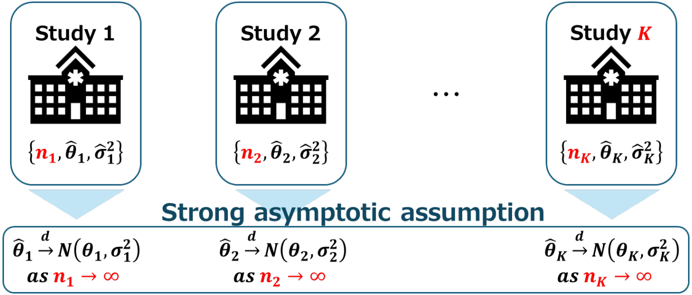
(A3) Within-study normality: $\hat{\theta}_k \mid \theta_k \sim N(\theta_k, \sigma_k^2)$.

(A4) Between-study normality: $\theta_k = \theta_0 + V_k,\ V_k \sim N(0, \tau^2)$.

Are assumptions (A1)–(A3) really reasonable?

✓ (A1)–(A3) require asymptotic normality of study estimates.

✗ Asymptotic normality requires large n_k in every study.

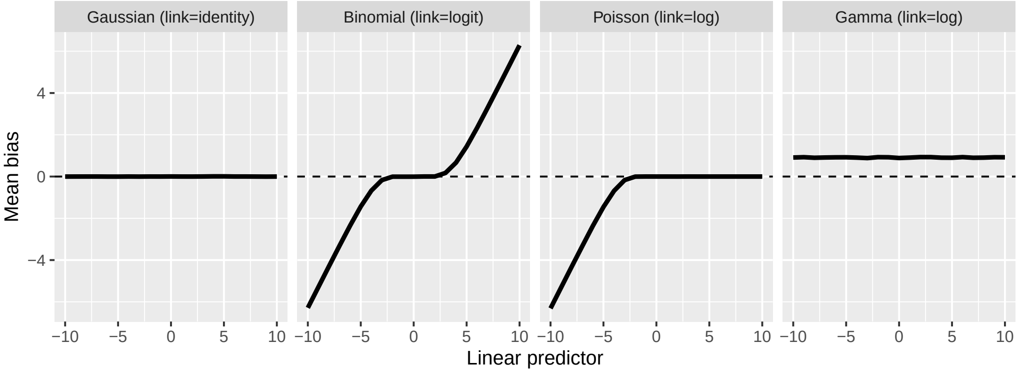


But in real meta-analyses, n_k is fixed

Small studies are often included, so (A1)–(A3) may fail.

Bias from nonlinear transformations

- ✗ Nonlinear transformations cause bias when n_k is small.
- ✗ Assumption (A1) fails: $E[\hat{\theta}_k \mid V_k] \neq \theta_k$.
- ✗ Bias remains even when $K \to \infty$.



Small-sample correction is necessary under GLMM meta-analysis.

Extension of random-effects meta-analysis to GLMM

- Individual outcomes Y_{ki} are not observed in practice in meta-analysis.
- Study k model:

$$\mu_k = E(Y_{ki} \mid V_k), \quad \theta_k = g(\mu_k) = \boldsymbol{x}_k^\top \boldsymbol{\beta} + V_k.$$

- $\boldsymbol{x}_k$ is a study-level covariate (not individual-level).

Individual-data likelihood (not available in practice)

$$L_{ki}(\boldsymbol{\beta}, \tau^2; y_{ki} \mid v_k) \propto \exp\left[\frac{y_{ki}(\boldsymbol{x}_k^\top \boldsymbol{\beta} + v_k) - b(\boldsymbol{x}_k^\top \boldsymbol{\beta} + v_k)}{a(\varphi_k)}\right]$$

- Between-study random effects:

$$V_k \overset{\text{i.i.d.}}{\sim} f_V(\cdot; \tau^2), \quad E(V_k) = 0, \; \mathrm{Var}(V_k) = \tau^2.$$

Sufficiency of aggregate data

- The study mean $\bar{y}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} y_{ki}$ is a sufficient statistic for $(\boldsymbol{\beta}, \tau^2)$.
- ✓ All information on $(\boldsymbol{\beta}, \tau^2)$ is contained in $\{n_k, \bar{y}_k, \boldsymbol{x}_k\}$.
- Individual data are unnecessary to form the likelihood.

