Methods Research Report

Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10055-I

Prepared by:
Tufts Medical Center Evidence-based Practice Center
Boston, MA

Investigators:
Thomas A. Trikalinos, M.D., Ph.D.
Paul Trow, Ph.D.
Christopher H. Schmid, Ph.D.

AHRQ Publication No. 13(14)-EHC084-EF
November 2013

This report is based on research conducted by the Tufts Medical Center Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10055-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact: EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Trikalinos TA, Trow P, Schmid CH. Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates. Methods Research Report. (Prepared by the Tufts Medical Center Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 13(14)-EHC084-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard G. Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

Stephanie Chang M.D., M.P.H.
Director, EPC Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Parivash Nourjah, Ph.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Simulation-Based Comparison of Methods for Meta-Analysis of Proportions and Rates

Structured Abstract

Background. In many systematic reviews it is appropriate to summarize proportions and rates (e.g., incidence rates) using meta-analysis. For example, researchers commonly perform meta-analyses of sensitivity and specificity to summarize medical test performance, or of adverse or harmful events. Many statistical methods can be used for meta-analysis of rates and proportions.

Purpose. To help provide guidance for meta-analysts, we performed an extensive simulation study to assess the statistical properties of alternative approaches to meta-analysis of proportions and incidence rates.

Methods. We simulated a large number of scenarios for meta-analyses of proportions and incidence rates (n=792 scenarios for each). The distinct scenarios were defined by combinations of various factors, including the distributional form for the true summary proportion or rate and its defining parameters (mean, variance), the number of studies per meta-analysis, and the number of patients per study.

For each scenario we generated 1000 random meta-analyses, on which we applied fixed and random effects analyses for two families of methods: (1) methods that approximate within-study variability with a normal distribution—not using a transformation, using a canonical transformation (logit and logarithmic for proportions and rates, respectively), or using a variance stabilizing transformation (arcsine and square root for proportions and rates, respectively); and (2) "discrete likelihood" methods that use the theoretically motivated binomial or Poisson distribution to model within study variability. We measured the performance of each method relative to the true values set in the simulation by their mean squared error, bias, and coverage.

Results. In general, and for both proportions and rates, the discrete likelihood approaches performed better than the approximate methods in terms of the three metrics.

Of the approximate methods, the variance stabilizing variants (arcsine transformation for proportions and square root transformation for rates) performed better than the untransformed methods or the methods using a canonical link.

Continuity correction factors are necessary to calculate real-valued means or variances for some approximate methods. The bias, mean square error and coverage of these approximate methods are very sensitive to the choice of continuity correction factors.

Conclusions. Discrete likelihood methods are preferable for the meta-analyses of proportions and rates. We discourage the use of approximate methods that require continuity corrections, as the arbitrary choice of the correction factor can greatly impact on the performance of the method. If software for fitting the discrete likelihood methods is unavailable and expected counts are large enough that normal approximations are adequate, we recommend use of a variance stabilizing transformation.

Contents

Background 1
Methods 3
Overview 3
Description of Simulations 5
Simulation Parameters 5
Generation of Random Data 6
Meta-Analysis Methods 7
Presentation of Simulation Results 10
Results 11
Results for Proportions 12
Overview of Results Across All Methods 12
Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis 23
Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis 39
Comparison Between Fixed and Random-Effects Discrete Likelihood Binomial Methods 56
Results for Rates 59
Overview of Results Across All Methods 61
Practical Recommendations for Meta-Analysis of Proportions or Rates 72
Discussion 73
Bias Stemming From Continuity Corrections 73
Bias Stemming From the Correlation Between Estimate and Its Variance 73
Bias Stemming From a Misspecification of the Random Effects Distribution 74
Other Considerations 75
References 77
Abbreviations 79

Tables

Table 1. Description of performance metrics for the estimated summary proportions and rates 4
Table 2. Simulation parameters for proportions 5
Table 3. Simulation parameters for rates 6
Table 4. Pseudo-algorithm for generating random data 7
Table 5. Formulas for the approximate methods for meta-analysis of proportions and rates in a study 9
Table 6. Comparison of proportion bias across random effects methods (selected scenarios with high heterogeneity) 14
Table 7. Comparison of proportion RMSE across random effects methods (selected scenarios with high heterogeneity) 17
Table 8. Comparison of coverage across random effects methods (selected scenarios with high heterogeneity) 20
Table 9. Mapping of simulation scenarios in the figures of the results section. 23
Table 10. Comparison of proportion bias across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) 63

Table 11. Comparison of proportion RMSE across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) 66
Table 12. Comparison of coverage across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) 69

Figures

Figure 1. Outline of the simulation analysis 3
Figure 2. Comparison of proportion bias between approximate methods: no transformation versus canonical transformation (logit) 25
Figure 3. Comparison of proportion RMSE between approximate methods: no transformation versus canonical transformation (logit) 26
Figure 4. Difference in RMSE between small and large heterogeneity scenarios: no transformation versus canonical transformation (logit) 27
Figure 5. Comparison of coverage between approximate methods: no transformation versus canonical transformation (logit) 28
Figure 6. Comparison of proportion bias between approximate methods: no transformation versus arcsine-transformation 30
Figure 7. Comparison of proportion RMSE between approximate methods: no transformation versus variance stabilizing (arcsine) transformation 31
Figure 8. Difference in RMSE between small and large heterogeneity scenarios: no transformation versus variance stabilizing (arcsine) transformation 32
Figure 9. Comparison of coverage between approximate methods: no transformation versus variance stabilizing (arcsine) transformation 33
Figure 10. Comparison of proportion bias between approximate methods: canonical (logit) versus variance stabilizing (arcsine) transformation 35
Figure 11. Comparison of proportion RMSE between approximate methods: canonical (logit) versus variance stabilizing (arcsine) transformation 36
Figure 12. Difference in RMSE between small and large heterogeneity scenarios: canonical (logit) versus variance stabilizing (arcsine) transformation 37
Figure 13. Comparison of coverage between approximate methods: canonical (logit) versus variance stabilizing (arcsine) transformation 38
Figure 14. Bias induced by correction factors in fixed effects meta-analysis of untransformed data 39
Figure 15. Comparison of proportion bias: approximate method (no transformation) versus discrete likelihood 42
Figure 16. Comparison of proportion RMSE: approximate method (no transformation) versus discrete likelihood 43
Figure 17. Difference in RMSE between small and large heterogeneity scenarios for analyses using the approximate method (no transformation) and the discrete likelihood method 44
Figure 18. Comparison of coverage: approximate method (no transformation) versus discrete likelihood 45
Figure 19. Comparison of proportion bias: approximate method (logit transformation) versus discrete likelihood 47
Figure 20. Comparison of proportion RMSE: approximate method (logit transformation) versus discrete likelihood 48
Figure 21. Difference in RMSE between small and large heterogeneity scenarios for analyses using the approximate methods (logit transformation) and the discrete likelihood method....49

Figure 22. Comparison of coverage: approximate method (logit transformation) versus discrete likelihood 50
Figure 23. Comparison of proportion bias: approximate method (arcsine transformation) versus discrete likelihood 52
Figure 24. Comparison of proportion RMSE: approximate method (arcsine transformation) versus discrete likelihood method..53
Figure 25. Difference in RMSE between small and large heterogeneity scenarios for analyses using approximate methods (arcsine transformation) and the discrete likelihood method 54
Figure 26. Comparison of coverage: approximate method (arcsine transformation) versus discrete likelihood method...55
Figure 27. Comparison of proportion bias: fixed versus random effects with the discrete likelihood method 57
Figure 28. Comparison of proportion RMSE: fixed versus random effects with the discrete likelihood method 58
Figure 29. Comparison of coverage: fixed versus random effects with the discrete likelihood method...59
Figure 30. Side by side comparison of results from simulations for proportions and rates: Comparison of proportion bias for approximate methods, canonical versus variance stabilizing transformations 60

Appendixes

Appendix A. Additional Descriptions of Methods and Results

Background

Meta-analyses of proportions or rates (e.g., incidence rates) are very often included in reports generated by the Effective Health Care Program, and in systematic reviews in general. For example, one would use a meta-analysis of proportions to calculate the summary frequency of adverse or harmful events. Depending on the clinical context, adverse events can be rare (e.g., the incidence rate of rhabdomyolysis in statin treated patients is approximately 3.3 events per 100,000 patient years[1]) or quite common (e.g., the average percentage of nausea and vomiting in chemotherapy-treated cancer patients was 48 percent in a meta-analysis of randomized trials[2]). Meta-analysis of proportions can also be applied to evaluate medical test performance. Although a bivariate meta-analysis of sensitivity and specificity or summarization with a hierarchical summary operating characteristic curve is commonly recommended, separate (univariate) meta-analyses may be sufficient when there is little variation in either sensitivity or specificity.[3,4]

Assuming that it is appropriate to perform a meta-analysis of proportions or rates, many open questions remain about which statistical methods are best to use. Generally speaking, we can group meta-analytic approaches into two families according to how they model within-study variability:

1. Discrete likelihood methods, which model the proportion of events or the incidence rate in a study using the binomial or Poisson distribution, respectively. These are theoretically motivated choices.
2. Approximate methods, which approximate within-study variability with a normal distribution. Of the many variants that have been used, the three that are most interesting[a] use a normal distribution to approximate the distribution of:
    a. Untransformed proportions or rates.
    b. Canonical transformations for proportions (logit transformations) and rates (logarithmic transformations).[b]
    c. Variance stabilizing transformations for proportions (arcsine transformations) and rates (square root transformations).[c]

The approximate methods have known shortcomings. First, they rely on the normal approximation to the binomial or Poisson distributions, which may introduce a bias or have other poor statistical properties when the proportion or rate is close to zero (or one), or when the study sample sizes are relatively small. Both situations are not uncommon in practice. A way to reduce this bias in the extremes (near zero or one) is to use the "canonical" transformations of proportions and rates. Intuitively, these non-linear transformations change the "spacing" of proportions and rates near the extremes. The logit transformation for proportions "expands" values near zero or one (mapping them to the whole real axis) and the logarithmic transformation for rates "expands" values tending to zero (rates are typically small numbers).

A second problem is that, for untransformed and canonically transformed proportions and rates, the sample variance is a function of, and is therefore correlated with, the sample mean. Failure to account for this correlation in the model may bias the summary estimate and its variance.[5-7] Corrections for this bias have been proposed,[5] but are not used in practice.[6,8]

---

[a]Because they have been used commonly, or for theoretical reasons—see later paragraphs.
[b]These are termed "canonical" as they are the typical transformations for the binomial and Poisson families in the framework of Generalized Linear Models. See McCullough P, Nelder JA. Generalized linear models. 2 ed: Chapman & Hall, 1989.
[c]Variance stabilizing transformations are discussed in Chapter 15 of Snedecor GW and Cochran WG. Statistical methods, 7ed. Ames: The Iowa State University Press, 1980.

Alternatively, variance stabilizing transformations, i.e., transformations for which the variance is independent of the estimated mean (arcsine transformations for proportions and square root transformations for rates) are a standard fix for the bias induced by this correlation.[9]

Discrete likelihood methods for random effects are fit as generalized linear mixed (GLMM) models,[7] (e.g., as random effects logistic or Poisson regressions). They cannot be implemented with non-iterative methods in simple spreadsheets,[d] and require programming in a general statistical package such as Stata, R, or SAS. While straightforward to implement for those with solid understanding of GLMMs and statistical languages, they are not easily accessible to meta-analysts without programming skills. Moreover, the routines that fit the discrete likelihood methods use numerical integration algorithms[10] that require examination of the robustness of the results to the algorithms' default settings.[e,11,12] These algorithms can also fail to converge. Inexperienced users may be unable to properly assess and cope with these complexities. As of this writing, no stand-alone meta-analysis software implements GLMMs for meta-analysis of binary data, including the meta-analysis of proportions.[13,14] Therefore, unless the discrete likelihood methods are substantially better than the approximate methods, it is unlikely they will be heavily used.

So which of the above methods should one prefer? A satisfactory answer to this question cannot be a purely empirical one. For example, it cannot be obtained by contrasting these methods in a large number of real-life datasets, because there is no way to know which method is closer to the unobserved "truth." The best approach is to perform a comprehensive simulation study.

We describe an extensive simulation study to assess the statistical properties of the above approaches to meta-analysis of proportions and incidence rates. This study is much broader than previous related investigations. For instance, several large simulation studies have extensively explored different methods for analyzing proportions,[15-18] but did not address meta-analysis across studies. Several others have investigated meta-analysis of proportions,[6] but did not compare all of the above methods in a wide range of realistic simulation scenarios. Stijnen et al[19] discussed the Poisson-normal random effects model for rates, but did no simulations. However, they do provide example SAS code for the discrete likelihood methods.

---

[d] Such as MS Excel™.

[e] Specifically, one has to check the sensitivity of the quadrature or adaptive quadrature method for numerically integrating the likelihood during model fit. Typically, one increases the number of integration points used by the algorithm (which results in slower but more accurate calculations) and checks for any numerical differences in the results.

Methods

Overview

Figure 1 outlines our simulation approach. Briefly, we generated 792 distinct scenarios for meta-analyses of proportions, and 792 scenarios for meta-analyses of incidence rates. These are formed by all possible combinations of choices for the parameters of our generative model.

For each scenario we generated 1,000 random meta-analyses, and analyzed all of them with the methods of interest using both fixed and random effects approaches. We quantified the performance of each method by calculating its bias, mean square error and coverage probability. Table 1 defines these performance metrics.

In the following sections we describe the simulation parameters, the meta-analytic methods compared and our choices for conveying the results of the simulations.

Figure 1. Outline of the simulation analysis

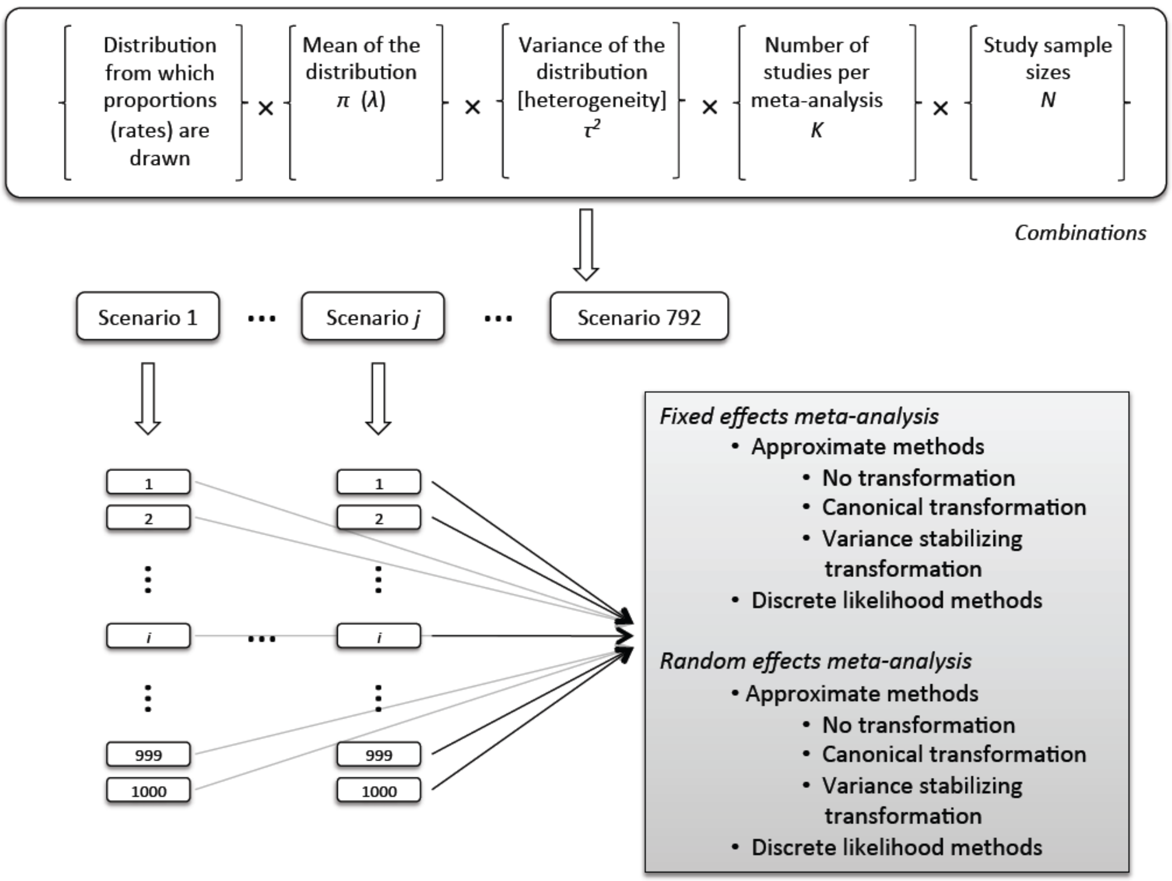


Table 1. Description of performance metrics for the estimated summary proportions and rates

Metric	Formula	Description	Comment
Bias	$\sum_{i=1}^{1000} \frac{\left(\pi_j - \hat{\pi}_{ij}\right)}{1000}$ $\sum_{i=1}^{1000} \frac{\left(\lambda_j - \hat{\lambda}_{ij}\right)}{1000}$	The average difference between the true (simulated) proportion and its estimate across the 1000 simulation replicates in scenario *j*. Similar definition for rates.	• Desirable to have bias near zero.
Proportion Bias	$\sum_{i=1}^{1000} \frac{\left(\pi_j - \hat{\pi}_{ij}\right)}{1000\pi_j}$ $\sum_{i=1}^{1000} \frac{\left(\lambda_j - \hat{\lambda}_{ij}\right)}{1000\lambda_j}$	Bias scaled by π_j or λ_j to make it a proportion of the average effect	• To compare across scenarios with different true proportions and rates
Square root of the mean squared error (RMSE)	$\sum_{i=1}^{1000} \frac{\left(\pi_j - \hat{\pi}_{ij}\right)^2}{1000}$ $\sum_{i=1}^{1000} \frac{\left(\lambda_j - \hat{\lambda}_{ij}\right)^2}{1000}$	The (square root of the) average squared difference between the true (simulated) proportion and its estimate across the 1000 simulation replicates in scenario *j*. Similar definition for rates	• Desirable to have RMSE near zero. • To compare across scenarios with different true proportions π_j, scale RMSE by π_j (Similarly for rates λ_j) • RMSE can be high even if bias is 0, because positive and negative deviations of the estimates from the true mean do not cancel out. • Mean squared error is the sum of the variance of the estimates plus the square of their bias.
Proportion RMSE	$\sum_{i=1}^{1000} \frac{\left(\pi_j - \hat{\pi}_{ij}\right)^2}{1000\pi_j}$ $\sum_{i=1}^{1000} \frac{\left(\lambda_j - \hat{\lambda}_{ij}\right)^2}{1000\lambda_j}$	Bias scaled by π_j or λ_j to make it a proportion of the average effect	• To compare across scenarios with different true proportions and rates.
Coverage probability	$\sum_{i=1}^{1000} I\left(\pi_j \in \left(\text{95\% CI of } \hat{\pi}_j\right)_i\right)/1000$ $\sum_{i=1}^{1000} I\left(\lambda_j \in \left(\text{95\% CI of } \hat{\lambda}_j\right)_i\right)/1000$	The proportion of times the two-sided 95% confidence interval (Wald-type based on normal distribution) of the estimated summary proportion contains the true proportion (or rate).	• Desirable to have coverage near 95%. o Coverage higher than 95% means that $\hat{\pi}$ or $\hat{\lambda}$ is an inefficient estimator o Coverage less than 95% indicates an inaccurate estimator Unlike bias and RMSE, coverage does not need scaling to compare across scenarios.

Simulation scenario is denoted by j; the specific random simulation under scenario j is denoted by i. π_j is the true proportion in scenario j, and its estimate by a meta-analysis method in simulation i for scenario j is denoted by $\widehat{\pi_{ij}}$. Similarly, $\lambda_j$is the true rate and its estimate is $\widehat{\lambda_{ij}}$, in simulation i.

Description of Simulations

Simulation Parameters

Table 2 shows the simulation parameters for proportions. Each simulation scenario represents combinations of the options in the table. The first row has options for the true distribution of the summary proportions across studies. We explicitly avoided choosing a logit-normal distribution for proportions or a log-normal distribution for rates, so as not to bias the results of the simulation study in favor of meta-analytic methods that assume these distributions for the random effects. Further, the examined values cover a wide range of realistic scenarios. For example, the true value of the proportion is varied from near the extreme of zero ($\pi = 0.001$) to $\pi = 0.50$. Because all meta-analysis methods examined for proportions are symmetric around $\pi = 0.50$, it is not necessary to span the range of values between 0.50 and 1 in the simulations. For example, results and conclusions for $\pi = 0.001$ and $1 - \pi = 0.999$ are "mirror images" around $\pi = 0.50$. Regarding heterogeneity, τ^2, we assumed that it could take three levels: zero, small, and large. Zero heterogeneity corresponds to fixed-effects realities, and is probably not the norm in real-life applications. Positive heterogeneity values are more likely to be encountered in reallife applications.

Table 3 shows the respective simulation parameters for rates. In contrast to proportions, the vast majority of incidence rates in applications take very small values. In the simulations, the true rate λ was varied from 1 to 50 events per 1000 person-years.

Table 2. Simulation parameters for proportions

#	Parameter	Values
1	Distribution of summary proportions across studies	Beta, uniform[f] (see Appendix for details)
2	True summary proportion, π	0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50
3	Number of studies, K	5, 15, 30
4	Sample sizes, N*	Vectors of sample sizes for the following choices: All sample sizes small (5-50 patients per study) All sample sizes medium (51-200 patients per study) All sample sizes large (201-1000 patients per study) Mixed sample sizes—approximately 50% small, 40% medium, and 10% large.
5	Heterogeneity, τ^2	Three levels: zero, small, and large. To determine τ, the square root of the heterogeneity variance, true summary proportions were multiplied by 0.10 or 0.50 for small or large heterogeneity, respectively.
6	Correction factor, c†	0, 0.001, 0.01, 0.10, 0.5, 1, 2

For parsimony, in the Results section we present in detail scenarios corresponding to the underlined choices. The index *j* for the scenario has been dropped.
*****The exact values for sample sizes used in the simulations are given in the Appendix.
†Some meta-analysis methods require the use of correction factors. See Continuity Correction Factors for details. The correction factor is an analytic choice and not a simulation parameter; however it is listed here for parsimony.

[f] Strictly speaking, a uniform is a special case of the beta distribution, i.e., Beta(1, 1). The Appendix provides details on the parameters of the modal beta distributions used in the simulations.

Table 3. Simulation parameters for rates

#	Parameter	Description or Values
1	Distribution of summary rates across studies	Gamma, uniform (see Appendix for details)
2	True summary rate, λ	0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50
3	Number of studies, K	5, 15, 30
4	Exposures, **E***	Vectors of exposures for the following choices: a. All sample sizes small (50-200 person-years per study) b. All sample sizes medium (201-500 person-years per study) c. All sample sizes large (501-10000 person-years per study) d. Mixed sample sizes – approximately 50% small, 40% medium, and 10 % large.
5	Heterogeneity, τ^2	Three levels: zero, small, and large. To determine τ, the square root of the heterogeneity, true summary rates were multiplied by 0.10 or 0.50 for small or large heterogeneity, respectively.
6	Correction factor, c †	0, 0.001, 0.01, 0.10, 0.5, 1, 2

For parsimony, in the Results section we present in detail scenarios corresponding the underlined choices. The index *j* for the scenario has been dropped.
*****The vectors of the exact values for sample sizes used in the simulations are given in the Appendix.
†Some meta-analysis methods require the use of correction factors. See text for details. See Continuity Correction Factors for details. The correction factor is an analytic choice and not a simulation parameter; however it is listed here for parsimony.

