Maastricht University | Maastricht UMC+
Faculty of Health, Medicine and Life Sciences

Meta-Analyzing Stochastic Superiority: Methodological Decisions in an Applied Example

Bethany Hamilton Bhat[a] & Wolfgang Viechtbauer[a]
[a]Department of Psychiatry and Neuropsychology, Maastricht University

Background

Stochastic Superiority (θ): The probability that a randomly drawn outcome from one population exceeds a randomly drawn outcome from another, with ties counted as one half:

$$\theta = P(Y > X) + \frac{1}{2}P(Y = X)$$

Objective

In meta-analysis, raw data are not typically available. This illustration presents the methodological decisions a meta-analyst encounters when synthesizing estimates of stochastic superiority. Using individual participant data (IPD), we compare estimates for stochastic superiority and sampling variances computed from reported summary statistics against an unbiased benchmark sampling variance that uses the raw data

Methodological Considerations

Decision 1: Which estimate for θ in context of meta-analysis?

- Non-parametric $\hat{\theta}$:
  - Constructed from Mann-Whitney U, $\hat{\theta} = \mathrm{U}/(n_1 n_2)$ which counts $Y > X$ pairs or rank sums with group sizes
- Parametric $\hat{\theta}$:
  - Common Language Effect Size (CLES)[1]:
    - $\hat{\theta} = \Phi\left((\hat{\mu}_Y - \hat{\mu}_X)/\sqrt{\hat{\sigma}_Y^2 + \hat{\sigma}_X^2}\right)$

Decision 2: Which sampling variance for $\hat{\theta}$?

- We present two approaches to estimate $\mathrm{Var}(\hat{\theta})$ against an unbiased estimator as a benchmark that uses the raw data (Bamber, 1975[2]; Brunner & Konietschke, 2024[3]):

Approach	Estimator	Inputs
Parametric	Obuchowski (1997)[4]/Goddard (1990)[5]	Means, SDs, group sizes
Non-parametric	Hanley & McNeil (1982)[6,7]; Newcombe (2006)[8]; Bamber (1975) SE max[2]	$\hat{\theta}$ and group sizes

Decision 3: Transform $\hat{\theta}_i$?

- $\hat{\theta} \in [0,1]$ and the sampling distribution is skewed near 0 and 1
- Transformations for $\hat{\theta}_i$:

Transformation	$g(\hat{\theta})$	$g^{-1}(\mathrm{t})$
logit	$\ln\left(\frac{\hat{\theta}}{1-\hat{\theta}}\right)$	$\frac{1}{1+e^{-t}}$
arcsine	$\arcsin\left(\sqrt{\hat{\theta}}\right)$	$\sin^2(t)$
probit	$\Phi^{-1}(\hat{\theta})$	$\Phi(t)$

Decision 4: How to back-transform to obtain $\hat{\mu}_\theta$?

- Back-transformation approaches to estimate $\hat{\mu}_\theta$:
  - Standard back-transformation ($g^{-1}(\mathrm{t})$):
    - Returns **median** of the distribution of true effects (Jensen's inequality)
  - Integral back-transformation[9]:
    - Estimates the **mean** of the distribution of true effects

$$\mu_\theta = \int_{-\infty}^{\infty} g^{-1}(t)\ \phi\left(t \mid \mu_g, \tau_g^2\right) dt$$

Applied Example: Huber et al., 2000

- IPD from five RCTs of Uro-Vaxom vs. placebo[10]; outcome is the number of urinary tract infection recurrences
- Stochastic superiority (θ): probability that a randomly chosen placebo subject has more recurrences than a randomly chosen Uro-Vaxom subject, with ties counted as one half