Likelihood for study k (aggregate-data only)

$$L_k(\boldsymbol{\beta}, \tau^2) \propto \int \exp\left[\frac{n_k\{\bar{y}_k(\boldsymbol{x}_k^\top\boldsymbol{\beta} + v_k) - b(\boldsymbol{x}_k^\top\boldsymbol{\beta} + v_k)\}}{a(\varphi_k)}\right] f_V(v_k; \tau^2)\, dv_k.$$

Maximum likelihood estimator

$$(\hat{\boldsymbol{\beta}}, \hat{\tau}^2) = \underset{(\boldsymbol{\beta}, \tau^2)}{\arg\max} \log L(\boldsymbol{\beta}, \tau^2), \quad L = \prod_{k=1}^{K} L_k.$$

Confidence interval from aggregate-data likelihood

Profile likelihood ratio statistic for β_ℓ

Let β_ℓ be the parameter of interest. The profile likelihood ratio statistic is

$$T(\beta_\ell^0) = -2\{\log L(\hat{\boldsymbol{\beta}}(\beta_\ell^0), \tilde{\tau}^2(\beta_\ell^0)) - \log L(\hat{\boldsymbol{\beta}}, \hat{\tau}^2)\} \xrightarrow{d} \chi_1^2 \quad (K \to \infty)$$

where $\hat{\boldsymbol{\beta}}(\beta_\ell^0)$ and $\tilde{\tau}^2(\beta_\ell^0)$ are constrained MLEs under the constraint $\beta_\ell = \beta_\ell^0$.

- ✓ Aggregate-data likelihood allows full MLE and CI without individual data.
- ✗ The χ^2 approximation requires large K.
- ✗ Meta-analyses often have only $K \leq 10$.

  $\Rightarrow$ Small-sample correction with respect to K is essential.

Classical Bartlett correction

Bartlett correction [Lawley, 1956]

Bartlett correction modifies the test statistic T as follows:

$$T_{BC}(\beta_\ell^0) = \frac{T(\beta_\ell^0)}{1 + 2C_{BC}(\beta_\ell^0)}, \quad C_{BC} = \frac{1}{2K}\left\{ l_2^{-2}\left(\tfrac{1}{4}l_4 - l_{31} + l_{22}\right) - l_2^{-3}\left(\tfrac{5}{12}l_3^2 - 2l_3 l_{21} + 2l_{21}^2\right)\right\}$$

$$l_r = E\left[\frac{\partial^r l}{\partial \beta_\ell^r}\right], \quad l_{rs} = \frac{\partial^s l_r}{\partial \beta_\ell^s}, \quad l(\beta_\ell) = \sum_{k=1}^{K} \log L_k(\hat{\boldsymbol{\beta}}(\beta_\ell), \tilde{\tau}^2), \quad r = 1,2,3,4, \quad s = 1,2,$$

where all expectations and derivatives are evaluated under $\beta_\ell = \beta_\ell^0$.

- ✓ Improves convergence from $O(K^{-1})$ to $O(K^{-2})$.
- ✗ Requires 3rd- and 4th-order derivatives of the profile likelihood.
  $\rightarrow$ Derivatives depend on link functions, exponential-family forms, and random effects, making analytical computation very difficult.

Contribution: simplified Bartlett correction (SBC)

- Approximate $C_{BC}(\beta_\ell^0)$ by the normal–normal correction term [Noma, 2011]:

$$C_{SBC}(\beta_\ell^0) = \frac{\sum_{k=1}^{K}(\sigma_k^2+\tilde{\tau}^2)^{-3}}{\left\{\sum_{k=1}^{K}(\sigma_k^2+\tilde{\tau}^2)^{-1}\right\}\left\{\sum_{k=1}^{K}(\sigma_k^2+\tilde{\tau}^2)^{-2}\right\}} > 0.$$

- PLSBC statistic:

$$T_{SBC}(\beta_\ell^0) = \frac{T(\beta_\ell^0)}{1+2C_{SBC}(\beta_\ell^0)}.$$

Theorem (Approximation error of SBC)

Under Y_{ki} from an exponential family, $V_k \sim N(0,\tau^2)$, $n_k = na_k$, $\sum a_k = 1$, $a_k > 0$:

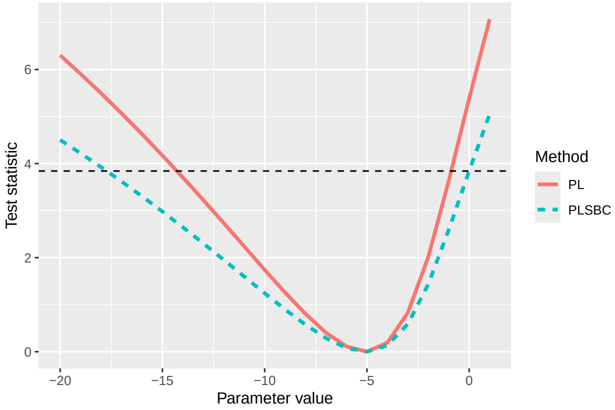
$$T_{SBC}(\beta_\ell^0) = T_{BC}(\beta_\ell^0) + O_p(n^{-1/2}) = \chi_1^2 + O_p(n^{-1/2} + K^{-2}).$$

✓ Approximation error is small: SBC retains second-order accuracy.

Real data example with 7 studies [Chu et al., 2020]

✓ SBC reduces $T(\beta_\ell)$ and widens the confidence interval.

$$\mathrm{CI}_{SBC} = \{\beta_\ell : T_{SBC}(\beta_\ell) \le q_{\chi_1^2,\alpha}\}.$$



✓ PLSBC is more conservative than PL, especially when K is small.

Simulation setup (main scenario)

- Methods compared:
  - **Normal–Normal (NN)**:
    - nDL: inverse-variance estimator [DerSimonian and Laird, 1986]
    - nPLBC: profile likelihood + Bartlett correction [Noma, 2011]
    - nMI: exact CI under NN model [Michael et al., 2019]
  - **GLMM**:
    - gPL: profile likelihood under GLMM
    - gPLSBC: PL + simplified Bartlett correction (proposed)
- Outcome types: normal, binomial, Poisson, gamma.
- Data-generating model:

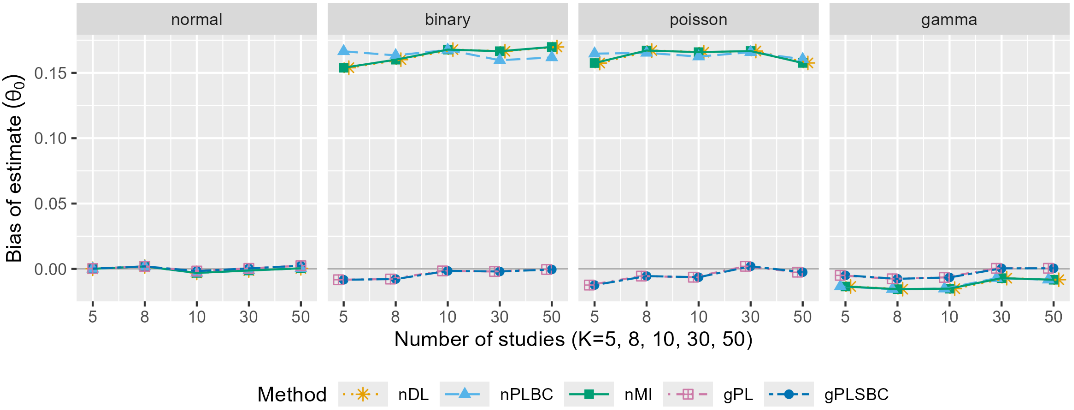
$$g(\mu_k) = \theta_0 + V_k, \quad V_k \sim N(0, \tau^2), \quad g(\cdot) : \text{link function.}$$

- Parameters:

$$\theta_0 = -2, \quad \tau^2 = 1, \quad K = 5, 8, 10, 30, 50, \quad n_k \sim \lfloor U(15, 150) \rfloor.$$

- Repetitions: 10,000 per setting.