Generation of Random Data

For each scenario we generated 1000 random meta-analyses of proportions or rates. The following pseudo-algorithm describes the process for scenario *j* (all subscripts for the scenario have been dropped for notational simplicity). The pseudo-algorithm in Table 4 refers to parameters in Table 2 or Table 3, as applicable. We assumed a binomial sampling distribution for proportions and a Poisson sampling distribution for rates so that the model fitting algorithms based on these assumptions are exact.

We verified the fidelity of the simulations by comparing the mean and variance of the empirical distributions of the true values of the simulated proportions or rates for all scenarios versus the respective simulation parameters.

Table 4. Pseudo-algorithm for generating random data

Step	Description	Simulation Table Row #	Comment
	*For a given simulation scenario j:**		
1	Choose the form of the distribution for true proportion (or true rate) in each study in the simulated meta-analyses	1	
2	Choose the average true proportion π (or average true rate λ)	2	
3	Choose the number of studies, *K*	3	
4	Choose the sample sizes for the *K* studies to be a vector of sample sizes $\mathbf{N}=(n_1,\ldots,n_K)$ for proportions [or exposures $\mathbf{E}=(e_1,\ldots,e_K)$ for rates]	4	N and E were drawn once randomly for each K and exposure level , and were then kept fixed for all simulations.
5	Choose τ^2	5	
6	Using the distribution specified in steps 1, 2 and 5, generate 1000 random vectors of true probabilities (or true rates) for the *K* studies: $\Pi_i=(\pi_{i1},\ldots,\pi_{iK})$ for proportions, or $\Lambda_i=(\lambda_{i1},\ldots,\lambda_{iK})$ for rates, where i indexes a random vector, $i=1,\ldots,1000$.	1, 2, 5	If $\tau^2=0$ in Step 5, all $\pi_{ik}=\pi$ for proportions, the value chosen in Step 2. Similarly, all $\lambda_{ik}=\lambda$ for rates. Formulas for calculating the parameters of beta and uniform distributions for proportions (or gamma and uniform distributions for rates) with a given mean and standard deviation are given in the Appendix.
7	Generate 1000 vectors of random data, corresponding to the numbers of events in each study: For proportions the random vector of events, $\mathbf{X}_i=(x_{i1},\ldots,x_{iK})$, is drawn from binomial distributions with probabilities Π_i and sample sizes $\mathbf{N}$. For rates, the random vector of counts, $\mathbf{Y}_i=(y_{i1},\ldots,y_{iK})$, is drawn from Poisson distributions with true rates Λ_i and exposures $\mathbf{E}$.		

*For ease of read, the index *J* for the simulation scenario has been dropped from the notation in the table. The symbol *i* indexes a random meta-analysis, and *k* indexes a study in a meta-analysis.

Meta-Analysis Methods

The underlying outcomes are either binary events (for which the proportion is the mean event rate) or counts of events (for which the incidence rate is the mean event rate per unit of exposure). Assuming these events are independent with nonvarying rates, the binary outcomes follow binomial distributions and the count outcomes follow Poisson distributions. Methods that use these distributions are called “discrete likelihood” methods. When sample sizes are large enough,[g] however, the logit transformation of proportions and the logarithmic transformation of rates approximate normal distributions, and methods for normal distributions may be used. The latter are called “approximate” methods.

We performed both fixed and random effects meta-analyses for both approximate and discrete likelihood methods. For approximate methods, fixed effects meta-analysis used inverse variance weights,[20] and random effects meta-analysis used the moment-based (non-iterative) estimator for between study variance as per DerSimonian and Laird.[21]

For discrete likelihood methods, the meta-analysis was performed in the generalized linear models framework for fixed effects, and the GLMM framework for random effects. For GLMMs we used 12 integration points after examining the robustness of results obtained with 1, 4, 8, 12, 16, and 20 integration points in two example meta-analyses (one with 5 and one with 30 studies of medium sample size, a true proportion or rate of 0.20 and large heterogeneity).

Approximate Methods

Table 5 shows formulas for the three approximate methods used in the simulation: untransformed rates and proportions, canonical transformations (logit and log) and variance stabilizing (arcsine and square root). From the formulas in the table, zero events ($x_{j,ik} = 0$) in, study *k* of simulation *i* in scenario *j* will result in an estimated variance of zero for untransformed proportions and rates, and an undefined estimate and variance for proportions and rates with the canonical transformation. In such pathological cases, inverse variance weights are undefined and continuity correction factors are typically used to make calculations possible, as explained below. Note that because of symmetry, the correction is also needed when $x_{j,ik} = n_{j,k}$ as well (but only for proportions). We opted to examine the behavior of approximate methods using the continuity correction factors, because this is a strategy that many meta-analyses follow in practice.

The canonical or the variance stabilizing transformations yield summary proportions (rates) that have to be back-transformed to the raw proportion (rate) scale using the inverse transformation of the respective functions in Table 5. Corresponding back-transformations are needed for the methods that use the discrete likelihood. In the main analyses we simply back-transformed the point estimates and the confidence interval bounds. Because these back-transformations are nonlinear, becomes a skewed distribution in the raw proportion (rate) scale. The back-transformed point estimate is no longer the mean of the back-transformed distribution of the summary proportion (summary rate), and this introduces a bias. In sensitivity analyses we removed this bias by obtaining the mean the back-transformed values using numerical integration. We calculated the performance metrics in Table 1 using both the simple back-transformation, and with numerical integration. Because results were very similar, we report the main analyses in the main text and present the sensitivity analyses in the Appendix.

[g] More accurately, when the expected numbers in all cells (for proportions) or the expected number of counts (for rates) is relatively large, e.g., >5.

Table 5. Formulas for the approximate methods for meta-analysis of proportions and rates in a study

	Proportions Estimated Mean	Proportions Estimated Variance	Rates Estimated Mean	Rates Estimated Variance
Untransformed proportions or rates	$p_{j,ik} = \frac{x_{j,ik}}{n_{j,k}}$ [cc optional]	$\frac{p_{j,ik}(1-p_{j,ik})}{n_{j,k}}$ [cc needed]	$l_{j,ik} = \frac{y_{j,ik}}{e_{j,k}}$ [cc optional]	$l_{j,ik}$ [cc needed]
Canonical transformations*	$\log \frac{p_{j,ik}}{1-p_{j,ik}}$ [cc needed]	$\frac{1}{x_{j,ik}} + \frac{1}{n_{j,k} - x_{j,ik}}$ [cc needed]	$\log(l_{j,ik})$ [cc needed]	$\frac{1}{y_{j,ik}}$ [cc needed]
Variance stabilizing transformations†	$\arcsin(\sqrt{p_{j,ik}})$ [no cc]	$\frac{1}{4n_{j,k}}$ [no cc]	$\sqrt{l_{j,ik}}$ [no cc]	$\frac{1}{4e_{j,k}}$ [no cc]

cc: continuity correction.

In the Table j indexes the scenario. $\boldsymbol{p_{ik}} = \hat{\boldsymbol{\pi}}_{\boldsymbol{ik}}$ and $\boldsymbol{l_{ik}} = \hat{\boldsymbol{\lambda}}_{\boldsymbol{ik}}$ are estimates of the true proportions and rates, respectively, in study $\boldsymbol{k}$ of simulation $\boldsymbol{i}$. The notation is the same as in Table 4 and in the text.
*Canonical transformations: the logit transformation for proportions and the logarithmic transformation for rates.
†Variance stabilizing transformations: the arcsine transformation for proportions and square root transformation for rates.

Continuity Correction Factors

When $x_{j,ik} = 0$ or $x_{j,ik} = n_{j,k}$, the respective proportion was estimated as $\tilde{\boldsymbol{p}}_{\boldsymbol{j,ik}} = \frac{\boldsymbol{x_{j,ik}} + \boldsymbol{c}}{\boldsymbol{n_{j,k}} + 2\boldsymbol{c}}$, where *c* is the continuity correction factor, or correction factor for short. This adjusted estimate $\tilde{\boldsymbol{p}}_{\boldsymbol{j,ik}}$ was used in the formulas of Table 5 instead of the unadjusted estimate $p_{j,ik}$ when a continuity correction was needed. Continuity corrections were performed only for pathological studies in a meta-analysis. Since for untransformed proportions, the correction factor is needed only in the variance estimate, we performed two sets of analyses in the untransformed case. The first used $\tilde{\boldsymbol{p}}_{\boldsymbol{j,ik}}$ for estimating mean and variance; the second used $p_{j,ik}$ to estimate the mean, but $\tilde{\boldsymbol{p}}_{\boldsymbol{j,ik}}$ to estimate the variance. For parsimony and clarity, we present results with the first set of analyses (which is what many meta-analysts have done), and briefly discuss how the second set of analyses differs from the first.

The same discussion applies to rates. In simulated studies with $y_{j,ik} = 0$ we estimated the respective rate as $\tilde{\boldsymbol{l}}_{\boldsymbol{j,ik}} = \frac{\boldsymbol{y_{j,ik}} + \boldsymbol{c}}{\boldsymbol{e_{j,k}} + 2\boldsymbol{c}} = \frac{\boldsymbol{c}}{\boldsymbol{e_{j,k}} + 2\boldsymbol{c}}$. Continuity corrections were applied only for pathological studies in a meta-analysis. As for proportions, we run two sets of analyses for methods using untransformed rates: the first used $\tilde{\boldsymbol{l}}_{\boldsymbol{j,ik}}$ for estimating mean and variance; the second used $l_{j,ik}$ to estimate the mean, but $\tilde{\boldsymbol{l}}_{\boldsymbol{j,ik}}$ to estimate the variance.

Discrete Likelihood Methods

We fit discrete likelihood methods by maximizing the likelihood in the generalized linear (mixed) models framework using canonical link functions. We used the R command (package) *lmer* with a binomial distribution (corresponding to the logit link function); all analyses were checked for correctness by independently coding them in Stata using the *xtmelogit* command. These methods do not need continuity corrections. For rates, the distribution function in the aforementioned R command is Poisson (whose canonical link function is the log), and the respective Stata command is *xtmepoisson*. When all (or almost all) simulated studies in a meta-analysis have 0 numerator, the random effects methods may not converge. This can happen when the true proportions or rates are small, the sample sizes or exposure sizes are small, or the number of studies in a meta-analysis is small. For each simulation, we recorded whether the random effects method converged or not. For completeness, we examined a hybrid strategy of attempting to fit a random effects model, and switching to a fixed effect model if the random effects model failed to converge.

The fixed effect model for both proportions and rates is mathematically equivalent to simple pooling, i.e., the summary proportion or the summary rate can be calculated by dividing the sum of the numerators by the sum of the denominators.

Presentation of Simulation Results

Categorizing Simulation Scenarios by Expected Counts per Study

The normal approximation to the binomial is adequate, conservatively, when the expected numbers of events (and no events) is at least 5, i.e., $n_{j,k}\pi_{j,ik} > 5$ (and $n_{j,k}(1-\pi_{j,ik}) > 5$). More liberally, one might use a cutoff of 1 event. Hence, for presentation purposes it is sensible to organize the simulated scenarios into 3 categories defined by expected counts of events, namely: those with ≤ 1; >1 and < 5; and ≥ 5 expected events per study. For any given scenario, the expected count of events per study is:

$$expected\ count_j = \frac{\pi_j}{K_j}\Sigma_{k=1}^{K_j}(n_{jk})$$

A similar categorization was used for organizing the simulations on rates, by grouping the scenarios according to the expected number of events in categories with ≤1; between 1 and ≤ 5; and at least 5. The expected number of events is given by a similar formula:

$$expected\ count_j = \frac{\lambda_j}{K_j}\Sigma_{k=1}^{K_j}(e_{jk})$$

Results

The first part of the Results section covers proportions, and the second part covers rates. For parsimony, and after careful examination of all scenarios, we decided to describe in detail findings on a representative subset of the performed analyses.[h] We start with comparisons and recommendations across all methods, and proceed to provide more details on pairwise comparisons between approximate methods and between approximate and discrete likelihood methods. Finally, results on rates were very similar to those from proportions. Thus, we report in detail results in proportions, and provide a succinct summary of the results on rates. We conclude with practical recommendations for applied meta-analysis.

1. Results for Proportions: Results for meta-analysis of proportions:
    a. Overview of Results Across All Methods: Overview of results from comparisons across all methods.
    b. Simulation results for random effects meta-analysis:
        i. Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis: Comparisons between the three approximate methods (untransformed, canonical and variance stabilizing transformation)
        ii. Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis: Comparisons between approximate versus discrete likelihood methods
    c. Comparison Between Fixed and Random-Effects Discrete Likelihood Binomial Methods: Simulation results on the comparison of fixed versus random effects meta-analysis for the discrete likelihood methods
2. Results for Rates: Results for meta-analysis of rates.
    a. Overview of Results Across All Methods: Overview of results from comparisons across all methods.
3. Practical Recommendations for Meta-Analysis of Proportions or Rates: Practical recommendations for meta-analysis.

Observations from simulations are listed as bullet points. An explanation is provided in a grey box after each observation:

- [An observation on the proportion bias, proportion RMSE or coverage probability of a meta-analysis method/strategy.]

[An explanation for the observation]

[h] We report simulation scenarios corresponding to the combinations of the underlined values in Table 2 for proportions, and Table 3 for rates.

Results for Proportions

Overview of Results Across All Methods

Table 6 shows the proportion bias with the random effects analysis methods for selected scenarios with high heterogeneity. Table 7 is the corresponding table for RMSE, and Table 8 for coverage. In the Appendix we present results for the corresponding scenarios with zero heterogeneity. We also conducted sensitivity analyses where the back-transformations of the meta-analysis point estimates were done with numerical integration instead of a simple back-transformation (see Methods section under Approximate Methods). The results of the sensitivity analyses were similar to the main analyses and are not shown.

In each table, scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4).

Results from the five different models are displayed in the last six columns along with the column labeled “Discrete (fraction converged)” which shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns “Discrete (random)” and “Discrete (hybrid)” are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

As an example, the first row of Table 6 describes the proportion bias for a simulation of five studies using a true proportion of 0.001 with small sample sizes (5-50 per study) in which the expected number of events (counts) is less than 0.1. This scenario is labeled S1. Results show that the inverse variance weight method using the proportion or the logit proportion estimates proportions that are between 12 and 15 times the size of the true proportion (i.e., estimated as 0.012 to 0.015, whereas the arcsine method estimates a mean that is 75 percent below the true proportion (i.e., is 0.00025). The discrete likelihood method converges only 16 percent of the time and when it does produces estimates that are nearly 5 times greater than 0.001 on average. Combining this with a fixed effect estimate when the method does not converge reduces this bias to only 5.8 percent.

We make the following general observations for scenarios with expected counts 1 or less, between 1 and 5, and 5 or more:

- For expected counts ≤1, the hybrid method has proportion bias and RMSE that are closer to zero and coverage probability closer to 95 percent compared to other methods.
- For expected counts between 1 and 5 the random effects discrete likelihood method and the approximate method with the variance stabilizing transformation have comparable performance, and perform better than other methods.

[For detailed descriptions and explanations, refer to these two sections: Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis, and Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis]

- For expected counts of 5 or more, the differences between methods become less evident.

For numerical reasons, the random effects discrete likelihood method does not always converge. The following general comments can be made:

For very small expected counts (<0.5) and $K \leq 15$ the random effects discrete likelihood method reached convergence for fewer than 90 percent of the simulations (see column "Discrete (fraction converged)"). For expected counts above 1 and K=30, the method converges practically for all simulations (when the fraction converged equals 1.0, random effects methods converged in all simulations).

- For expected counts above 1 the random effect discrete likelihood method converged (almost) always, and thus the performance of the hybrid strategy is identical to that with random effects (discrete likelihood).

> The random effects discrete likelihood method will not converge when all (or almost all) studies in a meta-analysis have 0 events. This is more common in simulation scenarios with very low expected counts, and when the number of studies is small.

We can make the following general comments on the preferred methods (hybrid strategy-discrete likelihood and variance-stabilizing transformation-approximate likelihood):

- For expected counts above 30, the approximate method using the variance stabilizing transformation has smaller absolute proportion bias and proportion RMSE than the discrete likelihood methods. For smaller counts, the methods using the discrete likelihood have smaller absolute proportion bias and proportion RMSE than the approximate method using the variance stabilizing transformation.

> [For detailed descriptions and explanations, refer to Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis]

- For very large expected counts (for example when the true proportion is 0.4) all compared methods converge in proportion bias and proportion RMSE.

> This is congruent with what we expect theoretically: the normal approximation to the binomial improves with increasing expected counts.

The hybrid strategy using the discrete likelihood has better coverage probabilities than the other methods for the widest range of scenarios.

Table 6. Comparison of proportion bias across random effects methods (selected scenarios with high heterogeneity)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	small	<0.1	12.380	14.529	-0.757	0.160	4.886	-0.058
M1	5	0.001	medium	0.1	2.979	3.778	-0.645	0.497	0.998	-0.007
S2	5	0.005	small	0.2	1.955	2.669	-0.659	0.583	0.550	-0.096
X1	5	0.001	mixed	0.2	0.446	6.042	-0.257	0.648	0.497	-0.030
S3	5	0.01	small	0.3	0.650	1.247	-0.563	0.797	0.099	-0.124
L1	5	0.001	large	0.6	0.094	0.687	-0.445	0.954	-0.046	-0.090
M2	5	0.005	medium	0.6	0.078	0.674	-0.448	0.946	-0.056	-0.107
X2	5	0.005	mixed	1.1	0.022	0.801	-0.280	0.980	-0.002	-0.022
M3	5	0.01	medium	1.3	-0.207	0.330	-0.300	0.997	-0.087	-0.089
S4	5	0.05	small	1.7	-0.240	0.198	-0.261	1.000	-0.101	-0.101
X3	5	0.01	mixed	2.2	-0.034	0.363	-0.263	0.998	-0.005	-0.007
L2	5	0.005	large	3.2	-0.221	0.091	-0.167	1.000	-0.093	-0.093
S5	5	0.1	small	3.4	-0.152	0.082	-0.119	1.000	-0.068	-0.068
L3	5	0.01	large	6.4	-0.134	0.014	-0.100	1.000	-0.089	-0.089
M4	5	0.05	medium	6.4	-0.117	0.012	-0.090	0.999	-0.082	-0.082
X4	5	0.05	mixed	10.9	-0.093	0.085	-0.139	1.000	-0.040	-0.040
M5	5	0.1	medium	12.8	-0.038	-0.021	-0.043	1.000	-0.066	-0.066
S6	5	0.4	small	13.7	-0.014	-0.020	-0.028	1.000	-0.044	-0.044
S7	5	0.5	small	17.1	-0.012	-0.014	-0.014	1.000	-0.016	-0.016
X5	5	0.1	mixed	21.9	-0.111	-0.007	-0.124	1.000	-0.088	-0.088
L4	5	0.05	large	31.8	-0.047	-0.087	-0.072	0.999	-0.110	-0.110
M6	5	0.4	medium	51.4	-0.002	-0.031	-0.018	1.000	-0.040	-0.040
L5	5	0.1	large	63.7	-0.013	-0.081	-0.049	1.000	-0.092	-0.092
M7	5	0.5	medium	64.2	0.001	0.000	0.001	1.000	0.001	0.001

Table 6. Comparison of proportion bias across random effects methods (selected scenarios with high heterogeneity) (continued)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
X6	5	0.4	mixed	87.5	-0.014	-0.023	-0.029	1.000	-0.041	-0.041
X7	5	0.5	mixed	109.4	-0.009	-0.007	-0.010	0.999	-0.012	-0.011
L6	5	0.4	large	254.6	0.003	-0.038	-0.016	1.000	-0.040	-0.040
L7	5	0.5	large	318.3	-0.005	-0.007	-0.006	1.000	-0.008	-0.008
S1	15	0.001	small	<0.1	14.558	19.528	-0.887	0.335	-0.477	-0.505
M1	15	0.001	medium	0.1	3.183	4.300	-0.826	0.815	-0.176	-0.328
X1	15	0.001	mixed	0.1	0.940	6.034	-0.661	0.864	-0.094	-0.217
S2	15	0.005	small	0.1	2.334	3.685	-0.805	0.847	-0.164	-0.292
S3	15	0.01	small	0.3	0.841	1.761	-0.695	0.977	-0.168	-0.187
M2	15	0.005	medium	0.6	0.096	0.815	-0.559	1.000	-0.140	-0.140
L1	15	0.001	large	0.6	0.010	0.794	-0.525	0.999	-0.125	-0.125
X2	15	0.005	mixed	0.7	-0.084	0.971	-0.465	1.000	-0.108	-0.108
M3	15	0.01	medium	1.1	-0.259	0.431	-0.389	1.000	-0.113	-0.113
X3	15	0.01	mixed	1.3	-0.226	0.445	-0.380	1.000	-0.101	-0.101
S4	15	0.05	small	1.4	-0.304	0.340	-0.350	1.000	-0.122	-0.122
S5	15	0.1	small	2.7	-0.237	0.153	-0.184	1.000	-0.087	-0.087
L2	15	0.005	large	3.1	-0.274	0.155	-0.191	0.999	-0.097	-0.097
M4	15	0.05	medium	5.7	-0.168	0.032	-0.118	1.000	-0.102	-0.102
L3	15	0.01	large	6.3	-0.169	0.029	-0.121	1.000	-0.104	-0.104
X4	15	0.05	mixed	6.7	-0.157	0.068	-0.150	1.000	-0.084	-0.084
S6	15	0.4	small	10.8	-0.007	-0.006	-0.029	1.000	-0.042	-0.042
M5	15	0.1	medium	11.5	-0.073	-0.026	-0.074	1.000	-0.096	-0.096
X5	15	0.1	mixed	13.4	-0.107	0.002	-0.098	1.000	-0.090	-0.090
S7	15	0.5	small	13.5	0.006	0.007	0.006	1.000	0.008	0.008
L4	15	0.05	large	31.4	-0.038	-0.074	-0.063	1.000	-0.105	-0.105
M6	15	0.4	medium	45.9	-0.009	-0.044	-0.030	1.000	-0.057	-0.057
X6	15	0.4	mixed	53.5	-0.005	-0.027	-0.025	1.000	-0.047	-0.047
M7	15	0.5	medium	57.4	-0.001	-0.001	-0.001	1.000	-0.001	-0.001
L5	15	0.1	large	62.8	-0.014	-0.085	-0.053	1.000	-0.101	-0.101
X7	15	0.5	mixed	66.9	0.001	0.001	0.002	1.000	0.003	0.003
L6	15	0.4	large	251.4	0.003	-0.043	-0.018	1.000	-0.046	-0.046
L7	15	0.5	large	314.2	-0.002	-0.002	-0.002	1.000	-0.002	-0.002
S1	30	0.001	small	<0.1	14.019	19.604	-0.937	0.546	-0.677	-0.823
M1	30	0.001	medium	0.1	3.359	4.686	-0.849	0.956	-0.346	-0.375
X1	30	0.001	mixed	0.1	1.113	6.608	-0.737	0.974	-0.225	-0.245
S2	30	0.005	small	0.1	2.248	3.744	-0.819	0.978	-0.311	-0.326
S3	30	0.01	small	0.3	0.786	1.790	-0.710	1.000	-0.190	-0.190

Table 6. Comparison of proportion bias across random effects methods (selected scenarios with high heterogeneity) (continued)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
M2	30	0.005	medium	0.5	0.129	0.902	-0.586	1.000	-0.135	-0.135
L1	30	0.001	large	0.5	0.081	0.922	-0.576	1.000	-0.126	-0.126
X2	30	0.005	mixed	0.6	-0.140	1.093	-0.474	1.000	-0.110	-0.110
M3	30	0.01	medium	1.1	-0.242	0.505	-0.404	1.000	-0.093	-0.093
X3	30	0.01	mixed	1.3	-0.287	0.539	-0.358	1.000	-0.096	-0.096
S4	30	0.05	small	1.4	-0.318	0.393	-0.334	1.000	-0.096	-0.096
S5	30	0.1	small	2.7	-0.243	0.167	-0.194	0.999	-0.096	-0.095
L2	30	0.005	large	2.7	-0.311	0.206	-0.220	1.000	-0.102	-0.102
M4	30	0.05	medium	5.4	-0.182	0.043	-0.128	1.000	-0.109	-0.109
L3	30	0.01	large	5.5	-0.190	0.056	-0.131	1.000	-0.103	-0.103
X4	30	0.05	mixed	6.4	-0.140	0.067	-0.157	1.000	-0.101	-0.101
M5	30	0.1	medium	10.8	-0.080	-0.024	-0.078	1.000	-0.100	-0.100
S6	30	0.4	small	10.9	-0.005	-0.005	-0.029	1.000	-0.043	-0.043
X5	30	0.1	mixed	12.7	-0.091	-0.006	-0.108	1.000	-0.102	-0.102
S7	30	0.5	small	13.6	0.002	0.002	0.003	1.000	0.003	0.003
L4	30	0.05	large	27.4	-0.043	-0.072	-0.066	1.000	-0.108	-0.108
M6	30	0.4	medium	43.0	-0.001	-0.037	-0.023	1.000	-0.050	-0.050
X6	30	0.4	mixed	50.8	-0.002	-0.026	-0.024	1.000	-0.045	-0.045
M7	30	0.5	medium	53.8	0.003	0.004	0.004	1.000	0.005	0.005
L5	30	0.1	large	54.9	-0.019	-0.087	-0.057	1.000	-0.106	-0.106
X7	30	0.5	mixed	63.5	0.002	0.001	0.001	1.000	0.001	0.001
L6	30	0.4	large	219.4	-0.005	-0.055	-0.028	1.000	-0.059	-0.059
L7	30	0.5	large	274.3	0.005	0.006	0.005	1.000	0.006	0.006

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4, 7=0.5).