Results

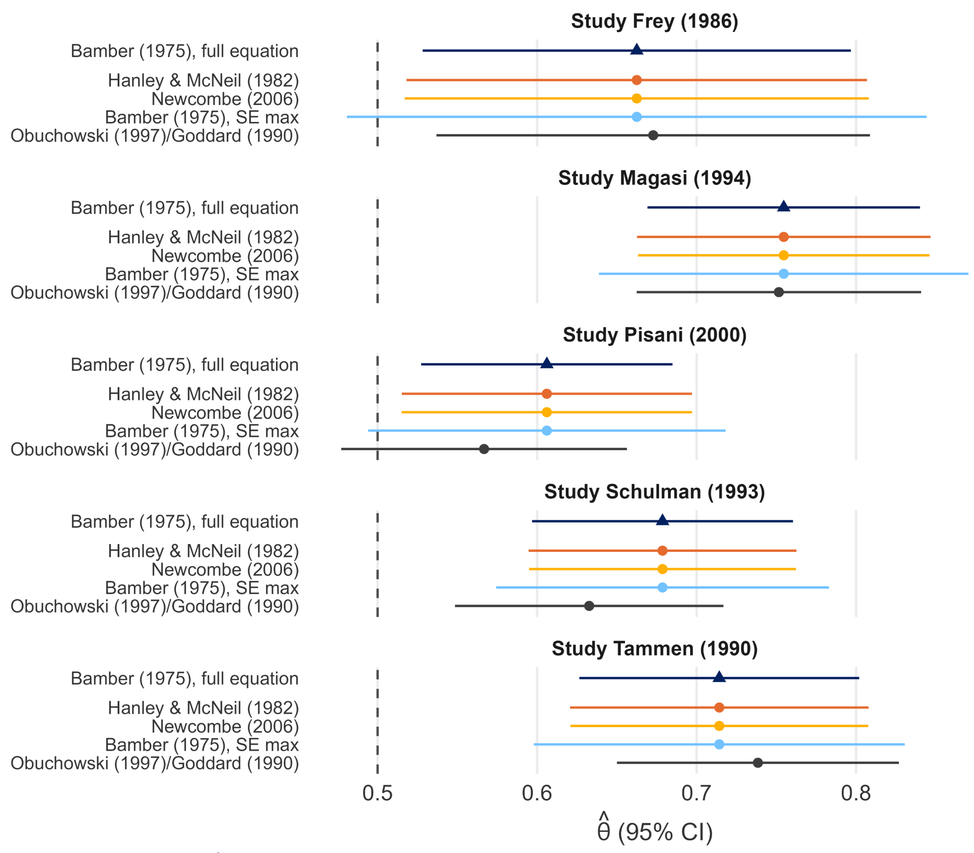


Figure 1: Study-level $\hat{\theta}_i$ estimates and 95% CI from Huber et al. (2000) by sampling-variance method. Triangle = unbiased benchmark

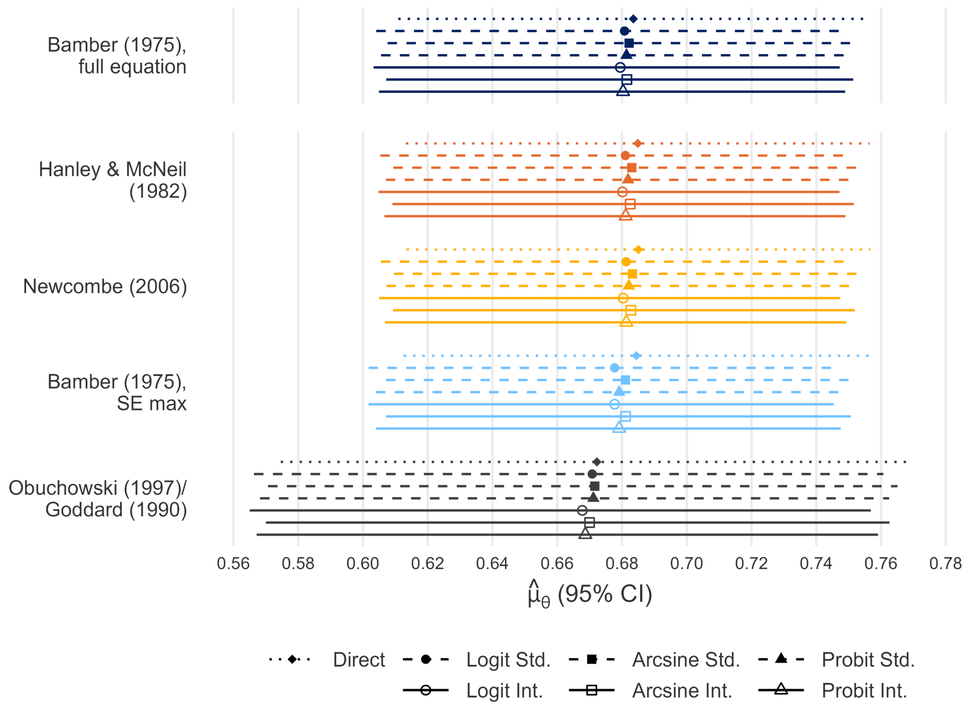


Figure 2: Pooled $\hat{\mu}_\theta$ estimates and 95% CI by sampling-variance method, transformation, and back-transformation

Conclusion

- Compared to the unbiased benchmark $\mathrm{Var}(\hat{\theta})$, the non-parametric and parametric sampling-variance estimators produce similar 95% CI widths (except Bamber's SE max, which overestimates the standard error), and the parametric $\hat{\theta}_i$ are slightly different from the non-parametric $\hat{\theta}_i$ across studies
- After accounting for between-study heterogeneity, differences between the pooled $\hat{\mu}_\theta$ estimates are minimal across meta-analytic approaches with only slight differences for the parametric estimates which are slightly smaller with wider 95% CIs
- A Monte Carlo simulation is currently underway to examine these methods under controlled and broader simulation settings

References

1. McGraw & Wong (1992). Psychol. Bull. 111:361–365. doi:10.1037/0033-2909.111.2.361.
2. Bamber (1975). J. Math. Psychol. 12:387–415. doi:10.1016/0022-2496(75)90001-2.
3. Brunner & Konietschke (2024). Statistical Papers 66:20. doi:10.1007/s00362-024-01635-0.
4. Obuchowski & McClish (1997). Stat. Med. 16:1529–1542. doi:10.1002/(SICI)1097-0258(19970715)16:13<1529::AID-SIM565>3.0.CO;2-H.
5. Goddard & Hinberg (1990). Stat. Med. 9:325–337. doi:10.1002/sim.4780090315.
6. Hanley & McNeil (1982). Radiology 143:29–36. doi:10.1148/radiology.143.1.7063747.
7. Cho, Matthews & Harel (2019). Int. Stat. Rev. 87:152–177. doi:10.1111/insr.12277.
8. Newcombe (2006). Stat. Med. 25:559–573. doi:10.1002/sim.2324.
9. Igelmann, Pauly & Viechtbauer (2026). Res. Synth. Methods:1–14. doi:10.1017/rsm.2026.10096.
10. Huber et al. (2000). Int. J. Immunopharmacol. 22:1103–1111. doi:10.1016/S0192-0561(00)00070-9

[29-P6-02]