A re-analysis of about 60,000 sparse data meta-analyses suggests that using an adequate method for pooling matters.

Maxi Schulz[1] **Malte Kramer**[2] **Oliver Kuss**[3] **Tim Mathes**[1]

[1]University Medical Center Göttingen, Germany [2]German Rheumatism Research Center Berlin, Germany
[3]German Diabetes Center (DDZ) Düsseldorf, Germany

Background

Challenges in sparse data meta-analyses: Meta-analyses that include only a few trials or trials with zero events in one or both arms are common.[1, 2] Such **sparse data** situations pose a challenge for statistical analyses. Simulations have shown that conventionally used inverse-variance statistical methods can lead to **anti-conservative results**.[1, 2, 3, 4] Although better-performing **one-stage methods** have become available in recent years, their implementation remains **limited in practice**. This study aimed to empirically compare conventional meta-analysis methods with one-stage methods for estimating the summary **odds ratio (OR)** in situations characterised by sparse data.

Methods

Data from the CDSR: We retrieved data from all available analyses in the Cochrane Database of Systematic Reviews (CDSR) up until October 25, 2021, to investigate the actual **impact of using conventional instead of one-stage methods** for sparse data scenarios in practice. We created two subsets with

- trials with **zero events in one or both arms** in subset (a), and
- **few trials** (2 to 5) in subset (b).

Meta-analyses with a few trials that reported zero events in one or both arms are included in both subsets.

Models: For each scenario, we computed the following **one-stage models**:

1. Generalized linear mixed model (GLMM),
2. Betabinomial model (BBM),
3. Bayesian binomial-normal hierarchical model using a weakly informative prior (BNHM-WIP)

and compared their results with those of **conventionally used methods** (Peto-Odds-Ratio (PETO), DerSimonianLaird method (DL) for zero event trials; DL, Paule-Mandel (PM), restricted maximum likelihood (REML) method for few trials).

Analyses and measures: We used the following **measures** for evaluation and comparison:

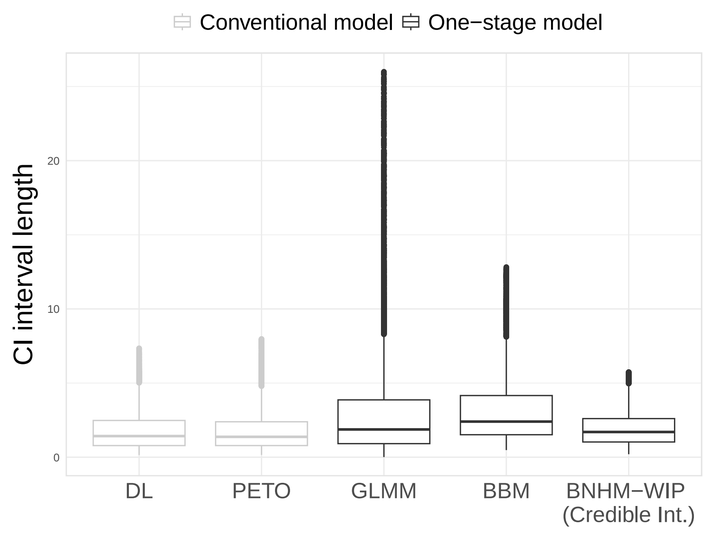
- Log-OR for the treatment effect
- Standard error (SE) for the log-OR
- Length of the 95%-CIs of the log-OR
- Number of meta-analyses for which the treatment effect changed from statistically significant to insignificant, when a one-stage method is used compared to a conventional method.

Results

In both scenarios, we observed:

- Models failed rarely.
- Log-ORs similarly across all models, with both average and median at approximately 0. GLMM with the widest range in computed log-OR.

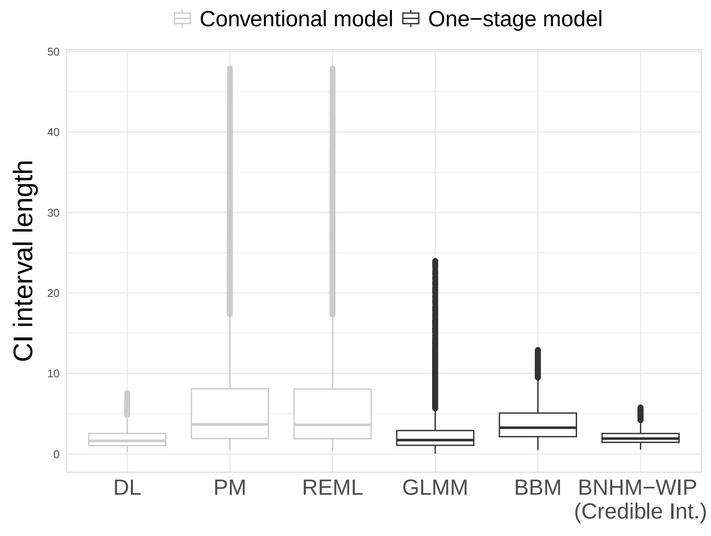
In the **zero event trials** scenario, both GLMM and BBM demonstrate great variability in the length of the CI alongside higher average SEs, while DL, PETO and BNHM-WIP exhibit similar CI lengths.