The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Table 7. Comparison of proportion RMSE across random effects methods (selected scenarios with high heterogeneity)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	small	<0.1	12.407	14.631	1.025	0.160	5.211	2.277
M1	5	0.001	medium	0.1	3.018	3.924	0.905	0.497	1.483	1.263
S2	5	0.005	small	0.2	2.002	2.804	0.838	0.583	1.004	1.002
X1	5	0.001	mixed	0.2	1.095	6.098	0.838	0.648	1.022	1.014
S3	5	0.01	small	0.3	0.733	1.442	0.760	0.797	0.685	0.759
L1	5	0.001	large	0.6	0.317	0.919	0.658	0.954	0.576	0.602
M2	5	0.005	medium	0.6	0.289	0.912	0.660	0.946	0.567	0.598
X2	5	0.005	mixed	1.1	0.595	0.965	0.545	0.980	0.576	0.587
M3	5	0.01	medium	1.3	0.366	0.569	0.528	0.997	0.458	0.461
S4	5	0.05	small	1.7	0.391	0.437	0.473	1.000	0.404	0.404
X3	5	0.01	mixed	2.2	0.466	0.609	0.513	0.998	0.491	0.493
L2	5	0.005	large	3.2	0.391	0.352	0.391	1.000	0.357	0.357
S5	5	0.1	small	3.4	0.354	0.316	0.352	1.000	0.323	0.323
L3	5	0.01	large	6.4	0.320	0.285	0.308	1.000	0.299	0.299
M4	5	0.05	medium	6.4	0.312	0.277	0.304	0.999	0.296	0.296
X4	5	0.05	mixed	10.9	0.340	0.366	0.373	1.000	0.364	0.364
M5	5	0.1	medium	12.8	0.266	0.252	0.260	1.000	0.264	0.264
S6	5	0.4	small	13.7	0.255	0.259	0.267	1.000	0.283	0.283
S7	5	0.5	small	17.1	0.247	0.262	0.268	1.000	0.295	0.295
X5	5	0.1	mixed	21.9	0.296	0.283	0.307	1.000	0.304	0.304
L4	5	0.05	large	31.8	0.240	0.241	0.239	0.999	0.253	0.253
M6	5	0.4	medium	51.4	0.242	0.266	0.254	1.000	0.275	0.275
L5	5	0.1	large	63.7	0.243	0.242	0.239	1.000	0.248	0.248
M7	5	0.5	medium	64.2	0.221	0.264	0.243	1.000	0.278	0.278
X6	5	0.4	mixed	87.5	0.257	0.267	0.270	1.000	0.284	0.284
X7	5	0.5	mixed	109.4	0.246	0.272	0.267	0.999	0.294	0.295
L6	5	0.4	large	254.6	0.222	0.253	0.235	1.000	0.256	0.256
L7	5	0.5	large	318.3	0.226	0.281	0.250	1.000	0.287	0.287

Table 7. Comparison of proportion RMSE across random effects methods (selected scenarios with high heterogeneity) (continued)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	15	0.001	small	<0.1	14.568	19.564	0.926	0.335	1.882	1.361
M1	15	0.001	medium	0.1	3.193	4.348	0.854	0.815	0.812	0.850
X1	15	0.001	mixed	0.1	1.233	6.072	0.759	0.864	0.721	0.765
S2	15	0.005	small	0.1	2.347	3.738	0.838	0.847	0.733	0.780
S3	15	0.01	small	0.3	0.866	1.834	0.744	0.977	0.564	0.578
M2	15	0.005	medium	0.6	0.179	0.910	0.617	1.000	0.405	0.405
L1	15	0.001	large	0.6	0.161	0.892	0.592	0.999	0.397	0.398
X2	15	0.005	mixed	0.7	0.362	1.042	0.543	1.000	0.410	0.410
M3	15	0.01	medium	1.1	0.294	0.532	0.463	1.000	0.299	0.299
X3	15	0.01	mixed	1.3	0.354	0.542	0.456	1.000	0.313	0.313
S4	15	0.05	small	1.4	0.347	0.434	0.430	1.000	0.291	0.291
S5	15	0.1	small	2.7	0.310	0.247	0.283	1.000	0.226	0.226
L2	15	0.005	large	3.1	0.331	0.250	0.279	0.999	0.222	0.222
M4	15	0.05	medium	5.7	0.240	0.167	0.207	1.000	0.195	0.195
L3	15	0.01	large	6.3	0.238	0.166	0.208	1.000	0.193	0.193
X4	15	0.05	mixed	6.7	0.239	0.203	0.243	1.000	0.209	0.209
S6	15	0.4	small	10.8	0.148	0.147	0.160	1.000	0.169	0.169
M5	15	0.1	medium	11.5	0.164	0.140	0.160	1.000	0.170	0.170
X5	15	0.1	mixed	13.4	0.196	0.164	0.192	1.000	0.186	0.186
S7	15	0.5	small	13.5	0.135	0.147	0.156	1.000	0.173	0.173
L4	15	0.05	large	31.4	0.149	0.154	0.153	1.000	0.173	0.173
M6	15	0.4	medium	45.9	0.135	0.157	0.148	1.000	0.167	0.167
X6	15	0.4	mixed	53.5	0.142	0.156	0.153	1.000	0.168	0.168
M7	15	0.5	medium	57.4	0.132	0.161	0.148	1.000	0.172	0.172
L5	15	0.1	large	62.8	0.136	0.156	0.142	1.000	0.166	0.166
X7	15	0.5	mixed	66.9	0.136	0.161	0.153	1.000	0.176	0.176
L6	15	0.4	large	251.4	0.128	0.154	0.138	1.000	0.156	0.156
L7	15	0.5	large	314.2	0.129	0.165	0.145	1.000	0.170	0.170

Table 7. Comparison of proportion RMSE across random effects methods (selected scenarios with high heterogeneity) (continued)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	30	0.001	small	<0.1	14.024	19.618	0.944	0.546	1.072	1.040
M1	30	0.001	medium	0.1	3.365	4.713	0.862	0.956	0.732	0.746
X1	30	0.001	mixed	0.1	1.225	6.628	0.775	0.974	0.623	0.635
S2	30	0.005	small	0.1	2.255	3.771	0.835	0.978	0.661	0.670
S3	30	0.01	small	0.3	0.798	1.825	0.734	1.000	0.459	0.459
M2	30	0.005	medium	0.5	0.173	0.957	0.615	1.000	0.325	0.325
L1	30	0.001	large	0.5	0.144	0.973	0.605	1.000	0.317	0.317
X2	30	0.005	mixed	0.6	0.271	1.126	0.513	1.000	0.295	0.295
M3	30	0.01	medium	1.1	0.261	0.561	0.440	1.000	0.223	0.223
X3	30	0.01	mixed	1.3	0.337	0.584	0.406	1.000	0.234	0.234
S4	30	0.05	small	1.4	0.340	0.441	0.379	1.000	0.214	0.214
S5	30	0.1	small	2.7	0.281	0.214	0.243	0.999	0.170	0.170
L2	30	0.005	large	2.7	0.339	0.254	0.266	1.000	0.179	0.179
M4	30	0.05	medium	5.4	0.217	0.120	0.174	1.000	0.157	0.157
L3	30	0.01	large	5.5	0.223	0.129	0.179	1.000	0.157	0.157
X4	30	0.05	mixed	6.4	0.192	0.149	0.205	1.000	0.166	0.166
M5	30	0.1	medium	10.8	0.136	0.107	0.133	1.000	0.147	0.147
S6	30	0.4	small	10.9	0.109	0.110	0.121	1.000	0.129	0.129
X5	30	0.1	mixed	12.7	0.144	0.111	0.156	1.000	0.151	0.151
S7	30	0.5	small	13.6	0.105	0.114	0.123	1.000	0.135	0.135
L4	30	0.05	large	27.4	0.107	0.117	0.114	1.000	0.142	0.142
M6	30	0.4	medium	43.0	0.092	0.109	0.101	1.000	0.118	0.118
X6	30	0.4	mixed	50.8	0.099	0.110	0.110	1.000	0.123	0.123
M7	30	0.5	medium	53.8	0.091	0.113	0.104	1.000	0.123	0.123
L5	30	0.1	large	54.9	0.095	0.125	0.107	1.000	0.140	0.140
X7	30	0.5	mixed	63.5	0.099	0.118	0.113	1.000	0.129	0.129
L6	30	0.4	large	219.4	0.092	0.121	0.103	1.000	0.124	0.124
L7	30	0.5	large	274.3	0.096	0.124	0.108	1.000	0.127	0.127

“Discrete” stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4, 7=0.5).

The column “Discrete (fraction converged)” shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns “Discrete (random)” and “Discrete (hybrid)” are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Table 8. Comparison of coverage across random effects methods (selected scenarios with high heterogeneity)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	small	<0.1	1.000	0.000	1.000	0.160	0.931	0.989
M1	5	0.001	medium	0.1	1.000	0.000	0.997	0.497	0.938	0.969
S2	5	0.005	small	0.2	1.000	0.417	1.000	0.583	0.938	0.964
X1	5	0.001	mixed	0.2	0.992	0.000	0.645	0.648	0.966	0.978
S3	5	0.01	small	0.3	0.999	0.766	0.796	0.797	0.976	0.981
L1	5	0.001	large	0.6	1.000	0.833	0.815	0.954	0.960	0.962
M2	5	0.005	medium	0.6	1.000	0.844	0.806	0.946	0.963	0.965
X2	5	0.005	mixed	1.1	0.814	0.798	0.826	0.980	0.941	0.922
M3	5	0.01	medium	1.3	0.926	0.896	0.842	0.997	0.972	0.969
S4	5	0.05	small	1.7	0.755	0.923	0.879	1.000	0.958	0.958
X3	5	0.01	mixed	2.2	0.763	0.824	0.797	0.998	0.841	0.839
L2	5	0.005	large	3.2	0.701	0.923	0.891	1.000	0.928	0.928
S5	5	0.1	small	3.4	0.769	0.938	0.906	1.000	0.918	0.918
L3	5	0.01	large	6.4	0.776	0.911	0.880	1.000	0.889	0.889
M4	5	0.05	medium	6.4	0.768	0.908	0.870	0.999	0.882	0.881
X4	5	0.05	mixed	10.9	0.717	0.740	0.828	1.000	0.702	0.702
M5	5	0.1	medium	12.8	0.830	0.878	0.866	1.000	0.849	0.849
S6	5	0.4	small	13.7	0.861	0.880	0.875	1.000	0.859	0.859
S7	5	0.5	small	17.1	0.853	0.860	0.864	1.000	0.850	0.850
X5	5	0.1	mixed	21.9	0.723	0.819	0.815	1.000	0.750	0.750
L4	5	0.05	large	31.8	0.825	0.866	0.861	0.999	0.842	0.842
M6	5	0.4	medium	51.4	0.835	0.832	0.841	1.000	0.815	0.815
L5	5	0.1	large	63.7	0.840	0.847	0.861	1.000	0.829	0.829
M7	5	0.5	medium	64.2	0.887	0.864	0.891	1.000	0.867	0.867
X6	5	0.4	mixed	87.5	0.827	0.854	0.835	1.000	0.828	0.828
X7	5	0.5	mixed	109.4	0.826	0.837	0.838	0.999	0.836	0.835
L6	5	0.4	large	254.6	0.877	0.852	0.872	1.000	0.853	0.853
L7	5	0.5	large	318.3	0.865	0.829	0.871	1.000	0.845	0.845

Table 8. Comparison of coverage across random effects methods (selected scenarios with high heterogeneity) (continued)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	15	0.001	small	<0.1	0.000	0.000	1.000	0.335	0.872	0.957
M1	15	0.001	medium	0.1	0.215	0.000	0.500	0.815	0.979	0.983
X1	15	0.001	mixed	0.1	0.964	0.000	0.653	0.864	0.956	0.962
S2	15	0.005	small	0.1	0.806	0.000	0.570	0.847	0.967	0.972
S3	15	0.01	small	0.3	0.996	0.023	0.479	0.977	0.988	0.965
M2	15	0.005	medium	0.6	0.999	0.463	0.551	1.000	0.975	0.975
L1	15	0.001	large	0.6	1.000	0.468	0.572	0.999	0.957	0.956
X2	15	0.005	mixed	0.7	0.788	0.295	0.636	1.000	0.953	0.953
M3	15	0.01	medium	1.1	0.781	0.653	0.690	1.000	0.974	0.974
X3	15	0.01	mixed	1.3	0.668	0.623	0.703	1.000	0.919	0.919
S4	15	0.05	small	1.4	0.527	0.711	0.739	1.000	0.946	0.946
S5	15	0.1	small	2.7	0.571	0.847	0.841	1.000	0.921	0.921
L2	15	0.005	large	3.1	0.506	0.839	0.828	0.999	0.928	0.928
M4	15	0.05	medium	5.7	0.684	0.925	0.878	1.000	0.885	0.885
L3	15	0.01	large	6.3	0.671	0.914	0.863	1.000	0.889	0.889
X4	15	0.05	mixed	6.7	0.695	0.862	0.841	1.000	0.900	0.900
S6	15	0.4	small	10.8	0.932	0.922	0.923	1.000	0.918	0.918
M5	15	0.1	medium	11.5	0.846	0.928	0.898	1.000	0.883	0.883
X5	15	0.1	mixed	13.4	0.774	0.909	0.860	1.000	0.873	0.873
S7	15	0.5	small	13.5	0.953	0.923	0.939	1.000	0.933	0.933
L4	15	0.05	large	31.4	0.860	0.868	0.874	1.000	0.844	0.844
M6	15	0.4	medium	45.9	0.934	0.885	0.912	1.000	0.904	0.904
X6	15	0.4	mixed	53.5	0.920	0.896	0.909	1.000	0.905	0.905
M7	15	0.5	medium	57.4	0.959	0.877	0.940	1.000	0.935	0.935
L5	15	0.1	large	62.8	0.890	0.857	0.897	1.000	0.860	0.860
X7	15	0.5	mixed	66.9	0.939	0.892	0.917	1.000	0.931	0.931
L6	15	0.4	large	251.4	0.939	0.880	0.931	1.000	0.920	0.920
L7	15	0.5	large	314.2	0.948	0.857	0.920	1.000	0.919	0.919

Table 8. Comparison of coverage across random effects methods (selected scenarios with high heterogeneity) (continued)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	30	0.001	small	<0.1	0.000	0.000	1.000	0.546	0.963	0.980
M1	30	0.001	medium	0.1	0.000	0.000	0.274	0.956	0.976	0.977
X1	30	0.001	mixed	0.1	0.826	0.000	0.405	0.974	0.972	0.947
S2	30	0.005	small	0.1	0.000	0.000	0.224	0.978	0.983	0.961
S3	30	0.01	small	0.3	0.857	0.000	0.197	1.000	0.977	0.977
M2	30	0.005	medium	0.5	1.000	0.112	0.284	1.000	0.974	0.974
L1	30	0.001	large	0.5	0.999	0.074	0.295	1.000	0.978	0.978
X2	30	0.005	mixed	0.6	0.814	0.009	0.420	1.000	0.963	0.963
M3	30	0.01	medium	1.1	0.666	0.326	0.501	1.000	0.965	0.965
X3	30	0.01	mixed	1.3	0.447	0.244	0.545	1.000	0.939	0.939
S4	30	0.05	small	1.4	0.266	0.412	0.580	1.000	0.936	0.936
S5	30	0.1	small	2.7	0.425	0.755	0.765	0.999	0.919	0.919
L2	30	0.005	large	2.7	0.245	0.667	0.707	1.000	0.916	0.916
M4	30	0.05	medium	5.4	0.553	0.925	0.831	1.000	0.862	0.862
L3	30	0.01	large	5.5	0.505	0.920	0.810	1.000	0.879	0.879
X4	30	0.05	mixed	6.4	0.666	0.871	0.752	1.000	0.861	0.861
M5	30	0.1	medium	10.8	0.800	0.930	0.852	1.000	0.823	0.823
S6	30	0.4	small	10.9	0.934	0.920	0.921	1.000	0.914	0.914
X5	30	0.1	mixed	12.7	0.789	0.936	0.847	1.000	0.869	0.869
S7	30	0.5	small	13.6	0.953	0.909	0.922	1.000	0.925	0.925
L4	30	0.05	large	27.4	0.877	0.876	0.892	1.000	0.800	0.800
M6	30	0.4	medium	43.0	0.960	0.902	0.930	1.000	0.916	0.916
X6	30	0.4	mixed	50.8	0.956	0.912	0.930	1.000	0.930	0.930
M7	30	0.5	medium	53.8	0.976	0.884	0.941	1.000	0.941	0.941
L5	30	0.1	large	54.9	0.911	0.825	0.901	1.000	0.813	0.813
X7	30	0.5	mixed	63.5	0.952	0.880	0.921	1.000	0.930	0.930
L6	30	0.4	large	219.4	0.953	0.850	0.927	1.000	0.906	0.906
L7	30	0.5	large	274.3	0.961	0.847	0.935	1.000	0.931	0.931

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4, 7=0.5). The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis

For random effects meta-analysis, results based on the variance stabilizing (arcsine) transformation have proportion bias and proportion RMSE closer to zero, and coverage probabilities closer to 95 percent, compared to those based on the canonical (logit) transformation or on untransformed data across a wide range of scenarios, as explained below.

Therefore, if one has to use approximate methods for random effects meta-analysis of proportions, we would recommend the variance stabilizing transformation.

No Transformation Versus Canonical Transformation (Logit Transformation)

Figure 2 compares the proportion bias for random effects meta-analysis on untransformed versus logit-transformed data. Its 9 panels are arranged in 3 columns, according to whether the mean expected number of events in the simulated studies is ≤1, between 1 and 5, or ≥ 5; and 3 rows, corresponding to the number of studies, *K*, in the simulated meta-analysis (5, 15, or 30).

In Figure 2, points represent simulation scenarios, and are coded with a letter-number pair, based on sample size and true proportion. Each letter-number pair appears exactly once in each row of plots, as mapped in Table 9.

Table 9. Mapping of simulation scenarios in the figures of the results section

Studies (K)	Expected count ≤ 1	1 < Expected count < 5	Expected count ≥ 5
5	S1, S2, S3, M1, M2, L1, X1	S4, S5, M3, L2, X2, X3	M4, M5, L3, L4, L5, X4, X5
15	S1, S2, S3, M1, M2, L1, X1, X2	S4, S5, M3, L2, X3	M4, M5, L3, L4, L5, X4, X5
30	S1, S2, S3, M1, M2, L1, X1, X2	S4, S5, M3, L2, X3	M4, M5, L3, L4, L5, X4, X5

Simulation scenarios are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). With a single exception, the various simulation scenarios map to the same range of expected count across all three choices of *K* (i.e., they map to the same row). The single exception pertains to X2 (i.e., scenarios with mixed sample sizes, and a true proportion equal to 0.002; underlined). For *K*=5, the expected count for X2 is 1.09; and for *K*=15 and *K*=30 the expected count is 0.67 and 0.64, respectively. This chance variation occurs because the sample sizes were chosen randomly for each *K*, and consequently the expected counts for all scenarios are different for 5 studies than for 15 or 30 studies. It so happens that for X2, this variation results in “changing columns.”

Figure 3 has a similar layout and, denotes the proportion RMSE, which behaves approximately the same as the absolute value of proportion bias (see Methods). Based on the two figures, we make the following observations:

- Within each column of Figure 2, the location of the graphed points (simulation scenarios) is quite similar, especially considering that the exact sample sizes (and thus the expected counts) differ across values of *K* (see Categorizing Simulation Scenarios by Expected Counts per Study). The same observation applies to Figure 3 as well. For both methods, the proportion bias and the proportion RMSE do not change materially with the number of studies, *K*.

> The influence of the number of studies on proportion bias (and the proportion RMSE) is small.

- For both methods, the proportion RMSE is generally highest for scenarios with expected counts <1, and tends to be smaller for higher expected counts. The *absolute value* of proportion bias follows a similar pattern.

> The normal approximations to the binomial are better as the expected count increases, and this translates to better values for proportion bias and proportion RMSE. Further, continuity corrections are more likely to be necessary when the expected count is less than 1; as discussed in the 5th bullet, continuity corrections introduce an upward bias for both methods.

- For expected counts less than 1, proportion bias and proportion RMSE are not very different between scenarios with smaller and larger heterogeneity: the red and black colored points are very near each other in Figure 2 and Figure 3. All other things being equal, proportion RMSE differs between smaller and larger heterogeneity scenarios when the expected counts are larger than 1 (Figure 4; and similarly for the absolute value of proportion bias—not shown).

> For small expected counts, the estimate of between study heterogeneity is very often 0 or close to 0, and this attenuates any differences between scenarios with smaller and larger simulated heterogeneity.