Simulation results: mean bias of overall effect $\hat{\theta}_0$

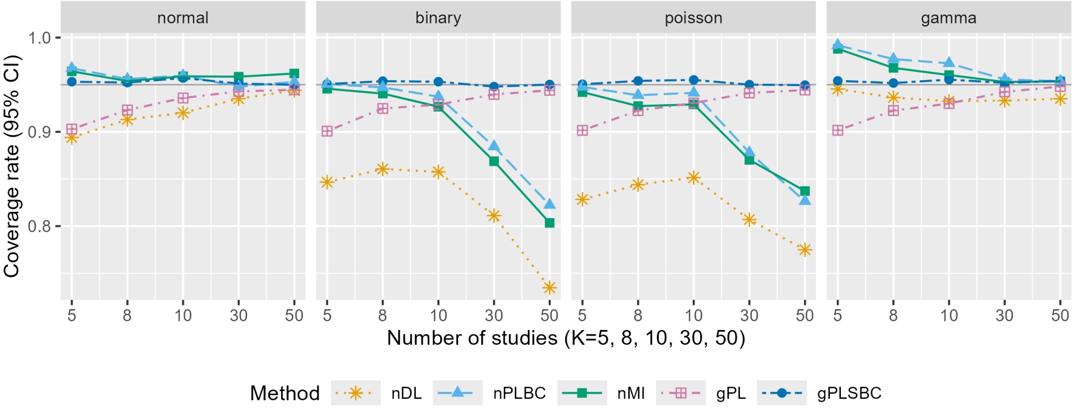


✓ GLMM methods (gPL, gPLSBC) remain consistent as $K \to \infty$.

✗ NN methods (nDL, nPLBC, nMI) show structural bias for non-normal outcomes.

$\Rightarrow$ Correct outcome modeling is essential for unbiased estimation.

Simulation results: coverage probability of nominal 95% CI



✓ gPLSBC maintains near-nominal coverage across all outcome types.

✗ NN methods (nDL, nPLBC, nMI) undercover for non-normal outcomes.

✗ gPL undercoveres when K is small.

$\Rightarrow$ Only gPLSBC is stable across both outcome types and small K.

Simulation summary

	Normal–Normal (nDL, nPLBC, nMI)	**GLMM** (gPL, gPLSBC)
Normal model correct	✓ Bias ≈ 0; coverage $\approx$ 95%	✓ Bias ≈ 0; coverage $\approx$ 95%
Non-normal outcomes (binomial, Poisson, gamma)	✗ Structural bias persists even for large K	✓ Correct model reduces structural bias
Coverage for small K (non-normal outcomes)	✗ nDL: undercoverage △ nPLBC/nMI: $\approx$ 95% when (A1)–(A3) hold; ✗ severely undercoverage when (A1) fails	✗ gPL: undercoverage ✓ gPLSBC: coverage $\approx 95\%$

Summary

- Meta-analyses often involve few studies.
- Normal–normal methods fail under non-normal outcomes.
- gPLSBC provides robust inference with near-nominal coverage even for small K.

For proofs, extended simulations, real-data analyses, and R packages, see:

arXiv:2508.08758

References I

Jonathan Davey, Rebecca M Turner, Mike J Clarke, and Julian PT Higgins. Characteristics of meta-analyses and their component studies in the cochrane database of systematic reviews: a cross-sectional, descriptive analysis. *BMC medical research methodology*, 11(1):160, 2011.

Hisashi Noma. Confidence intervals for a random-effects meta-analysis based on bartlett-type corrections. *Statistics in medicine*, 30(28):3304–3312, 2011.

Haben Michael, Suzanne Thornton, Minge Xie, and Lu Tian. Exact inference on the random-effects model for meta-analyses with few studies. *Biometrics*, 75(2):485–493, 2019.

Dan Jackson, Martin Law, Theo Stijnen, Wolfgang Viechtbauer, and Ian R White. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Statistics in medicine*, 37(7):1059–1085, 2018.

Derrick N Lawley. A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43(3/4):295–303, 1956.

Derek K Chu, Elie A Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J Schünemann, Amena El-Harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis. *The lancet*, 395(10242):1973–1987, 2020.

References II

Rebecca DerSimonian and Nan Laird. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3): 177–188, 1986.

Samuel S Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The annals of mathematical statistics*, 9(1):60–62, 1938.

Rabi N Bhattacharya, Jayanta K Ghosh, et al. On the validity of the formal edgeworth expansion. *Ann. Statist*, 6(2):434–451, 1978.

Megan Rutter, Jonathan Bowley, Peter C Lanyon, Matthew J Grainge, and Fiona A Pearce. A systematic review and meta-analysis of the incidence rate of takayasu arteritis. *Rheumatology*, 60 (11):4982–4990, 2021.

Rui Long, Junying Tian, Shasha Wu, Yang Li, Xiuhua Yang, and Jun Fei. Clinical efficacy of surgical versus conservative treatment for multiple rib fractures: a meta-analysis of randomized controlled trials. *International Journal of Surgery*, 83:79–88, 2020.