*Estimated interval lengths of the log-OR in the scenario of zero event trials (a).**

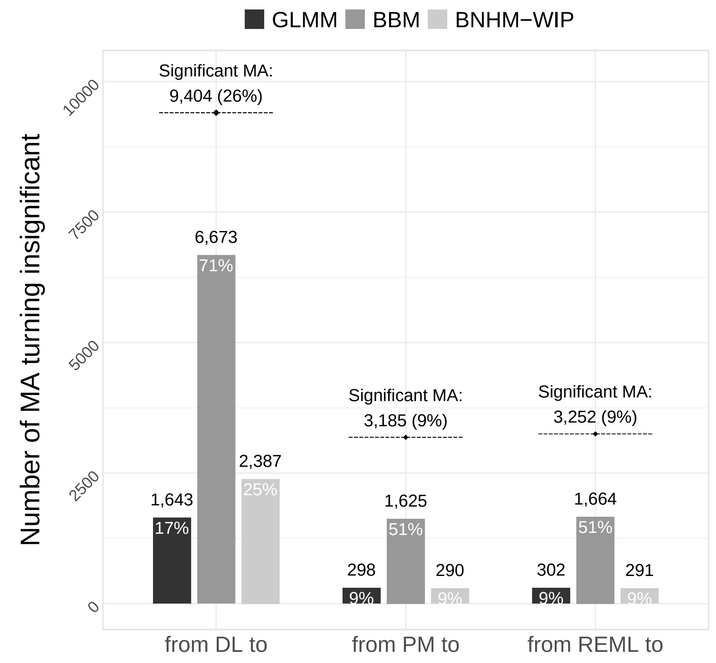
Statistical significance changes most frequently when analysed by the BBM (58%, 61%) and GLMM (24%, 27%) instead of both conventional methods DL and PETO.

In the **few trials** scenario, PM and REML exhibit very large CIs and also demonstrate the highest range in interval lengths, followed by the BBM and GLMM. The average and median 95%-CI lengths, however, were widest for the BBM. The interval length tends to be smaller and less variable in case of DL and BNHM-WIP.



*Estimated interval lengths of the log-OR in the scenario of few trials (b).**

The least frequent significant models are PM, REML and BBM. Significance changes most often when analysed by BBM, even for PM and REML. In case of statistically significant results of PM and REML, heterogeneity τ^2 was mostly estimated to be zero.



Change in significance in the scenario of few trials (b).

Discussion

In the zero event situation, higher average SEs of the one-stage methods GLMM and BBM suggest that established methods might not sufficiently account for **between-study heterogeneity** leading to smaller CIs.

In the few trials scenario, PM and REML resulted in very wide CIs despite ad-hoc modification to the HKSJ CIs. However, they occured mostly in situations where the CIs are fairly wide for all methods, i.e. imprecision is extremely high anyway. Significance often changed from BBM compared to PM and REML, despite the relatively wide CIs of the latter. An explanation could be that heterogeneity variance is often underestimated.

The narrow CIs of the BNHM-WIP across both scenarios may be a consequence of the selected **prior distribution** for the treatment effect.

Conclusion

Our results showed that **statistical precision is highly variable** depending on the method and therefore highlight the importance of **cautious model selection** for meta-analyses in sparse data situations. Using several models for assessing the robustness of findings should be considered.

References

[1] Tim Mathes and Oliver Kuss. A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Research Synthesis Methods*, 9(3):366–381, 2018. doi: 10.1002/jrsm.1296.

[2] Burak Kürsad Günhan, Christian Röver, and Tim Friede. Random-effects meta-analysis of few studies involving rare events. *Research Synthesis Methods*, 11(1):74–90, 2020. doi: 10.1002/jrsm.1370.

[3] Oliver Kuss. Statistical methods for meta-analyses including information from studies without any events: add nothing to nothing and succeed nevertheless. *Statistics in Medicine*, 34(7):1097–1116, 2015. doi: 10.1002/sim.6383.

[4] Chang Xu, Luis Furuya-Kanamori, and Lifeng Lin. Synthesis of evidence from zero-events studies: A comparison of one-stage framework methods. *Research Synthesis Methods*, 13(2):176–189, 2022. doi: 10.1002/jrsm.1521.

* Results are presented excluding extreme outliers (99%-Quartil of the data) and non-converged meta-analyses.

https://doi.org/10.1002/jrsm.1748