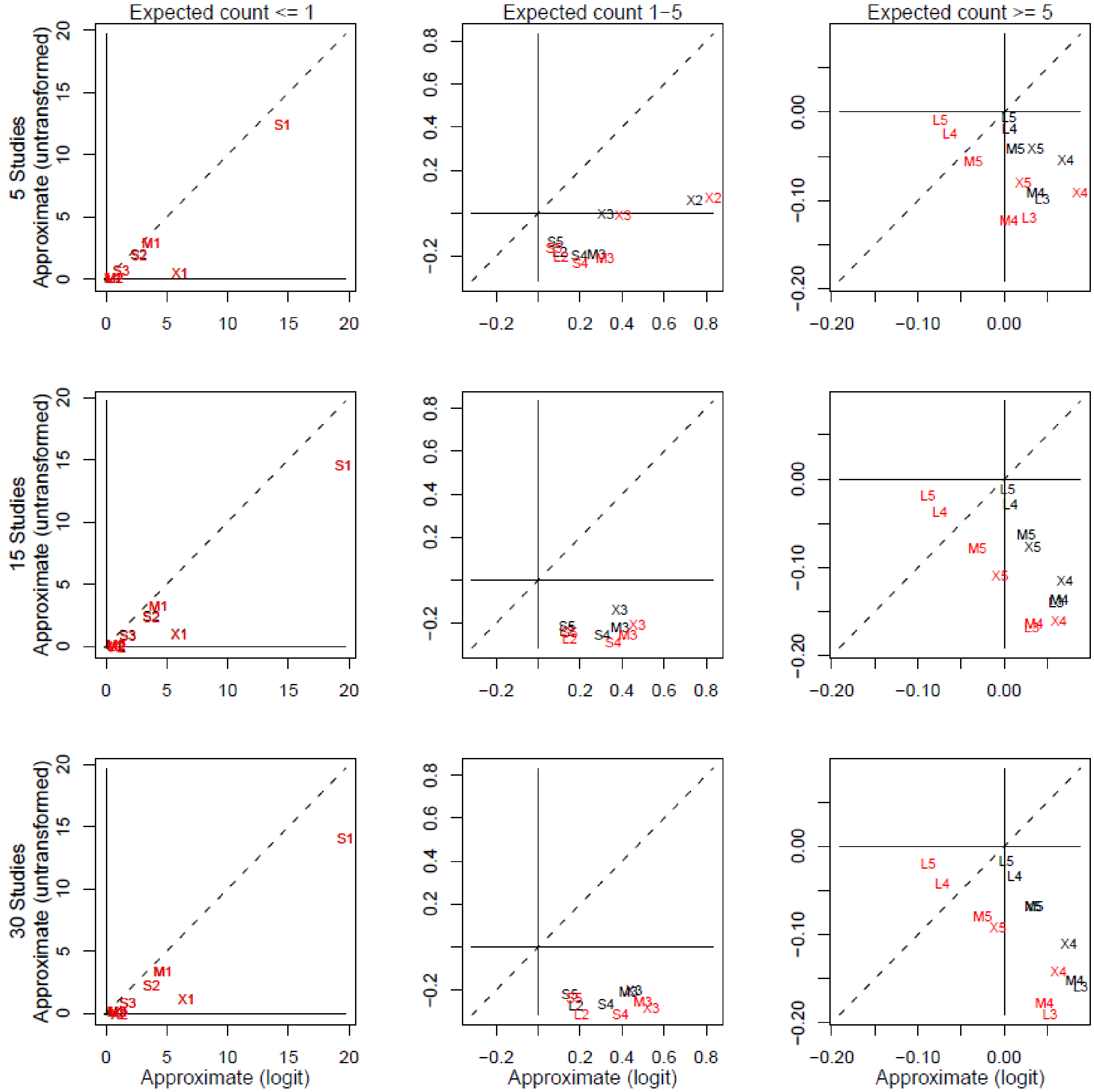
- When the expected count is less than 5, differences in proportion bias (and proportion RMSE) between small and large heterogeneity scenarios are most pronounced for the logit transformation: Within each plot, the difference between red and black colored labels is larger along the horizontal axis compared to the vertical.

> A likely explanation pertains to the "floor effect" of the untransformed data; small proportions are bounded by 0. In contrast, the logit transformation "expands" small proportion values to occupy the whole negative real axis, effectively removing the "floor effect". This allows the summary estimate to be more influenced by studies with smaller proportion values, which are more likely to be drawn in simulations with large heterogeneity.

- Analytically, we expect the proportion bias of the summary estimate to be negative for untransformed data, and positive for logit-transformed data. The use of continuity corrections (*cc*) adds an additional bias component. This additional bias component can be positive, zero, or negative, for $\frac{cc}{n_{j,k}} > \pi_j$, $\frac{cc}{n_{j,k}} = \pi_j$, or $\frac{cc}{n_{j,k}} < \pi_j$, respectively.

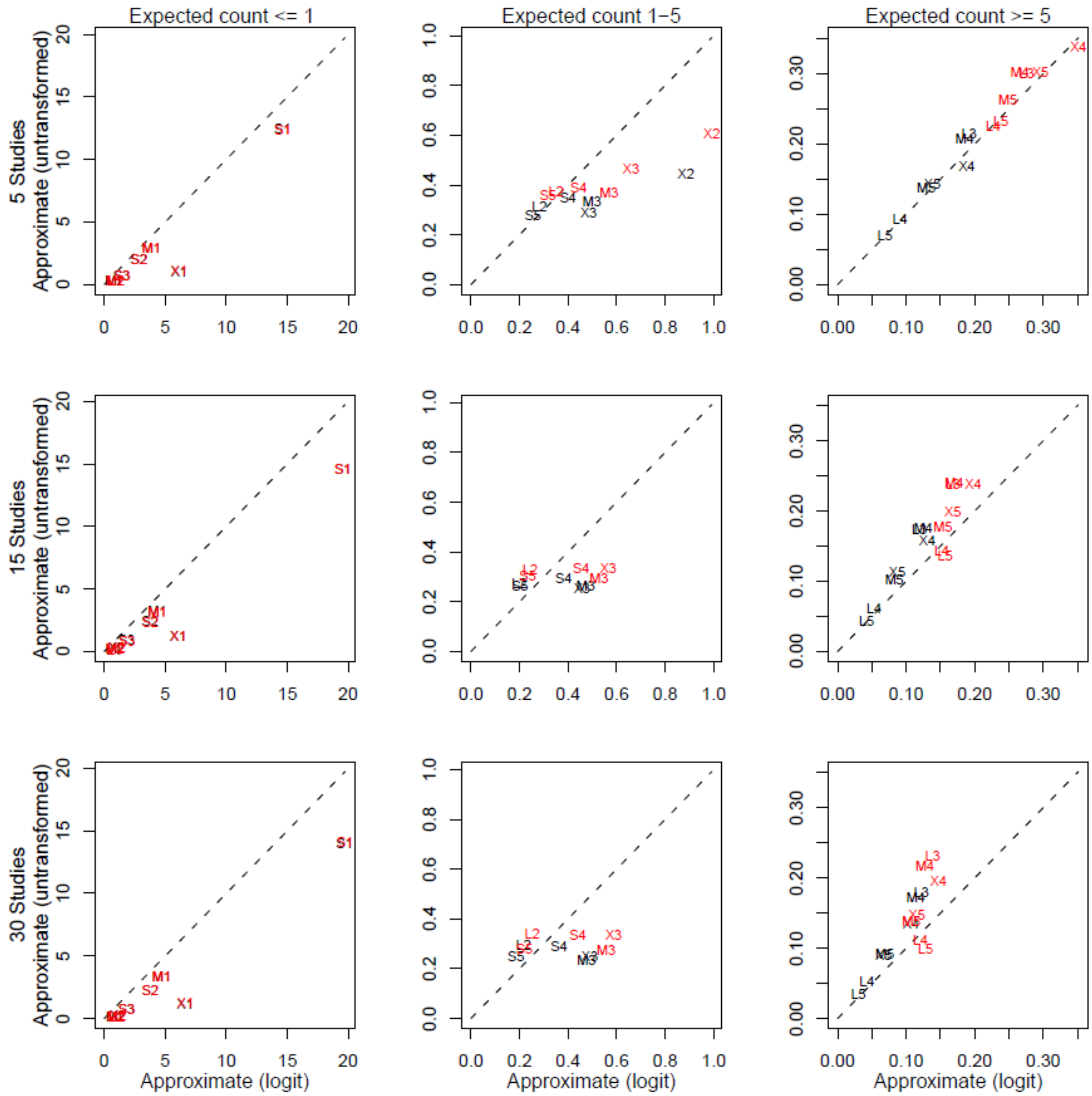
Proportion bias (for both transformed and untransformed data) can be positive, zero, or negative, depending on the relative magnitude of the two bias components, i.e., the bias component related to the mathematical transformation and that related to the correction factors. As heterogeneity increases, it is more difficult to make qualitative predictions for the proportion bias.

Figure 2. Comparison of proportion bias between approximate methods: no transformation versus canonical transformation (logit)



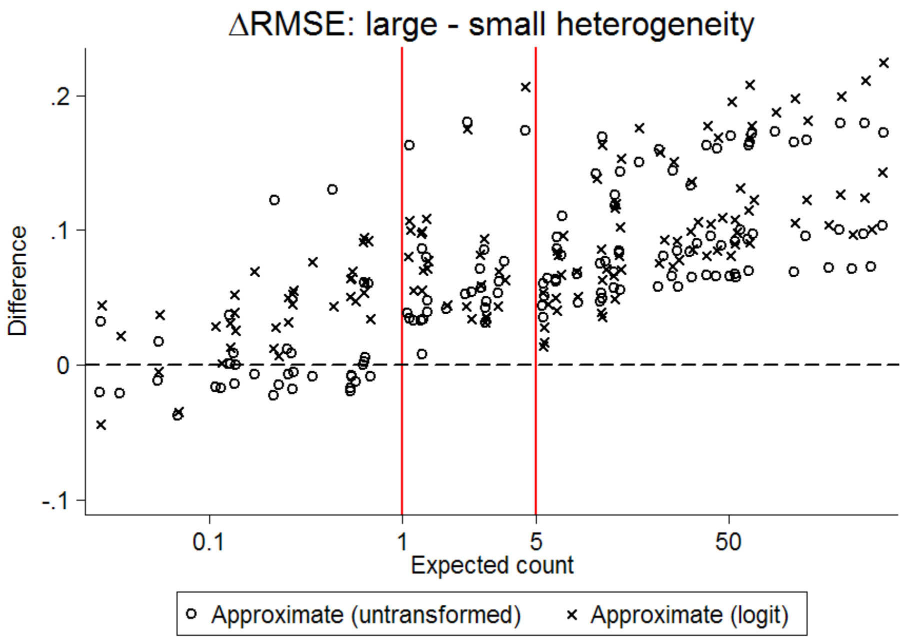
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 3. Comparison of proportion RMSE between approximate methods: no transformation versus canonical transformation (logit)



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 4. Difference in RMSE between small and large heterogeneity scenarios: no transformation versus canonical transformation (logit)



Shown are all simulation scenarios (not only the "representative" ones listed in Table 9). A similar pattern is observed for the absolute value of proportion bias. The horizontal dotted line at zero is the line of no difference. Vertical lines separate scenarios by expected counts categories.

- For both methods, coverage is better when the expected count is at least 5 (Figure 5), compared to ≤ 1 or between 1 and 5.

> The normal distribution approximates the binomial better as expected counts increase above 5.

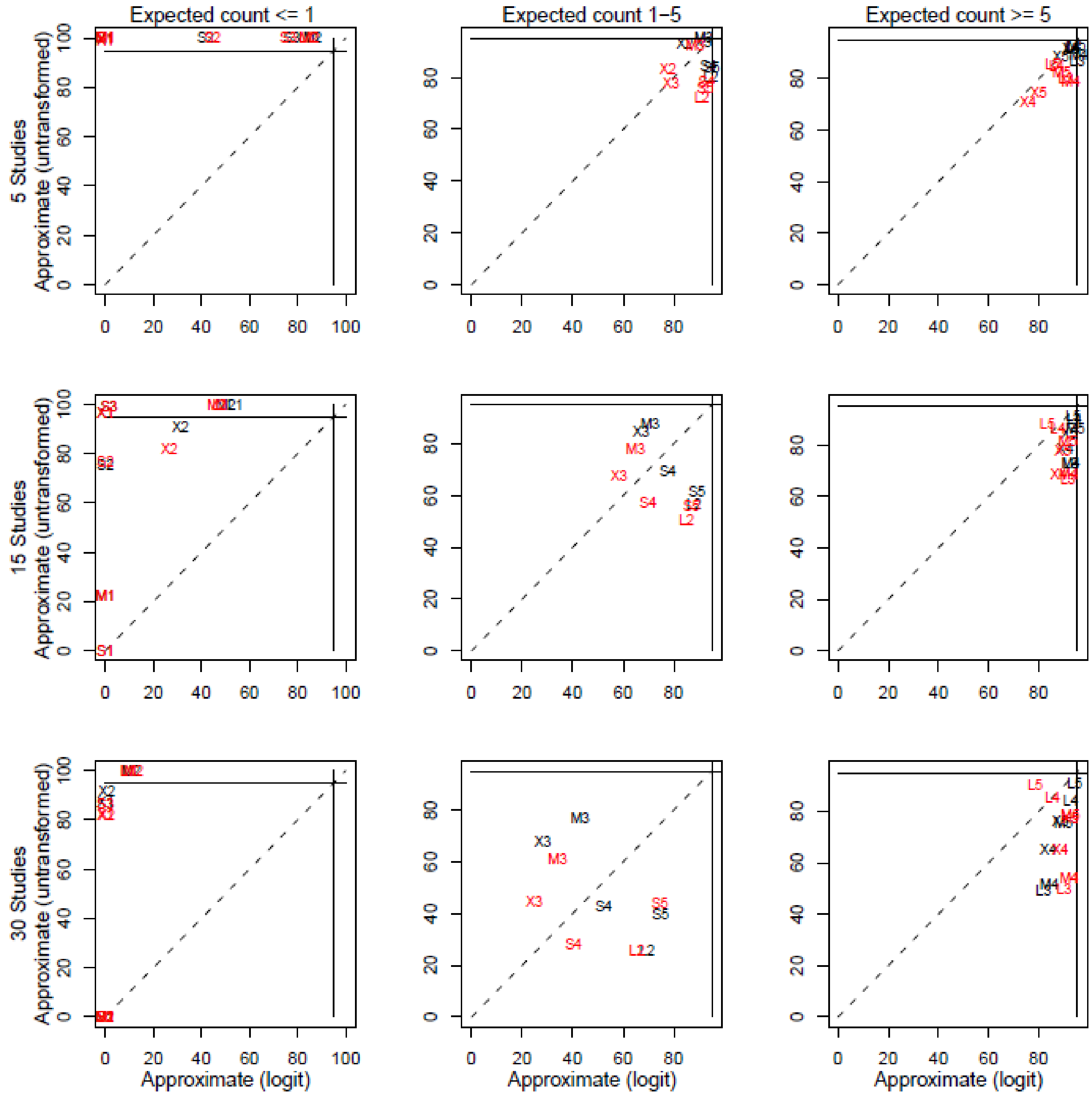
- For expected counts ≥5, coverage is better for the canonical (logit) transformation compared to the untransformed data.
- For expected counts ≤1, the coverage is very often 100 percent for the untransformed data or 0 percent for the logit transformation.

> For expected counts ≤ 1, continuity corrections are often needed for both approximate methods. For untransformed data, the combination of a positive bias (associated with the use of the continuity corrections) and the fact that the lower confidence interval can reach all the way to 0 ("floor effect") may explain the 100 percent coverage.

- Coverage appears to become worse with increasing K. This is more evident for expected counts between 1 and 5.

> We see no obvious explanation for this pattern.

Figure 5. Comparison of coverage between approximate methods: no transformation versus canonical transformation (logit)



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

No Transformation Versus Variance-Stabilizing Transformation (Arcsine Transformation)

Figure 6 and Figure 7 compare proportion bias and proportion RMSE, respectively for random effects meta-analysis on untransformed versus arcsine-transformed data. Many of the observations and explanations below are similar to those made in No Transformation Versus Canonical Transformation (Logit Transformation).

- For small expected counts (≤1), analyses based on the variance-stabilizing transformation has much smaller proportion bias and proportion RMSE than those based on untransformed data. This difference is attenuated for simulation scenarios with larger expected counts.

> Analyses based on untransformed data require continuity corrections (and most often when the expected counts are ≤1). As discussed in No Transformation Versus Canonical Transformation (Logit Transformation), the net effect is a large positive bias and a large proportion RMSE. However, for arcsine transformed data, no continuity corrections are needed, and thus the proportion bias and proportion RMSE are comparatively much smaller.

- For both methods, the proportion bias and the proportion RMSE do not change dramatically with the number of studies, *K*.

> The influence of the number of studies on proportion bias (and the proportion RMSE) is small, as in No Transformation Versus Canonical Transformation (Logit Transformation).

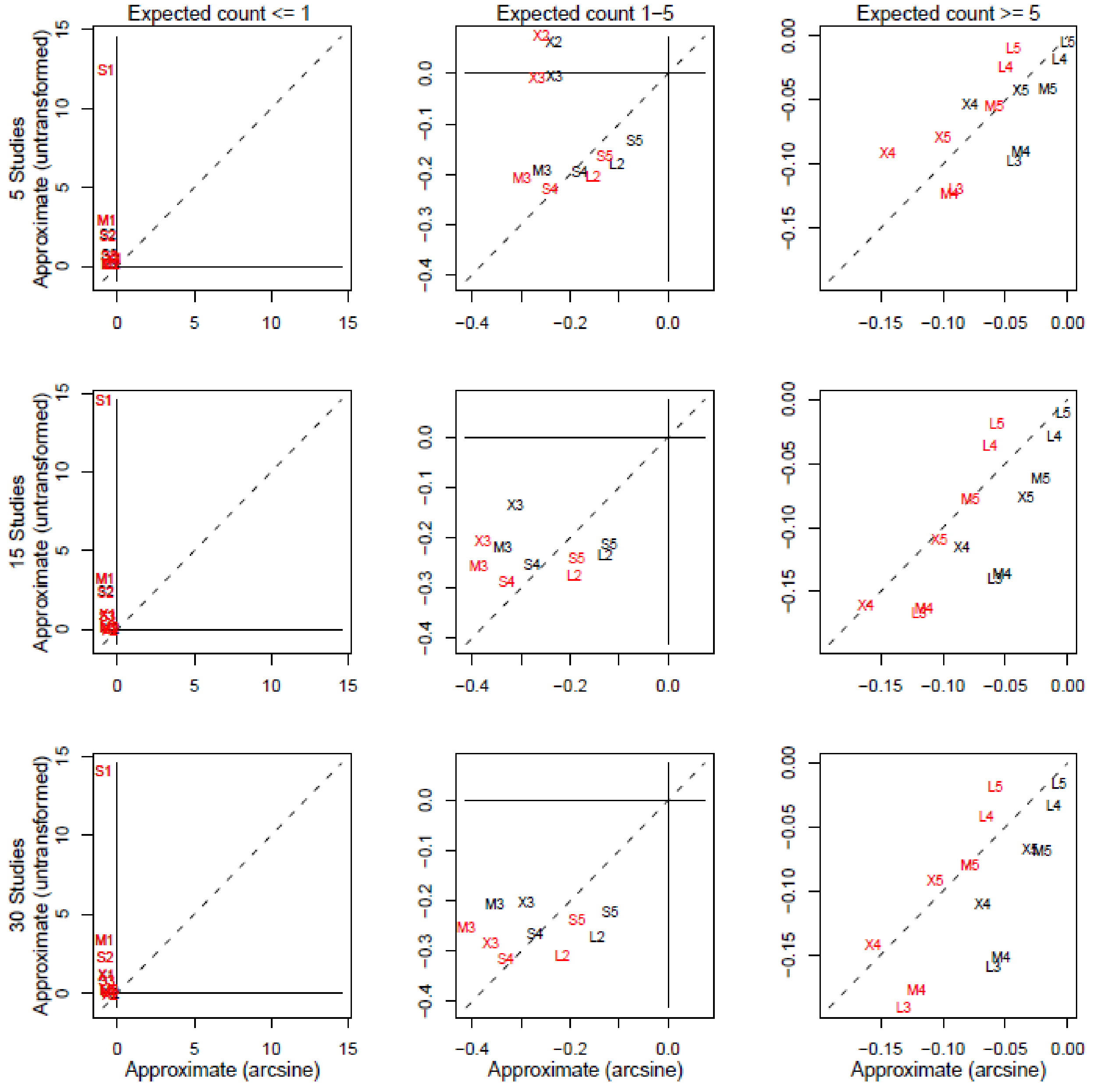
- For expected counts less than 1, proportion bias and proportion RMSE are not very different between scenarios with smaller and larger heterogeneity. All other things being equal, differences in proportion RMSE between smaller and larger heterogeneity scenarios are evident for expected counts larger than 1 (Figure 8; and similarly of the absolute value of proportion bias—not shown).

> A likely explanation is that for small expected counts, the estimate of between study heterogeneity is very often 0 or close to 0, and this attenuates any differences between scenarios with smaller and larger simulated heterogeneity.

- Analytically, we expect the proportion bias to be negative for analyses based on both untransformed and arcsine-transformed data (See appendix). As described in No Transformation Versus Canonical Transformation (Logit Transformation), the need for continuity corrections (*cc*) adds an additional bias component for analyses based on untransformed data. This additional bias component can be positive, zero, or negative, for $\frac{cc}{n_{j,k}} > \pi_j$, $\frac{cc}{n_{j,k}} = \pi_j$, or $\frac{cc}{n_{j,k}} < \pi_j$, respectively.
- For expected counts ≥5, proportion bias is slightly better on the arcsine scale under small amounts of heterogeneity, but no clear trend is evident for RMSE. For expected counts

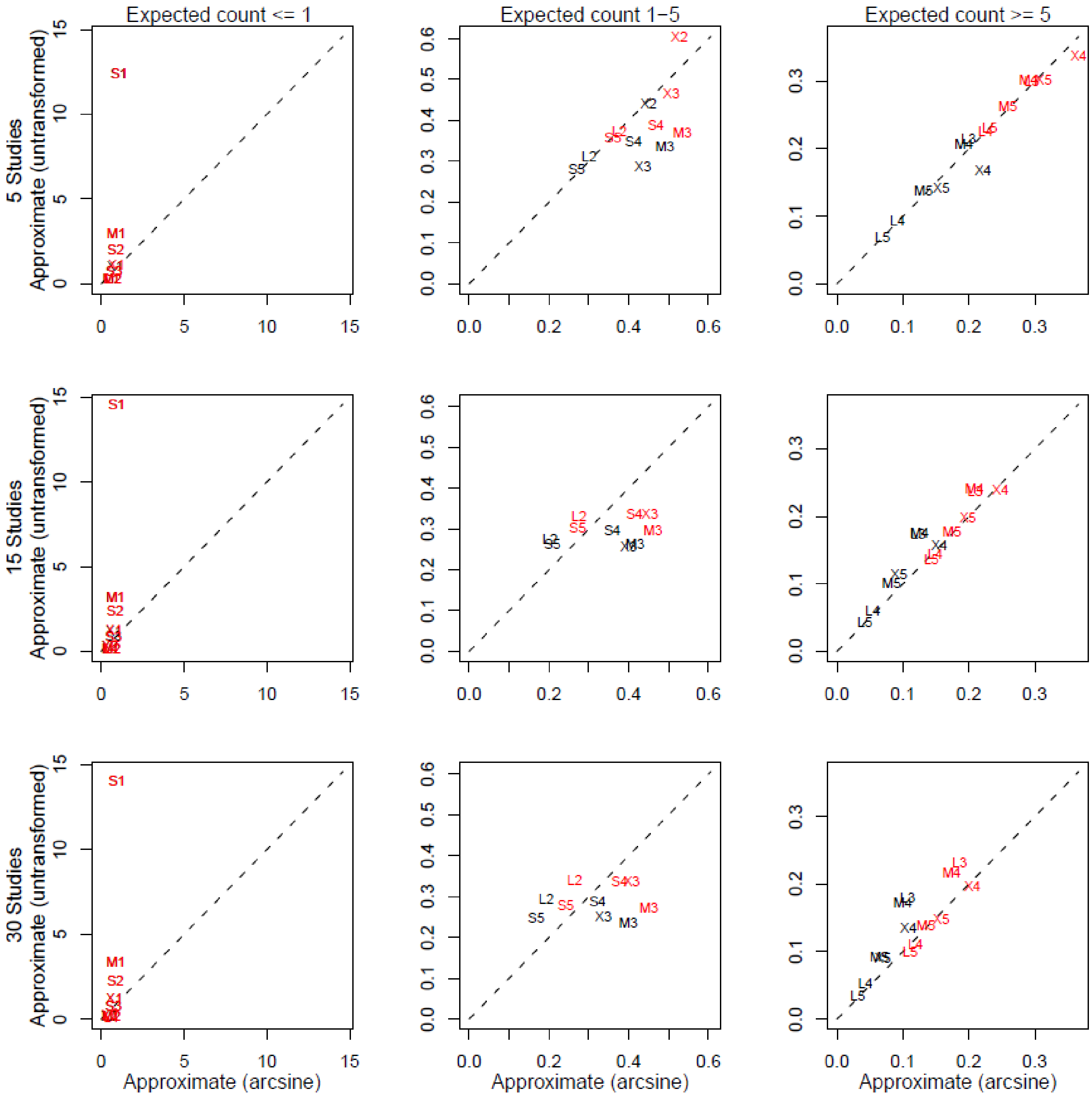
between 1 and 5, the two methods are comparable for proportion bias, but the arcsine scale may have a bit higher RMSE.

Figure 6. Comparison of proportion bias between approximate methods: no transformation versus arcsine-transformation



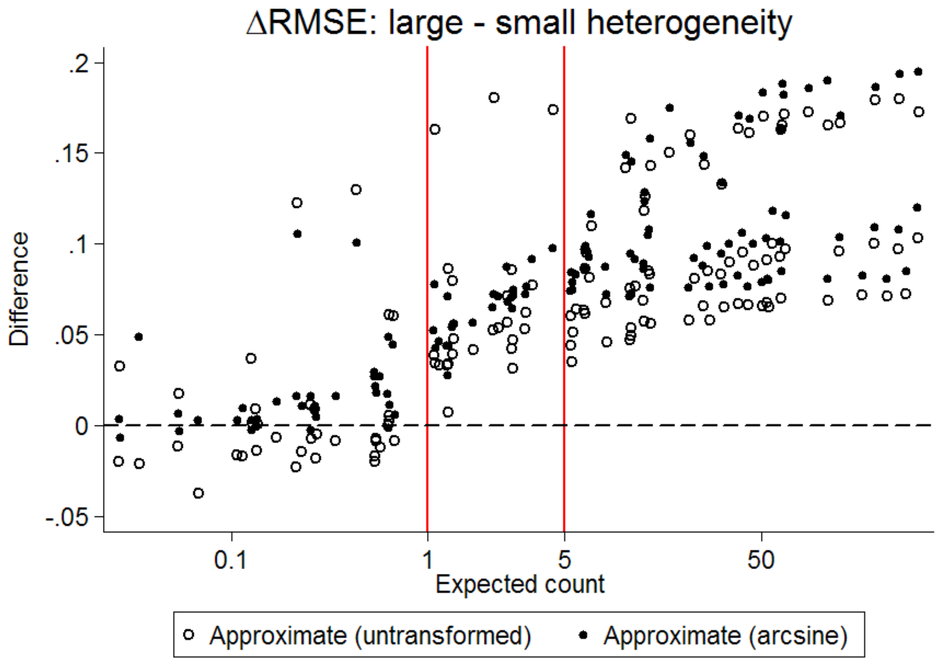
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 7. Comparison of proportion RMSE between approximate methods: no transformation versus variance stabilizing (arcsine) transformation



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 8. Difference in RMSE between small and large heterogeneity scenarios: no transformation versus variance stabilizing (arcsine) transformation



Shown are all simulation scenarios (not only the "representative" ones listed in Table 9). A similar pattern is observed for the absolute value of proportion bias. The horizontal dotted line at zero is the line of no difference. Vertical lines separate scenarios by expected counts categories.

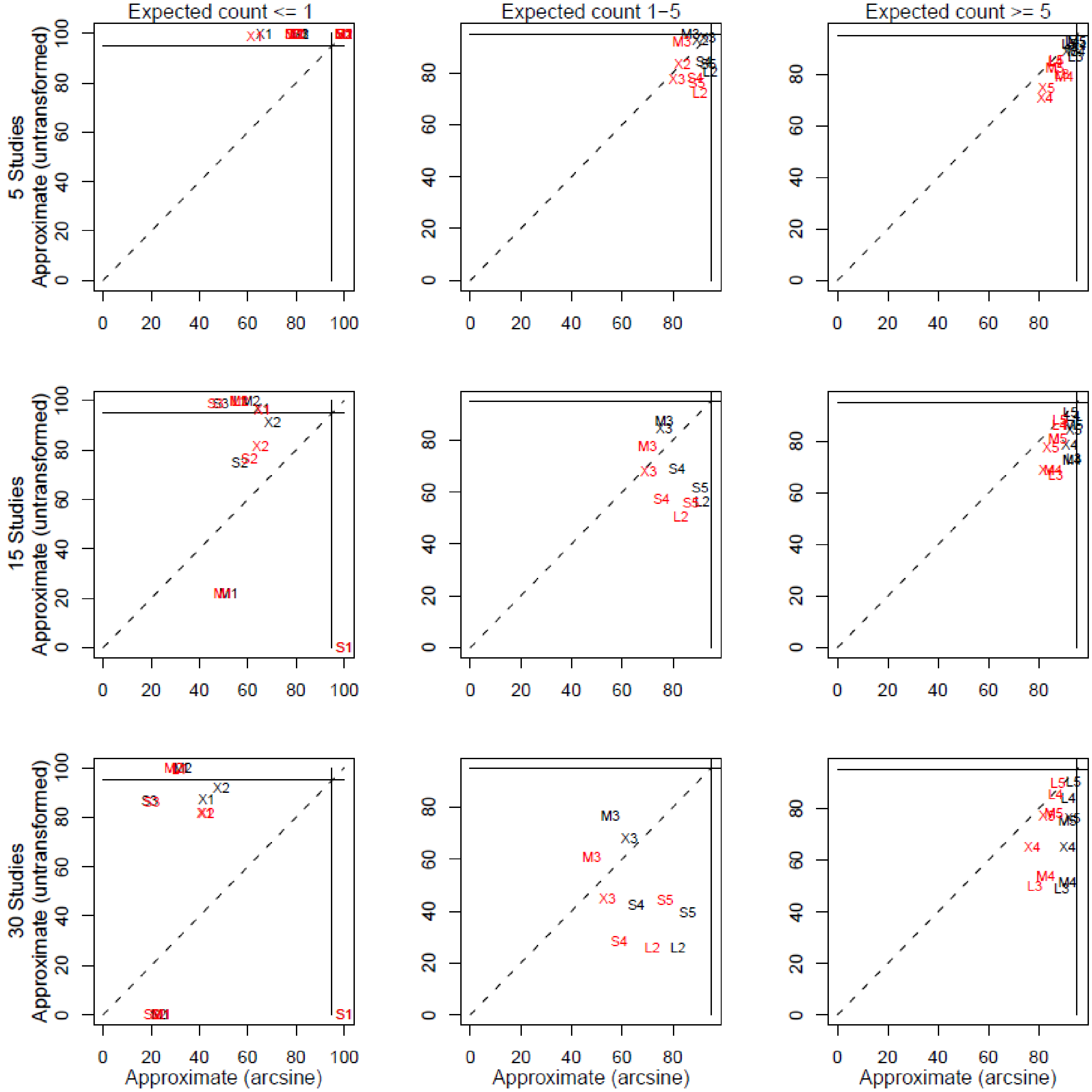
For both methods, coverage is better when the expected count is at least 5 (Figure 9), compared to ≤ 1 or between 1 and 5.

> The normal is a better approximation to the binomial as the expected counts increase above 5.

- For expected counts ≥5, coverage is better for the variance stabilizing transformation compared to the untransformed data.
- Coverage appears to become worse with increasing *K*, and more so for scenarios where heterogeneity is large. This is more evident for expected counts between 1 and 5.

> We have found no explanation for this pattern.

Figure 9. Comparison of coverage between approximate methods: no transformation versus variance stabilizing (arcsine) transformation



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

Canonical (Logit) Versus Variance-Stabilizing (Arcsine) Transformation

Figure 10 and Figure 11 compare proportion bias and proportion RMSE, respectively for random effects meta-analysis on logit-transformed versus arcsine-transformed data. Many of the observations and explanations below are similar to those made in two sections: No Transformation Versus Canonical Transformation (Logit Transformation) and No Transformation Versus Variance-Stabilizing Transformation (Arcsine Transformation).

- For small expected counts (≤1), analyses based on arcsine transformed data have much smaller proportion bias and proportion RMSE compared to those based on logit-transformed data. For larger expected counts, the differences are very small (most points in Figure 10 and Figure 11 are on the diagonal).

> Analyses based on logit-transformed data require continuity corrections (and most often when the expected counts are ≤1). As discussed in No Transformation Versus Canonical Transformation (Logit Transformation), the net effect is a large positive bias and a large proportion RMSE. However, for arcsine transformed data (No Transformation Versus Variance-Stabilizing Transformation [Arcsine Transformation]), no continuity corrections are needed, and thus the proportion bias and proportion RMSE are smaller.

- For both methods, the proportion bias and the proportion RMSE do not change dramatically with the number of studies, *K*.

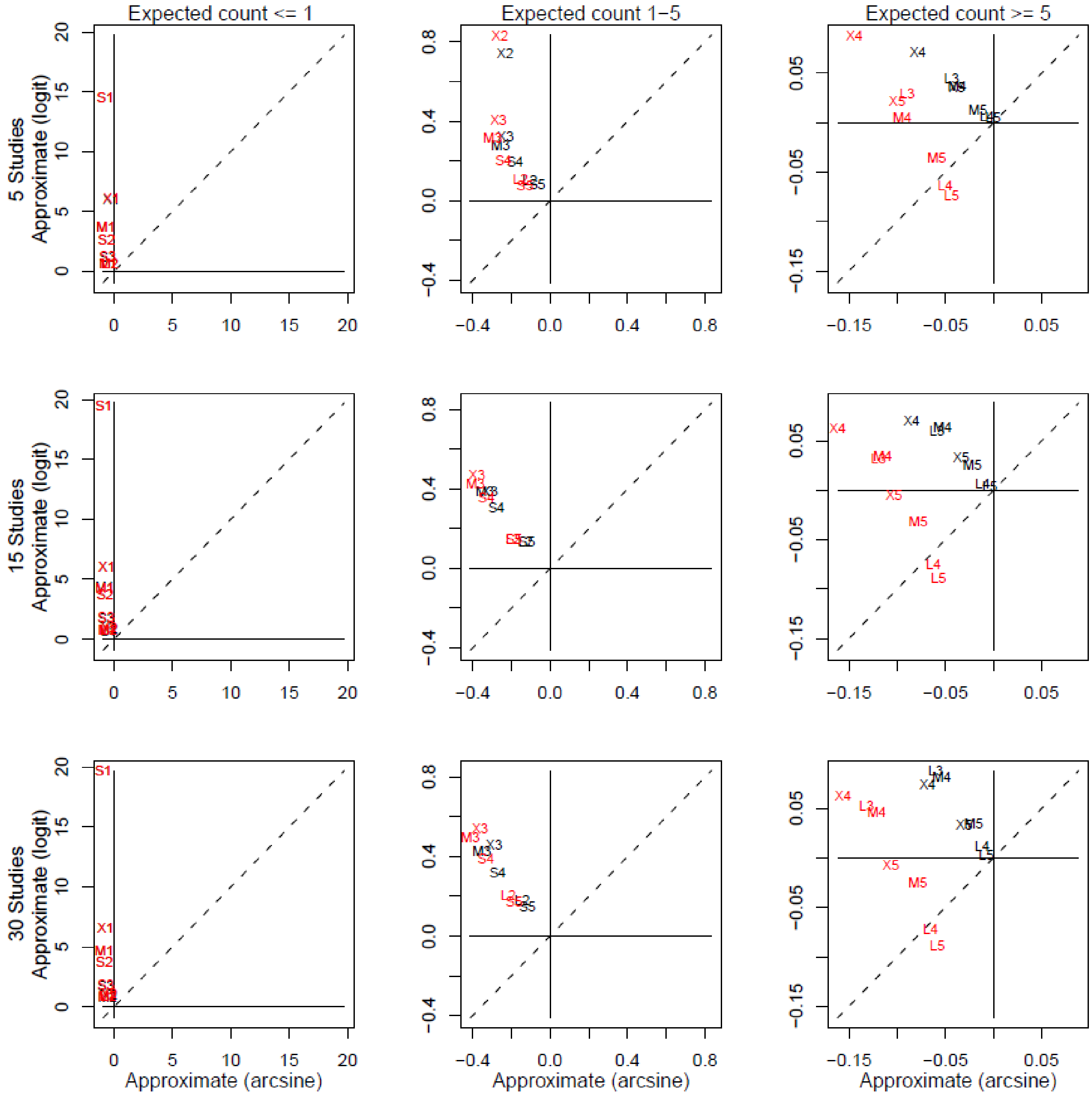
> The influence of the number of studies on proportion bias (and the proportion RMSE) is small, as in previous sections.

- For expected counts less than 1, proportion bias and proportion RMSE are not very different between scenarios with smaller and larger heterogeneity. All other things being equal, differences in proportion RMSE between smaller and larger heterogeneity scenarios are evident for expected counts larger than 1 (Figure 12; and similarly for the absolute value of proportion bias—not shown).

> For small expected counts, the estimate of between study heterogeneity is very often 0 or close to 0, and this attenuates any differences between scenarios with smaller and larger simulated heterogeneity.

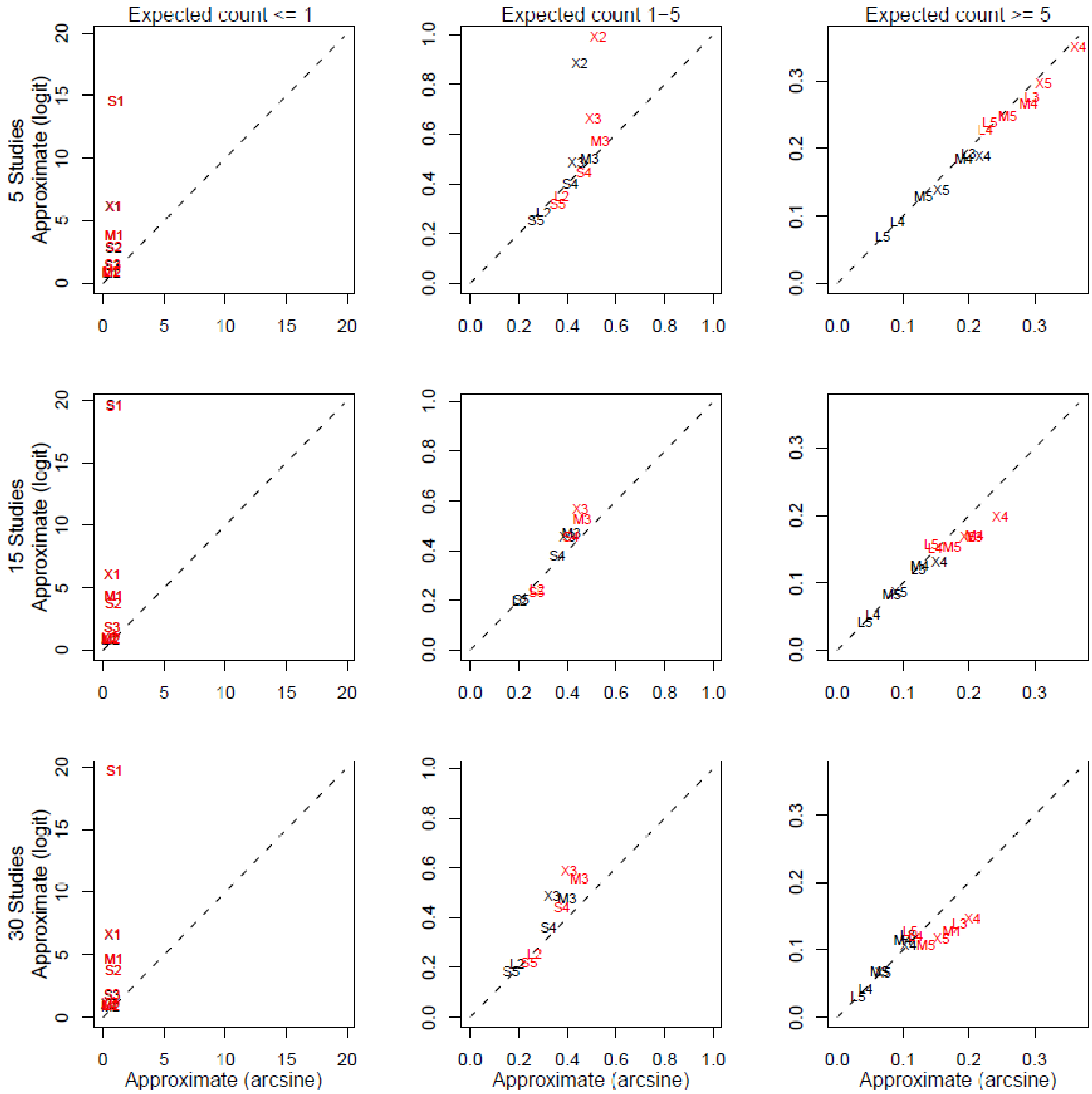
- Analytically, we expect the proportion bias to be negative for analyses based on arcsine-transformed data, and positive for analyses based on logit-transformed data (See appendix). As described in No Transformation Versus Canonical Transformation (Logit Transformation), the need for continuity corrections (*cc*) adds an additional bias component for analyses based on logit-transformed data. This additional bias component can be positive, zero, or negative, for $\frac{cc}{n_{j,k}} > \pi_j$, $\frac{cc}{n_{j,k}} = \pi_j$, or $\frac{cc}{n_{j,k}} < \pi_j$, respectively.

Figure 10. Comparison of proportion bias between approximate methods: canonical (logit) versus variance stabilizing (arcsine) transformation



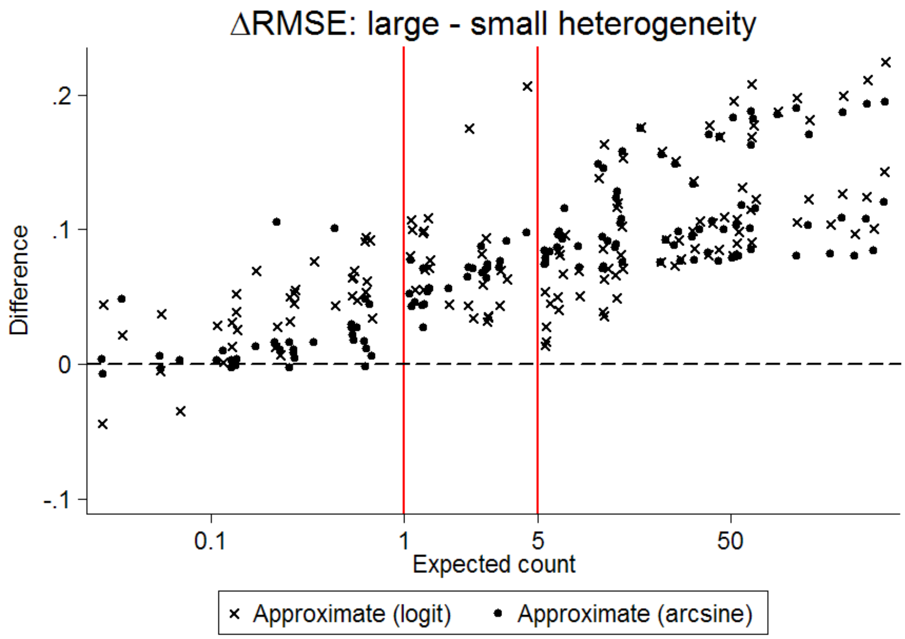
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 11. Comparison of proportion RMSE between approximate methods: canonical (logit) versus variance stabilizing (arcsine) transformation



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 12. Difference in RMSE between small and large heterogeneity scenarios: canonical (logit) versus variance stabilizing (arcsine) transformation



Shown are all simulation scenarios (not only the "representative" ones listed in Table 9). A similar pattern is observed for the absolute value of proportion bias. The horizontal dotted line at zero is the line of no difference. Vertical lines separate scenarios by expected counts categories.

For both methods, coverage is better when the expected count is at least 5 (Figure 13), compared to expected counts of 1 or less, or between 1 and 5.

> The normal distribution approximates the binomial better as expected counts increase above 5.

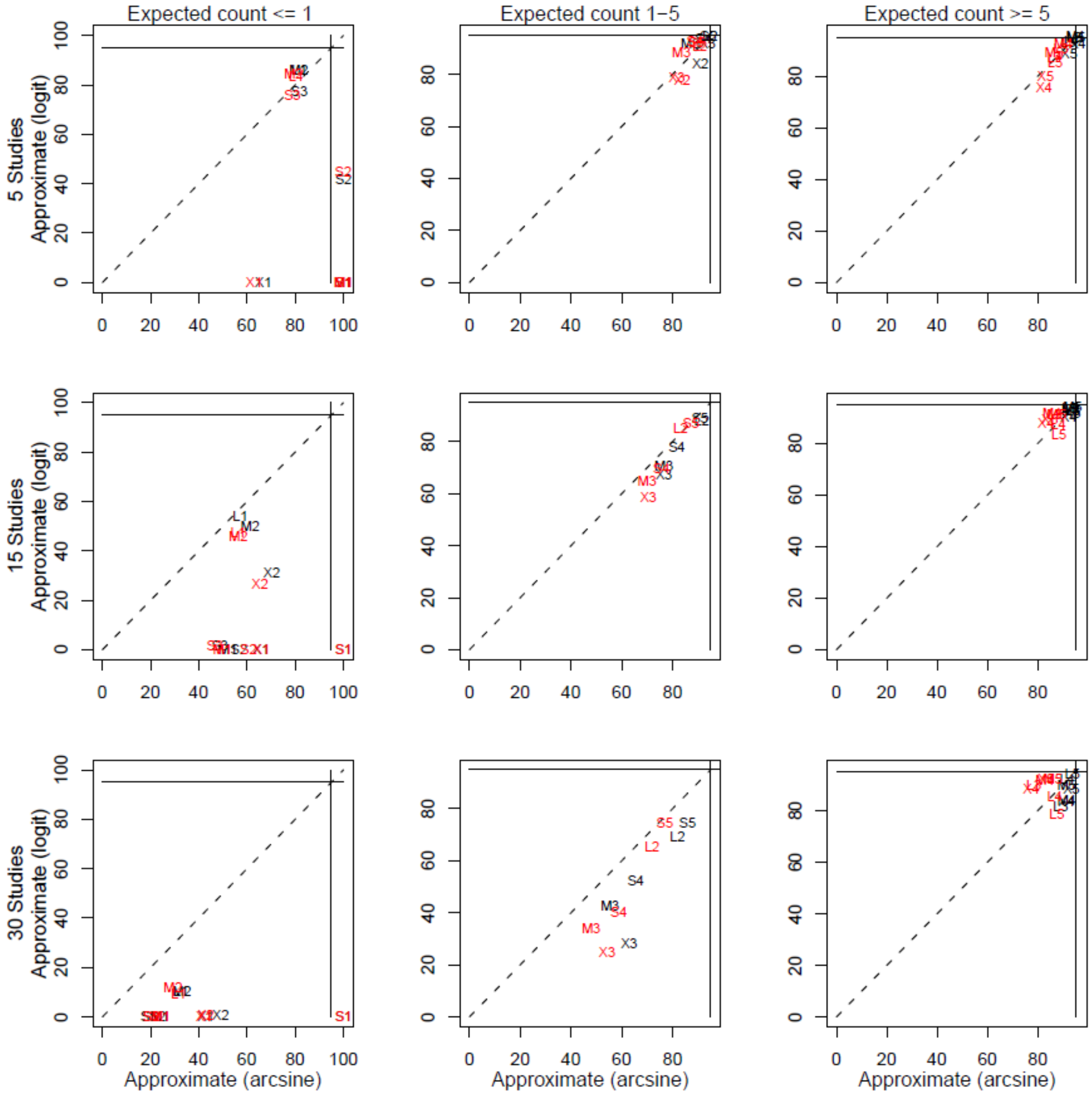
- For expected counts ≥5, coverage is similar for both methods.
- For expected counts ≤5, analyses based on the arcsine transformation have better coverage.

> For small expected counts, continuity corrections are often needed for logit-transformed data. The large positive bias (associated with the use of the continuity corrections) can displace the point estimate sufficiently to result in 0 percent coverage.

- For expected counts between 1 and 5, coverage appears to become worse with increasing *K*, and more so for scenarios where heterogeneity is large.

> We have found no explanation for this pattern.

Figure 13. Comparison of coverage between approximate methods: canonical (logit) versus variance stabilizing (arcsine) transformation



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

Effect of the Correction Factor for Zero-Event Studies

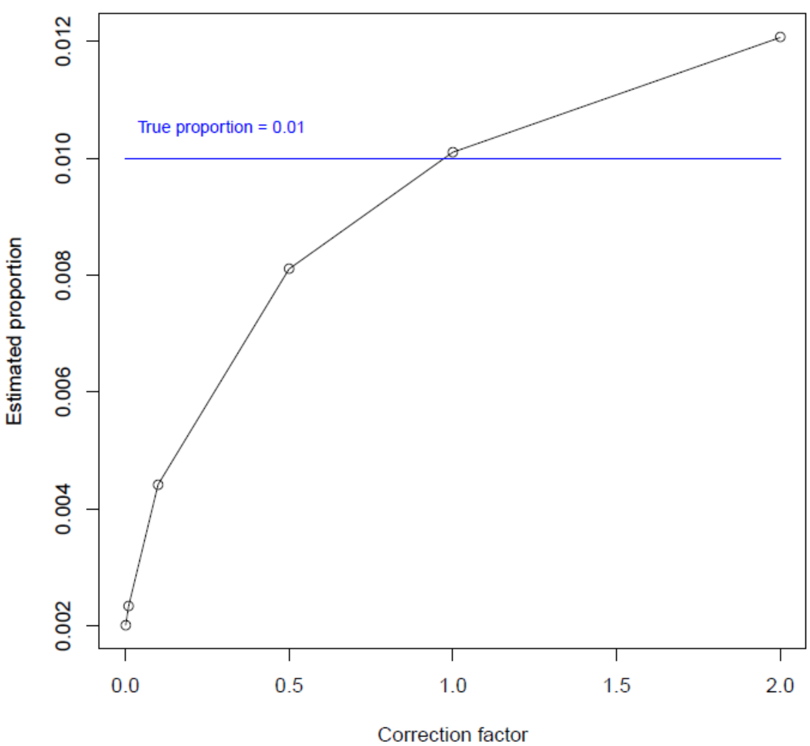
The approximate methods for untransformed and logit-transformed data apply a correction factor to adjust the estimated proportion for zero-event studies. The correction factor is arbitrary, in that it is not selected with a procedure that proposes a unique value. We report results using a continuity correction of 0.5, a very commonly used value.

As discussed in the Methods section, the correction factor introduces a bias that depends on the size of the correction factor and the sample size of the study.

Figure 14 shows the mean summary proportion for fixed effects inverse variance meta-analyses with 15 studies with average sample size equal to 115 for various values of the correction factor, assuming a true proportion of 0.010 and no between-study heterogeneity. Approximately 1/3 of the simulated studies have 0 events. Note the estimated proportion is less than the true proportion for small values of the correction factor, and increases as the correction factor gets larger. For some value of the correction factor (in this case 1), the estimated proportion will be exactly the true proportion, but this is of no use in practice since the “optimal” value of the correction factor depends on the true proportion and sample sizes.

Since the correction factor is unrelated to the true proportion, the overall meta-analysis estimate is skewed towards the adjusted estimate, resulting in greater absolute bias and RMSE, and poorer coverage. This typically occurs for small true proportions and sample sizes.

Figure 14. Bias induced by correction factors in fixed effects meta-analysis of untransformed data



Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis

For random effects meta-analysis, results based on the binomial likelihood (exact method) have proportion bias and proportion RMSE closer to zero and coverage probabilities closer to 95 percent, compared to those based on the approximate methods across a wide range of scenarios, as explained below.

Therefore, we recommend the discrete likelihood method for random effects meta-analysis over approximate methods.

Approximate Method With No Transformation Versus Discrete Likelihood Method

Figure 15 compares the proportion bias for random effects meta-analysis using the approximate method with untransformed data versus the discrete likelihood method that maximizes the binomial likelihood.

Figure 16 depicts proportion RMSE. Based on the two figures, we make the following observations:

- The absolute proportion bias (deviation from 0) is generally smaller for the discrete likelihood method compared to the approximate method using untransformed data, particularly when expected counts ≤1.

> Meta-analyses with approximate methods and untransformed data require continuity corrections (and most often when the expected counts are ≤1). As discussed in No Transformation Versus Canonical Transformation (Logit Transformation), the net effect is a large positive bias and a large proportion RMSE. Continuity corrections are not needed for the discrete likelihood method.

- For meta-analyses with 15 or 30 studies and expected counts larger than 1, the proportion bias of the discrete likelihood binomial method is approximately constant at each level of heterogeneity. For large heterogeneity, the proportion bias is approximately -10 percent of the respective true proportion, while for small heterogeneity, it is approximately 0. If the true proportion is exactly 0.50, the proportion bias is approximately 0.

> We have no definitive explanation for the approximately constant proportion bias of -10 percent, which is observed across all proportions when heterogeneity is large, and the true summary proportion is different than 0.50.
> A plausible conjecture is based on the observation that the random effects discrete likelihood methods assume a logit-normal distribution of the true proportions across studies. A range of e.g., 0.10 in the proportion scale corresponds to vastly different ranges in the logit scale, depending on its location. For example the interval [0.01, 0.11] in the proportion scale corresponds to a length of [-4.60, -2.09] in the logit scale, but the interval [0.45, 0.55] corresponds to the much narrower [-0.20, 0.20] in the logit scale. The further away a given interval is located from 0.50 in the proportion scale, the more it expands in the logit scale. "Averaging" in the logit scale will therefore result in a negative bias in the proportion scale when the true mean is less than 0.50; no bias if the true mean is 0.50, and (by symmetry) a positive bias if the true mean is more than 0.50.

- When all studies have large sample sizes and the heterogeneity is large, the approximate method with untransformed data has proportion bias closer to 0 compared to the discrete likelihood method. (See points L4 and L5 in red font on the rightmost column in Figure 15).

> This occurs because for large heterogeneity and expected counts ≥5 the discrete likelihood methods have a proportion bias of approximately -10 percent, regardless of the sample sizes of the studies. In these same scenarios, the normal approximation to the binomial is good enough to result in better proportion bias for the approximate compared to the discrete likelihood methods.

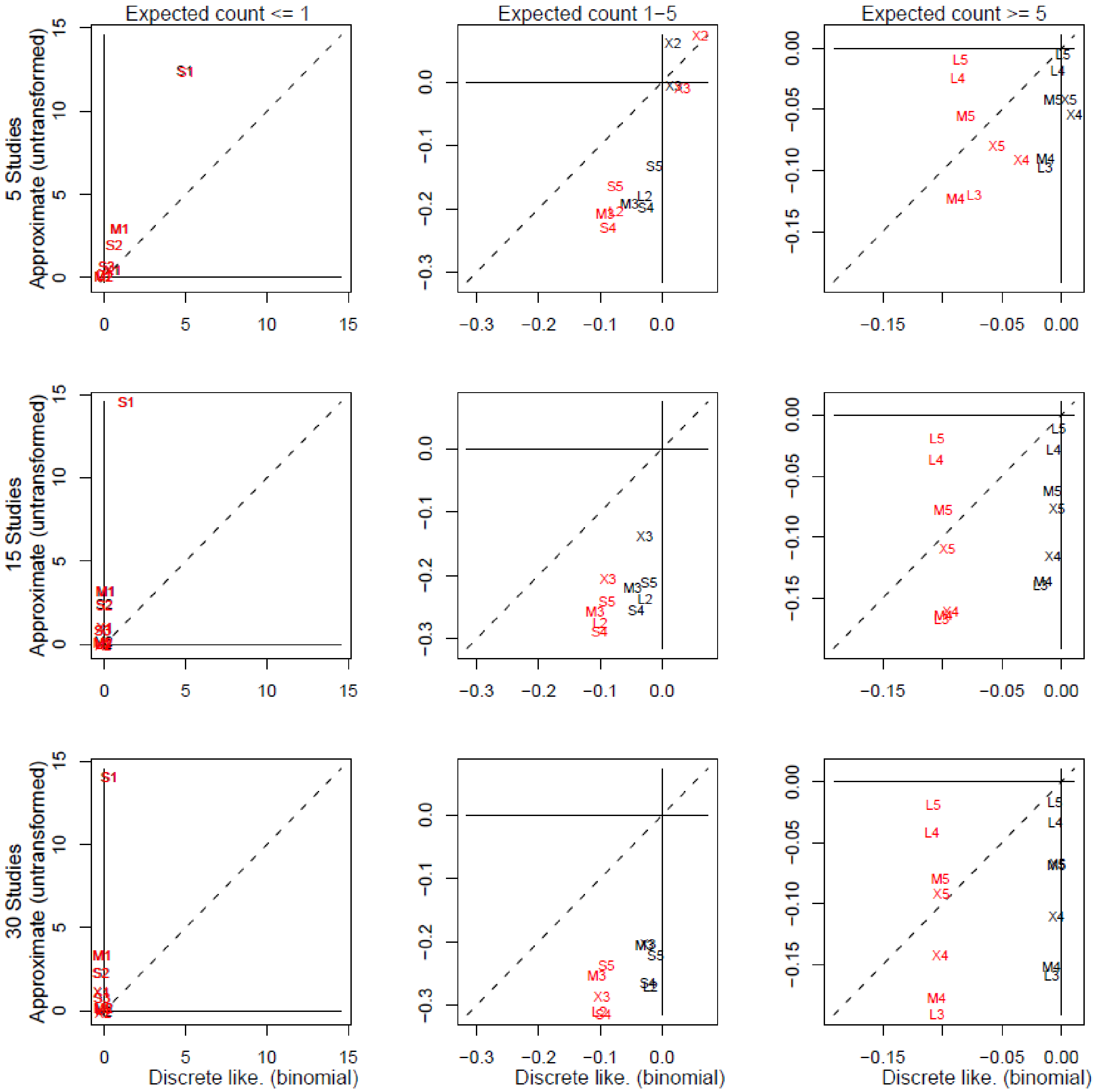
- For expected counts ≤1, primarily, the proportion bias of both the discrete likelihood and the approximate methods is positive. For larger expected counts the proportion bias for both methods becomes negative.

> For the approximate methods, the explanation has been given in No Transformation Versus Canonical Transformation (Logit Transformation): Analytically we expect the bias to be negative. However, when the expected count is ≤1, studies will often have 0 events, and a continuity correction factor, *cc*, is needed. This additional bias component can be positive, zero, or negative, for $\frac{cc}{n_{j,k}} > \pi_j$, $\frac{cc}{n_{j,k}} = \pi_j$, or $\frac{cc}{n_{j,k}} < \pi_j$, respectively.
>
> For the discrete likelihood method, the explanation is much simpler. When all studies have a 0 numerator, the discrete likelihood method (random effects logistic regression) fails to converge. Because simulations that fail to converge do not contribute to the calculations of proportion bias, proportion RMSE and coverage, these metrics should be interpreted with caution.

- Proportion RMSE is generally lower for the discrete likelihood method. The differences are largest for small expected counts. For expected counts less than 1, proportion bias and proportion RMSE are not very different between scenarios with smaller and larger heterogeneity. All other things being equal, differences in proportion RMSE between smaller and larger heterogeneity scenarios are evident for expected counts larger than 1 (Figure 17; and similarly of the absolute value of proportion bias—not shown).
- Proportion RMSE is larger for large amounts of heterogeneity.

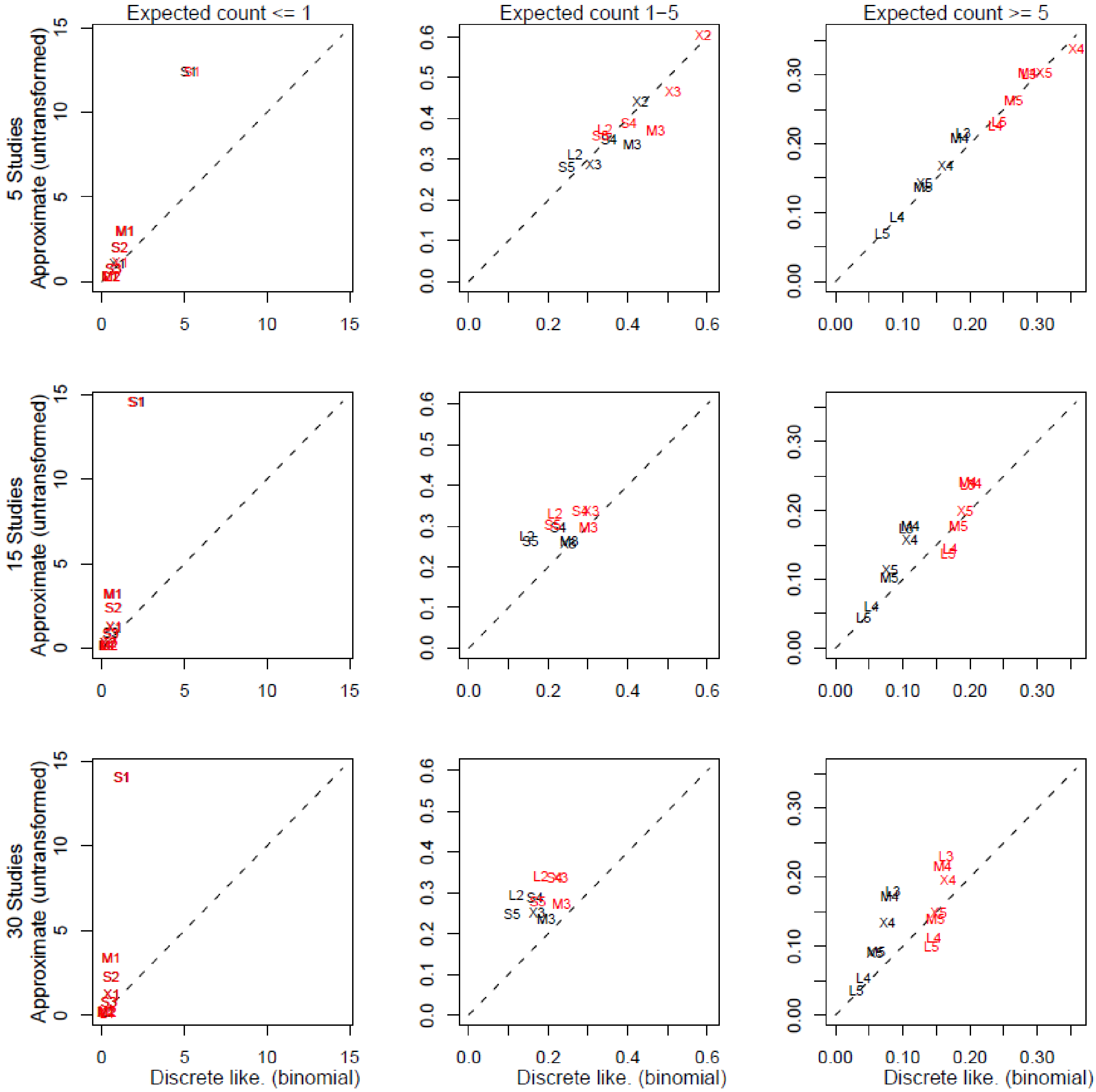
> Large amounts of heterogeneity increases variance, one of the components of mean squared error. When there is a lot of data, variability decreases and bias tends to dominate the mean squared error.

Figure 15. Comparison of proportion bias: approximate method (no transformation) versus discrete likelihood



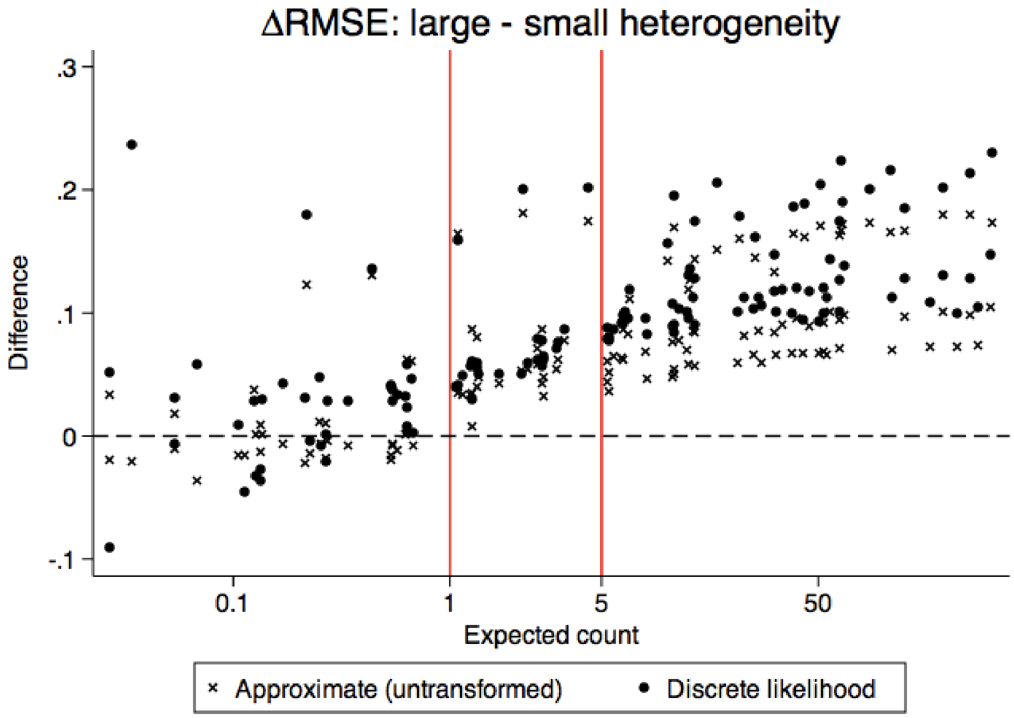
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 16. Comparison of proportion RMSE: approximate method (no transformation) versus discrete likelihood



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 17. Difference in RMSE between small and large heterogeneity scenarios for analyses using the approximate method (no transformation) and the discrete likelihood method

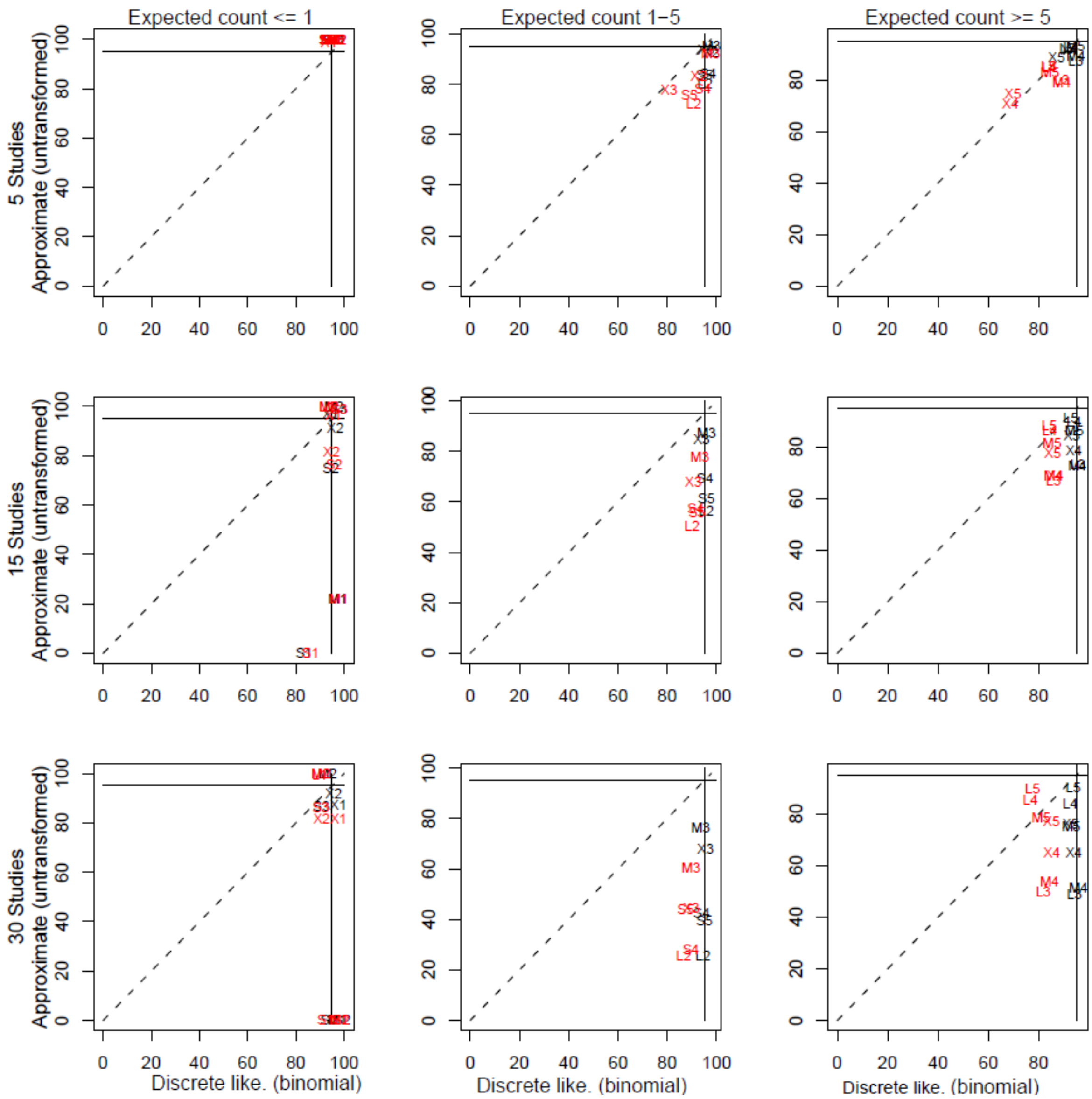


Shown are all simulation scenarios (not only the "representative" ones listed in Table 9). A similar pattern is observed for the absolute value of proportion bias. The horizontal dotted line at zero is the line of no difference. Vertical lines separate scenarios by expected counts categories.

- The coverage of the discrete likelihood binomial method (Figure 18) is much closer to the desired value of 95 percent than that of the approximate method on untransformed data. This is particularly noticeable for K=15 and 30 studies.

The normal approximation to the binomial is much better for large expected counts compared to smaller expected counts. Further, for small expected counts, numerators with zero counts are common and continuity corrections are needed for the approximate method. Thus, coverage is suboptimal with the approximate method. See these two sections: No Transformation Versus Canonical Transformation (Logit Transformation) and No Transformation Versus Variance-Stabilizing Transformation (Arcsine Transformation).

Figure 18. Comparison of coverage: approximate method (no transformation) versus discrete likelihood



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

Approximate Method With Logit Transformation Versus Discrete Likelihood Method

Figure 19 compares proportion bias for random-effects meta-analysis using the approximate method on logit-transformed data versus the discrete likelihood binomial method.

Figure 20 depicts the proportion RMSE. We make the following observations:

- The absolute proportion bias (deviation from 0) is generally smaller for the discrete likelihood method compared to the approximate method on logit transformed data. This is more profound for expected counts ≤5.

Meta-analyses with approximate methods and untransformed data require continuity corrections (and most often when the expected counts are ≤1). As discussed in No Transformation Versus Canonical Transformation (Logit Transformation), the net effect is a large positive bias and a large proportion RMSE. Continuity corrections are not needed for the discrete likelihood method.

- As discussed in Approximate Method With No Transformation Versus Discrete Likelihood Method, for meta-analyses with 15 or 30 studies and expected counts larger than 1, the proportion bias of the discrete likelihood binomial method is approximately constant at each level of heterogeneity. For large heterogeneity, the proportion bias is approximately -10 percent of the respective true proportion, while for small heterogeneity, it is approximately 0. If the true proportion is exactly 0.50, the proportion bias is approximately 0.

We have no definitive explanation for the approximately constant proportion bias of -10 percent, which is observed across all proportions when heterogeneity is large, and the true summary proportion is different than 0.50.
A plausible conjecture is based on the observation that the random effects discrete likelihood methods assume a logit-normal distribution of the true proportions across studies. A range of e.g., 0.10 in the proportion scale corresponds to vastly different ranges in the logit scale, depending on its location. For example the interval [0.01, 0.11] in the proportion scale corresponds to a length of [-4.60, -2.09] in the logit scale, but the interval [0.45, 0.55] corresponds to the much narrower [-0.20, 0.20] in the logit scale. The further away a given interval is located from 0.50 in the proportion scale, the more it expands in the logit scale. "Averaging" in the logit scale will therefore result in a negative bias in the proportion scale when the true mean is less than 0.50; no bias if the true mean is 0.50, and (by symmetry) a positive bias if the true mean is more than 0.50. The finding that for large expected counts (scenarios L4 and L5) the logit-transformed bias approaches that of the discrete likelihood method supports this conjecture.

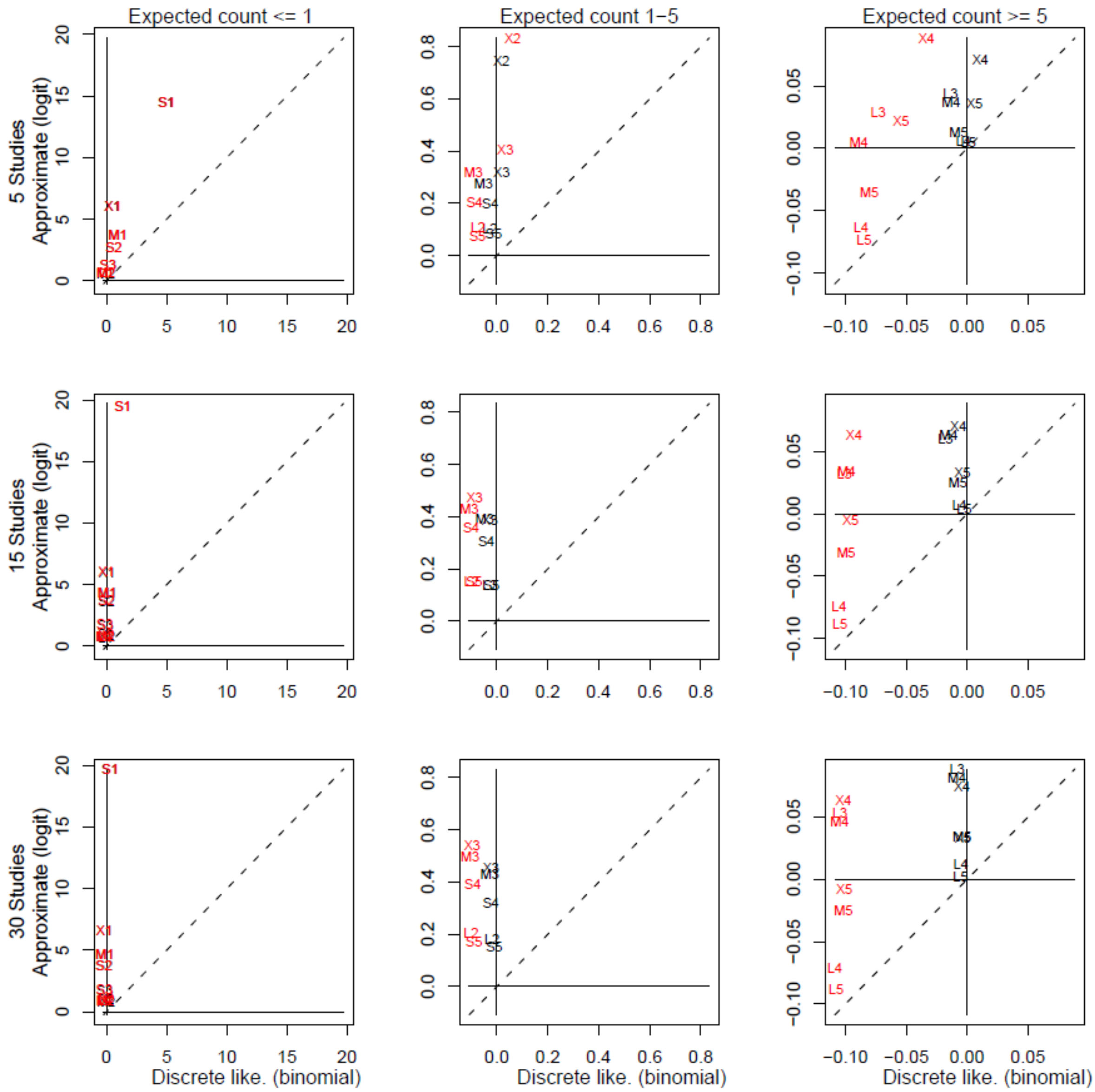
- For expected counts ≤1, primarily, the proportion bias for the discrete likelihood method is positive. It becomes negative for larger expected counts.

As described in Approximate Method With No Transformation Versus Discrete Likelihood Method, when all studies have a 0 numerator, the discrete likelihood method (random effects logistic regression) fails to converge. Because simulations that fail to converge do not contribute to the calculations of proportion bias, proportion RMSE and coverage, these metrics are biased.

- Proportion RMSE is generally lower for the discrete likelihood method. The differences are largest for small expected counts and small when expected counts ≥ 5.

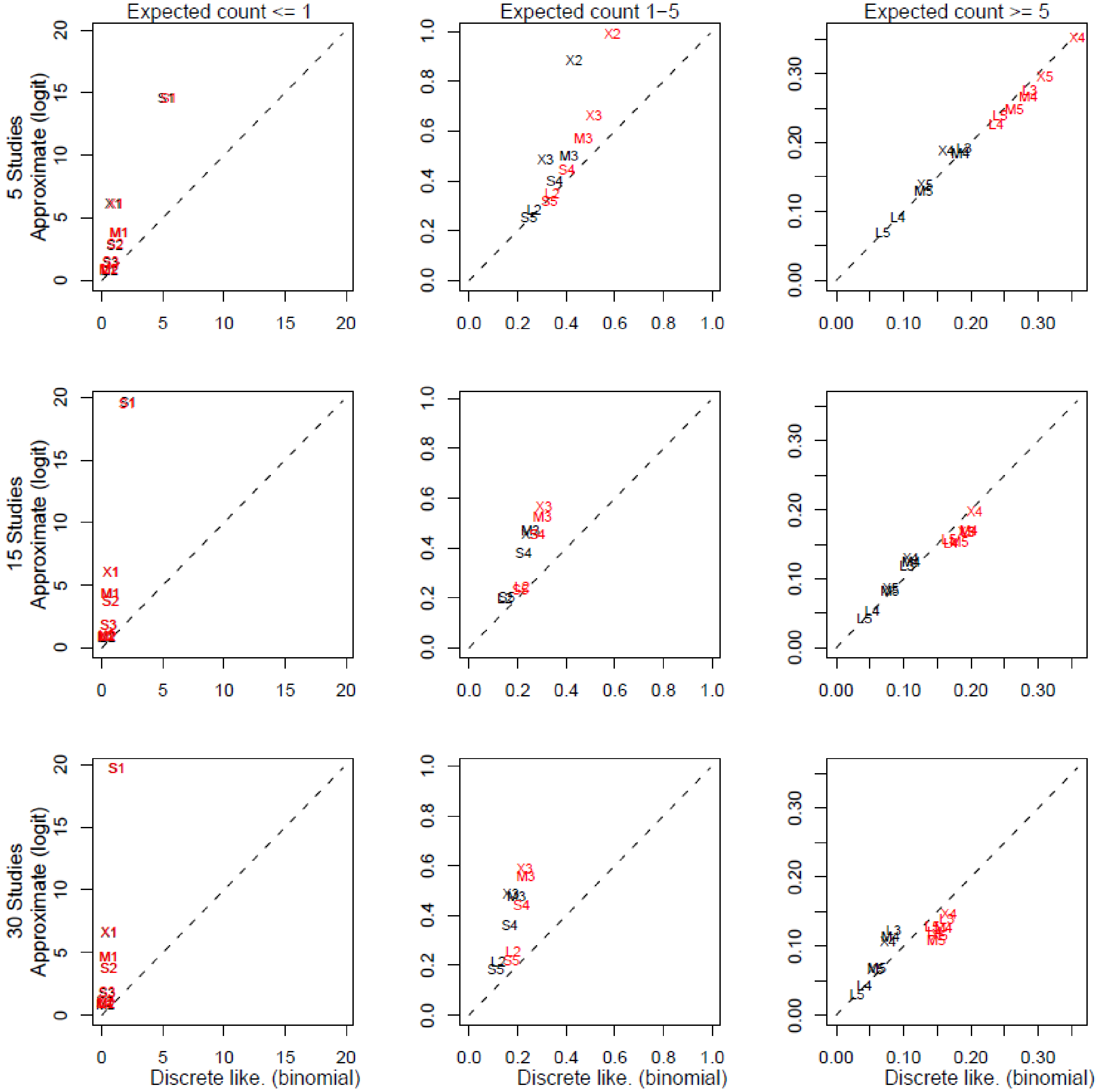
For expected counts less than 1, the proportion bias and the proportion RMSE are not very different between scenarios with smaller and larger heterogeneity. All other things being equal, differences in proportion RMSE between smaller and larger heterogeneity scenarios are evident for expected counts larger than 1 (Figure 21; and similarly for the absolute value of proportion bias—not shown).

Figure 19. Comparison of proportion bias: approximate method (logit transformation) versus discrete likelihood



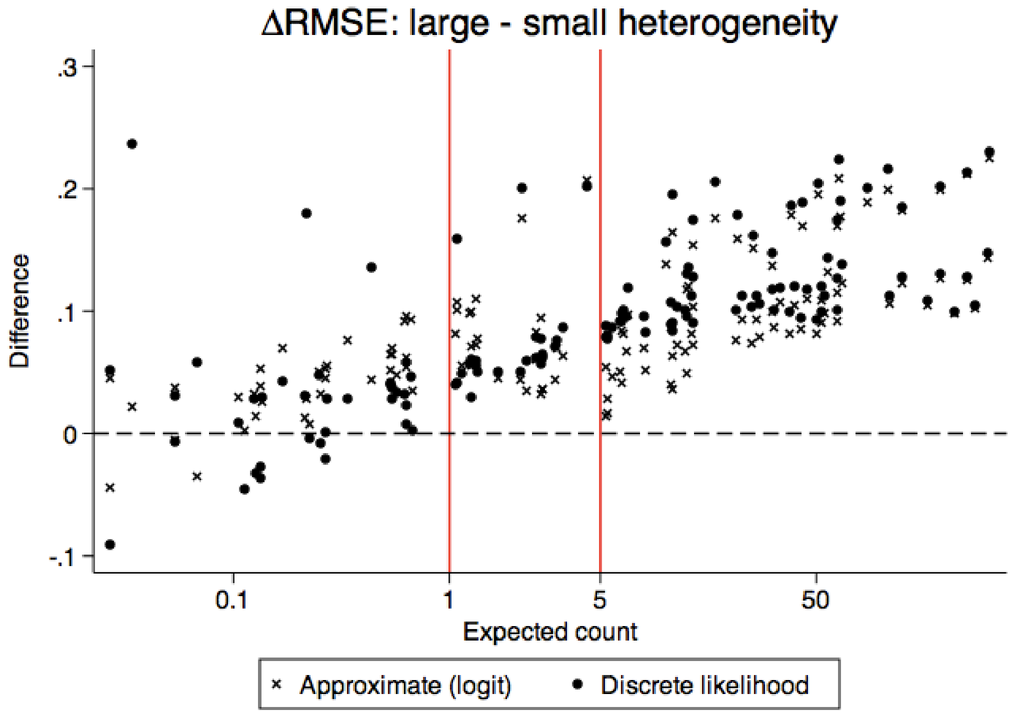
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 20. Comparison of proportion RMSE: approximate method (logit transformation) versus discrete likelihood



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 21. Difference in RMSE between small and large heterogeneity scenarios for analyses using the approximate methods (logit transformation) and the discrete likelihood method



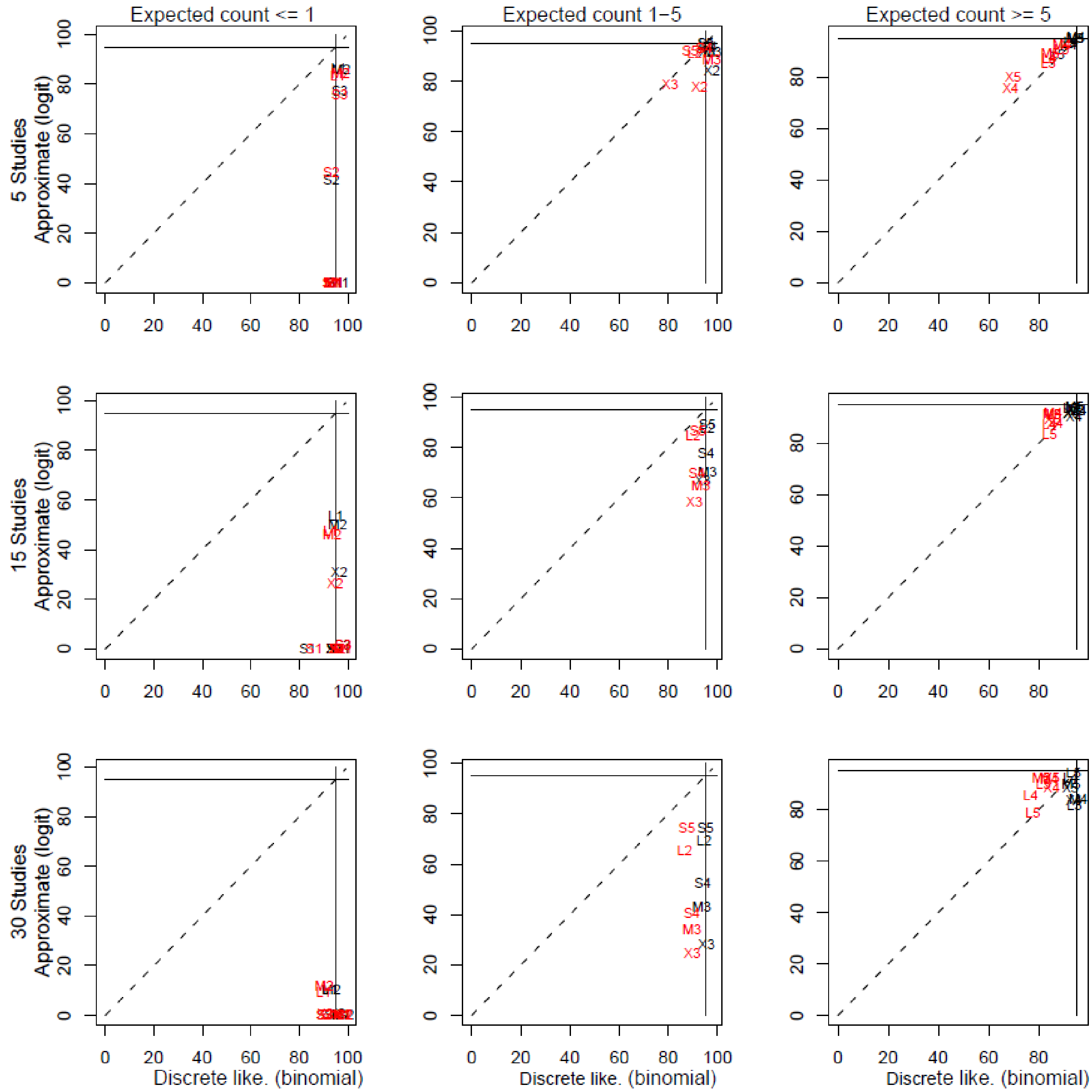
Shown are all simulation scenarios (not only the "representative" ones listed in Table 9). A similar pattern is observed for the absolute value of proportion bias. The horizontal dotted line at zero is the line of no difference. Vertical lines separate scenarios by expected counts categories.

Figure 22 shows coverage probabilities.

- The coverage of the discrete likelihood binomial method is much closer to the desired value of 95 percent than that of the approximate method on logit-transformed data. This is particularly noticeable for K=15 and 30 studies.

> The normal approximation to the binomial is much better for large expected counts compared to smaller expected counts. Further, for small expected counts, 0 numerators are common and continuity corrections are needed for the approximate method. Thus, coverage is suboptimal with the approximate method. See also two sections: No Transformation Versus Canonical Transformation (Logit Transformation) and No Transformation Versus Variance-Stabilizing Transformation (Arcsine Transformation).

Figure 22. Comparison of coverage: approximate method (logit transformation) versus discrete likelihood



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

Approximate Method With Arcsine Transformation Versus Discrete Likelihood Binomial Method

Figure 23 compares proportion bias for random-effects meta-analysis using the approximate method with the variance stabilizing (arcsine) transformation versus the discrete likelihood method. Figure 24 depicts the proportion RMSE. We make the following observations:

- The absolute proportion bias (deviation from 0) is generally smaller for the discrete likelihood method compared to the approximate method on arcsine transformed data. This is more pronounced for expected counts ≤5.

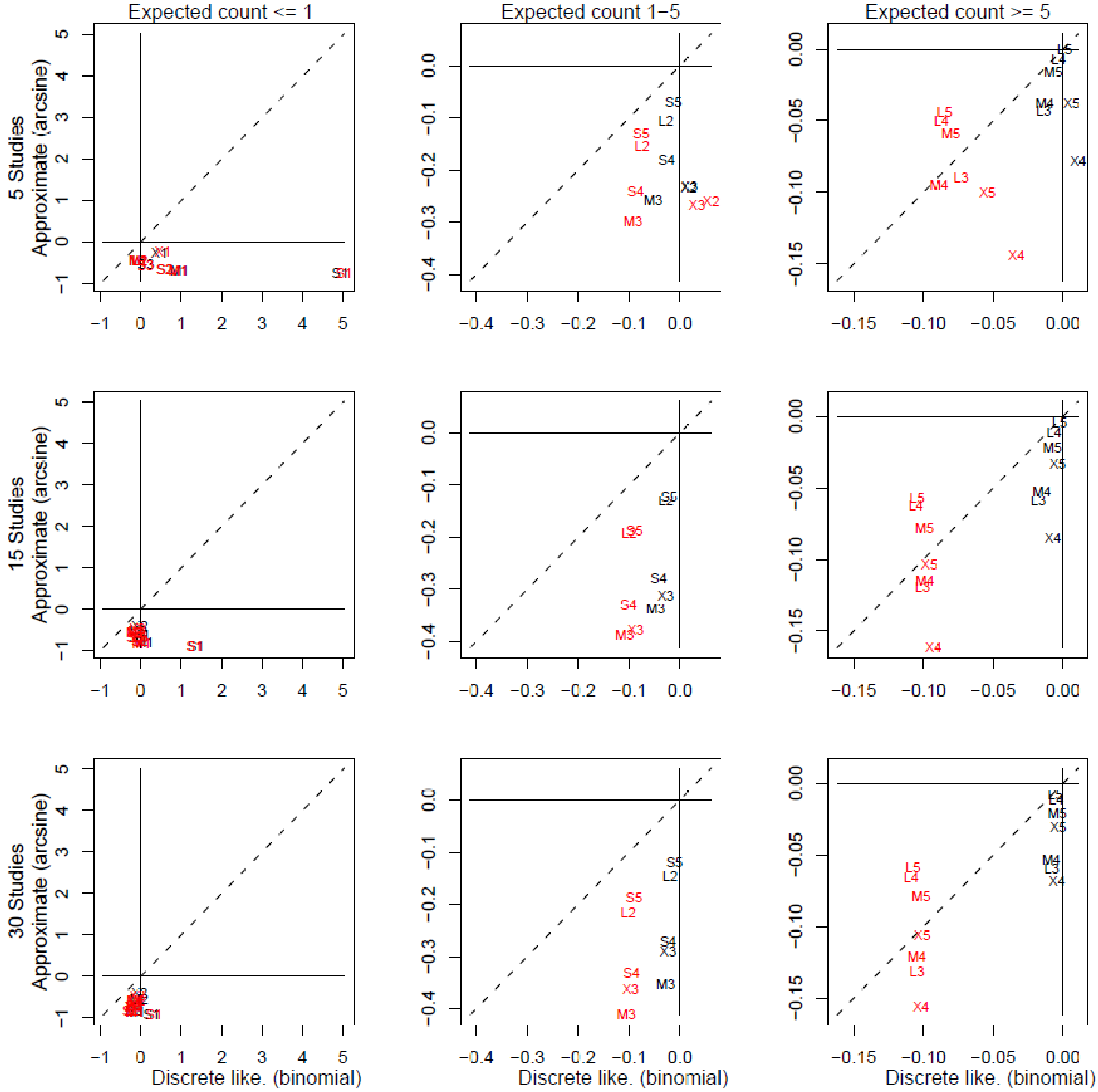
> The normal approximation to the binomial becomes better when the expected counts are >5 compared with smaller expected counts.

- As discussed in Approximate Method With No Transformation Versus Discrete Likelihood Method, for meta-analyses with 15 or 30 studies and expected counts larger than 1, the proportion bias of the discrete likelihood binomial method is approximately constant at each level of heterogeneity. For large heterogeneity, the proportion bias is approximately -10 percent of the respective true proportion, while for small heterogeneity, it is approximately 0. If the true proportion is exactly 0.50, the proportion bias is approximately 0.

> We have no definitive explanation for the approximately constant proportion bias of -10 percent, which is observed across all proportions when heterogeneity is large, and the true summary proportion is different than 0.50.
> A plausible conjecture is based on the observation that the random effects discrete likelihood methods assume a logit-normal distribution of the true proportions across studies. A range of e.g., 0.10 in the proportion scale corresponds to vastly different ranges in the logit scale, depending on its location. For example the interval [0.01, 0.11] in the proportion scale corresponds to a length of [-4.60, -2.09] in the logit scale, but the interval [0.45, 0.55] corresponds to the much narrower [-0.20, 0.20] in the logit scale. The further away a given interval is located from 0.50 in the proportion scale, the more it expands in the logit scale. “Averaging” in the logit scale will therefore result in a negative bias in the proportion scale when the true mean is less than 0.50; no bias if the true mean is 0.50, and (by symmetry) a positive bias if the true mean is more than 0.50. The finding that for large expected counts (scenarios L4 and L5) the logit-transformed bias approaches that of the discrete likelihood method supports this conjecture.

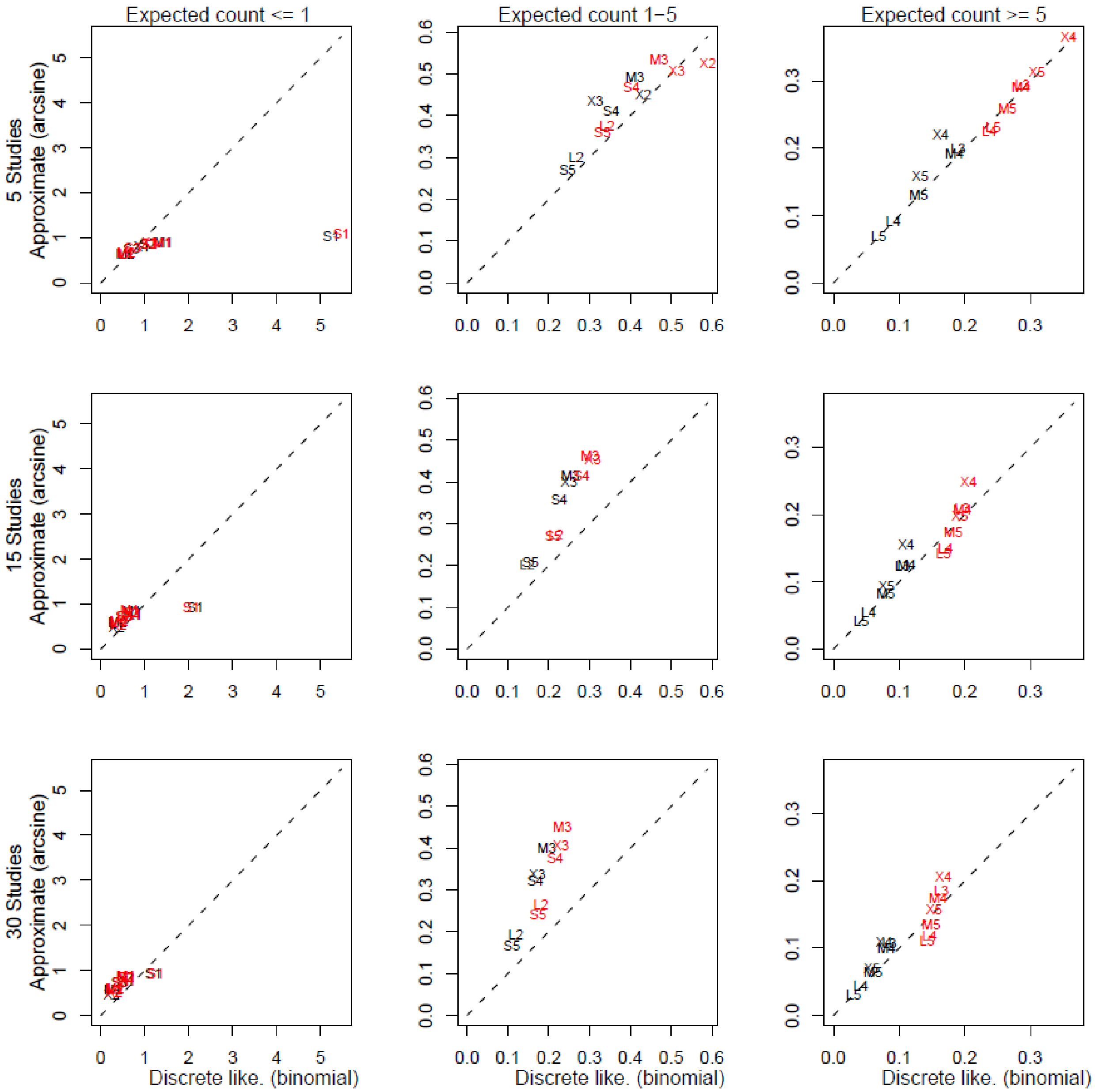
- RMSE is higher with the arcsine method for expected counts between 1 and 5.
- For expected counts less than 1, the proportion bias and the proportion RMSE differ little between scenarios with smaller and larger heterogeneity.
- All other things being equal, differences in proportion RMSE between smaller and larger heterogeneity scenarios are evident for expected counts larger than 1 (Figure 25; and similarly of the absolute value of proportion bias—not shown). RMSE is larger with more heterogeneity.

Figure 23. Comparison of proportion bias: approximate method (arcsine transformation) versus discrete likelihood



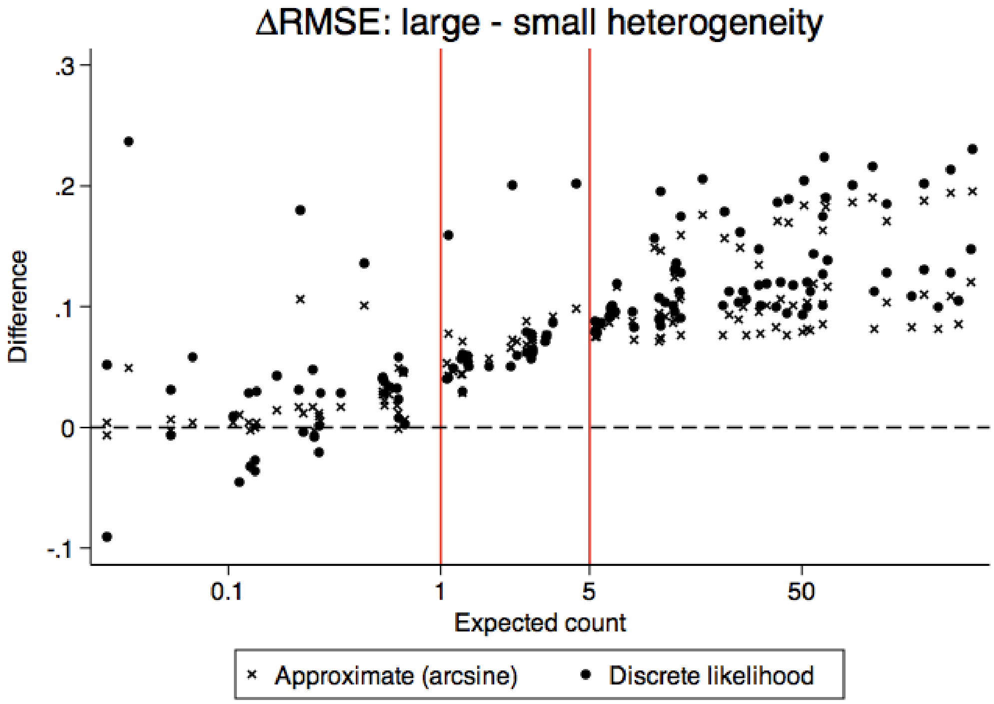
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 24. Comparison of proportion RMSE: approximate method (arcsine transformation) versus discrete likelihood method



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 25. Difference in RMSE between small and large heterogeneity scenarios for analyses using approximate methods (arcsine transformation) and the discrete likelihood method

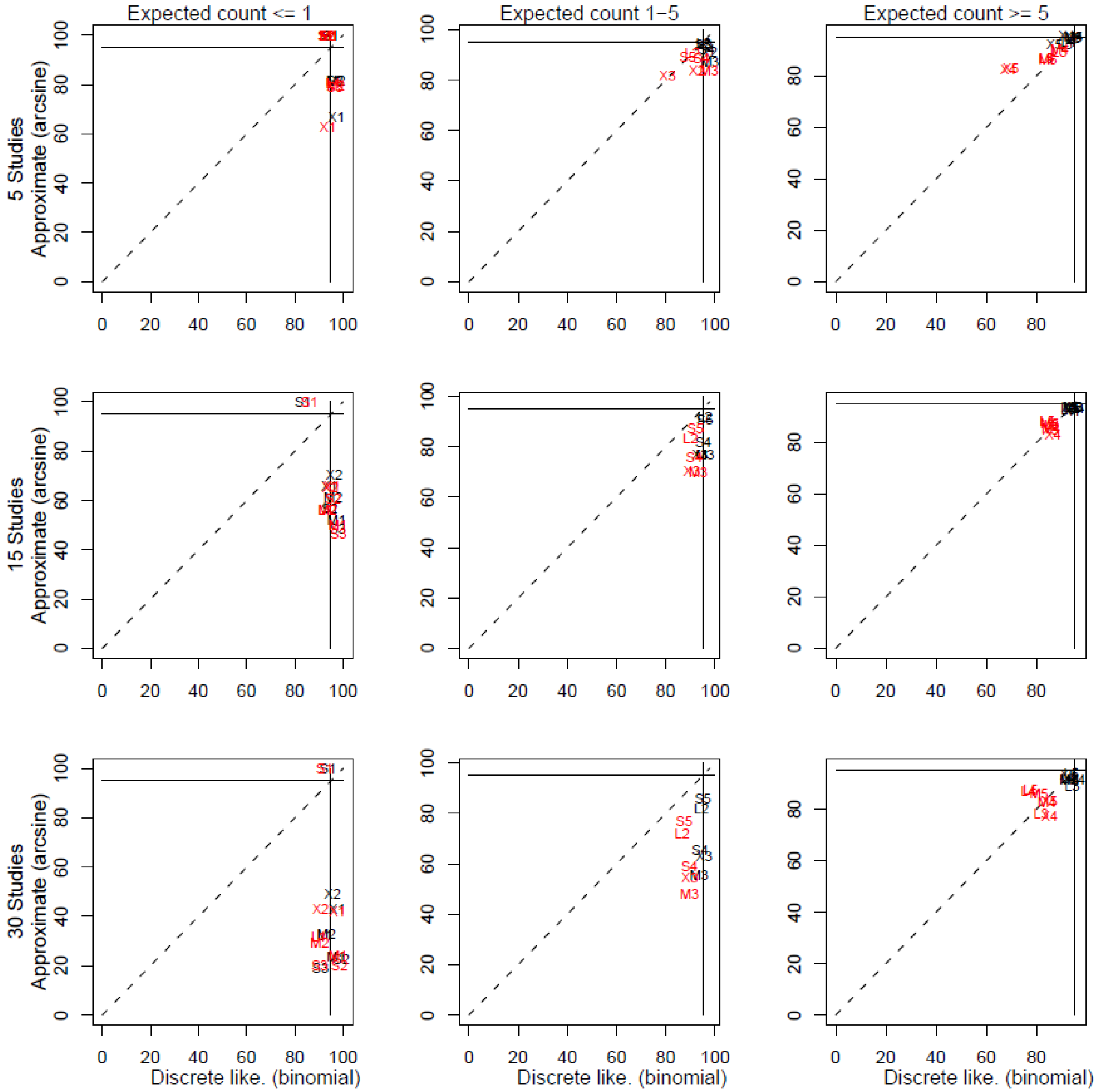


Shown are all simulation scenarios (not only the "representative" ones listed in Table 9). A similar pattern is observed for the absolute value of proportion bias. The horizontal dotted line at zero is the line of no difference. Vertical lines separate scenarios by expected counts categories.

Figure 26 shows coverage probabilities.

- For expected counts above 5, the coverage probabilities are very similar for the approximate methods on arcsine-transformed data and the discrete likelihood methods.
- Coverage is better under scenarios with less heterogeneity
- With large amounts of heterogeneity, both methods have inadequate coverage
- For expected counts less than 5, the coverage of the discrete likelihood binomial method is much closer to the desired value of 95 percent than that of the approximate method on arcsine-transformed data.

Figure 26. Comparison of coverage: approximate method (arcsine transformation) versus discrete likelihood method



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

Comparison Between Fixed and Random-Effects Discrete Likelihood Binomial Methods

Most meta-analyses use random effects methods. For random effects meta-analysis, discrete likelihood methods have less bias, smaller RMSE, and better coverage probabilities compared to approximate methods for a large range of simulation scenarios (See two sections: Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis and Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis), and are thus preferable.

Here we present comparisons between fixed and random effects meta-analysis using the discrete likelihood methods. Figure 27 compares the proportion bias and Figure 28 the proportion RMSE with fixed versus random effects analyses. We make the following observations:

- Results based on the fixed-effect discrete likelihood binomial method have smaller absolute proportion bias than results based on the random-effects discrete likelihood binomial method, particularly for data with high heterogeneity—see Figure 27, the whole left column.

> The fixed effects estimator with the discrete likelihood method is an unbiased estimator. Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis discussed the bias observed with the random effects methods: it is constant at each level of heterogeneity, and is small (around 0 percent) for smaller heterogeneity, and approximately -10 percent for larger heterogeneity scenarios when the true proportion is not 0.50.

- The proportion RMSE is approximately the same for the two methods—see Figure 28.

> When the expected counts are small, it common to have meta-analyses where all studies have zero counts in the numerator. In such cases, the discrete likelihood method (random effects logistic regression) fails to converge. Because simulations that fail to converge do not contribute to the calculations of proportion bias, proportion RMSE and coverage, these metrics should be interpreted with caution.

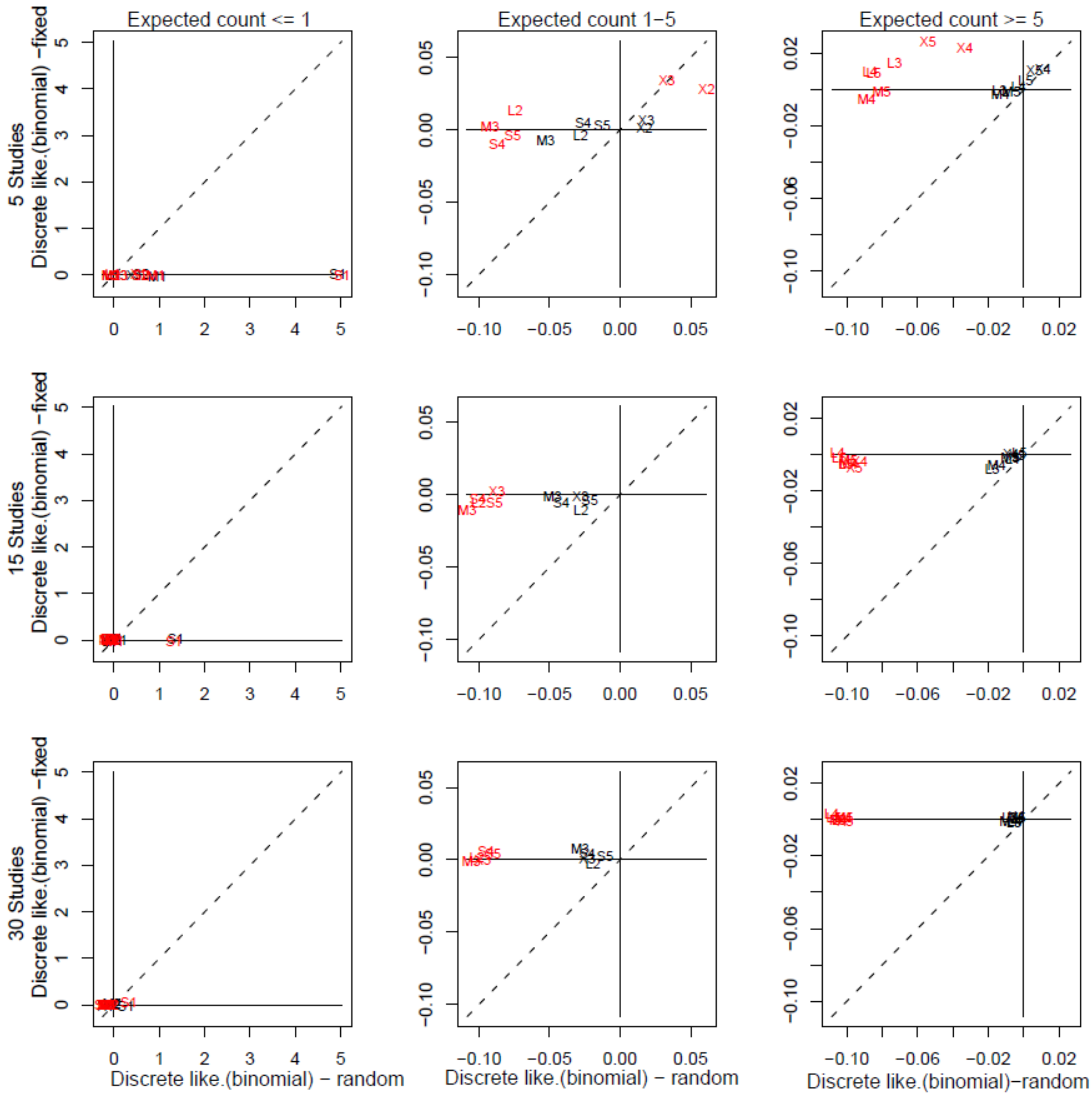
- The random-effects discrete likelihood method has better coverage than the fixed-effect method, particularly for data with expected counts at least 1—see Figure 29.

> Fixed effects analyses assume no heterogeneity. Random effects analyses estimate between-study heterogeneity and incorporate it in the calculations, resulting in wider confidence intervals compared with fixed effects analyses. Therefore, coverage is better with random effects models when the data have heterogeneity.

- The random-effects discrete likelihood method has only about 80 percent coverage when the expected counts are ≥ 5 and there is large heterogeneity.

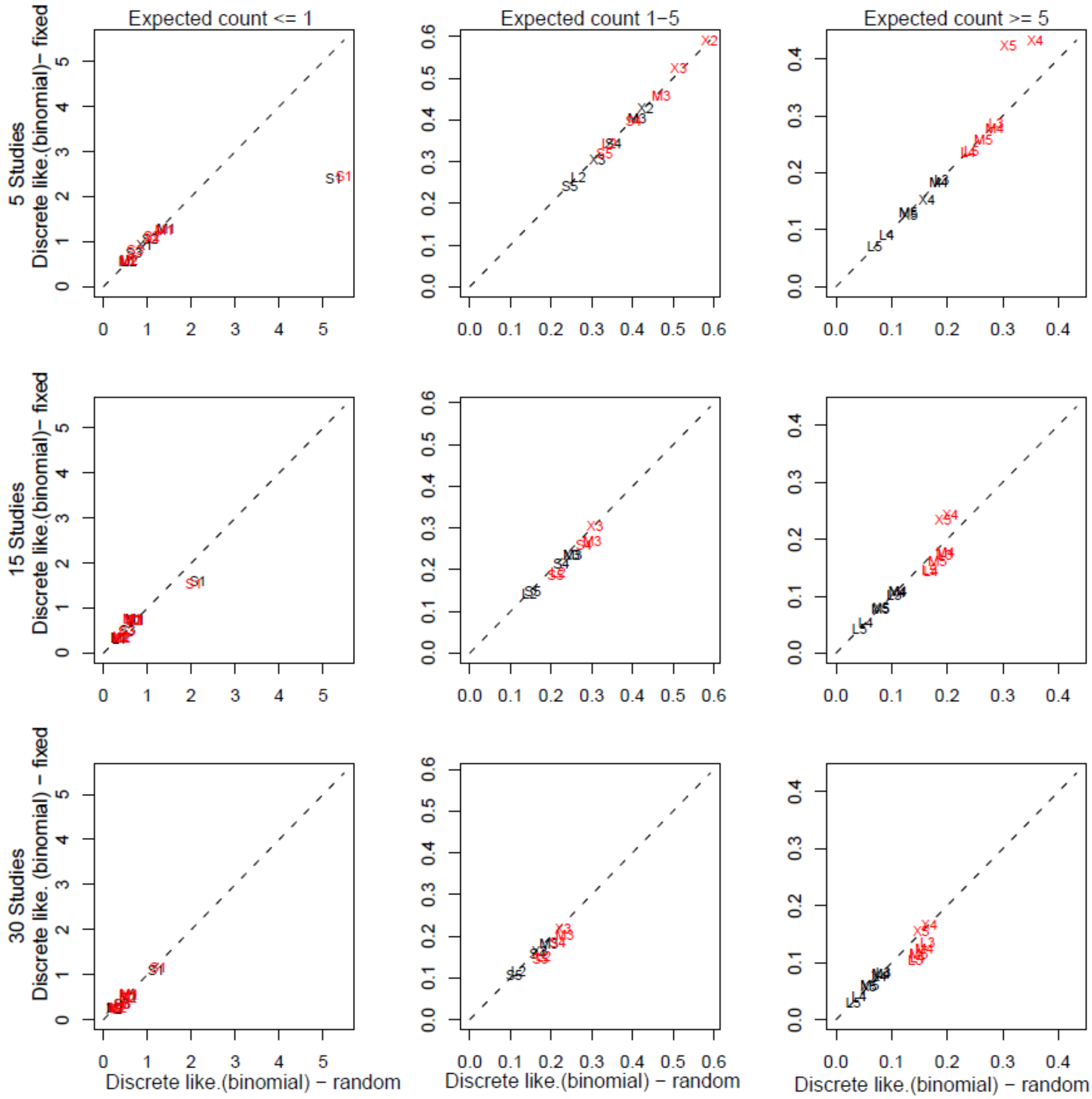
The undercoverage results from the bias exhibited by the random effects discrete likelihood method.

Figure 27. Comparison of proportion bias: fixed versus random effects with the discrete likelihood method



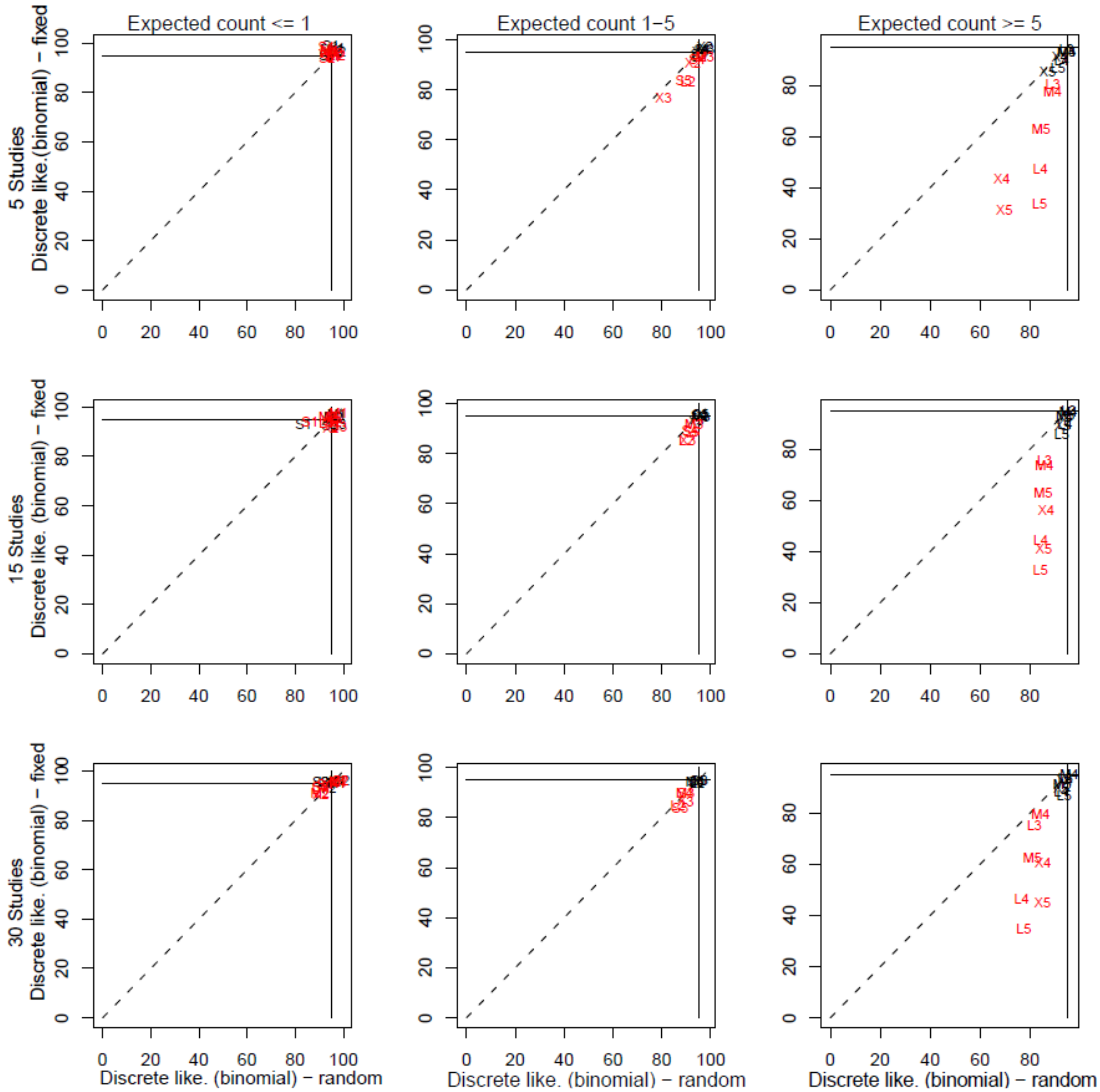
Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns.

Figure 28. Comparison of proportion RMSE: fixed versus random effects with the discrete likelihood method



Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal proportion RMSE for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. Note the change in scale across columns.

Figure 29. Comparison of coverage: fixed versus random effects with the discrete likelihood method

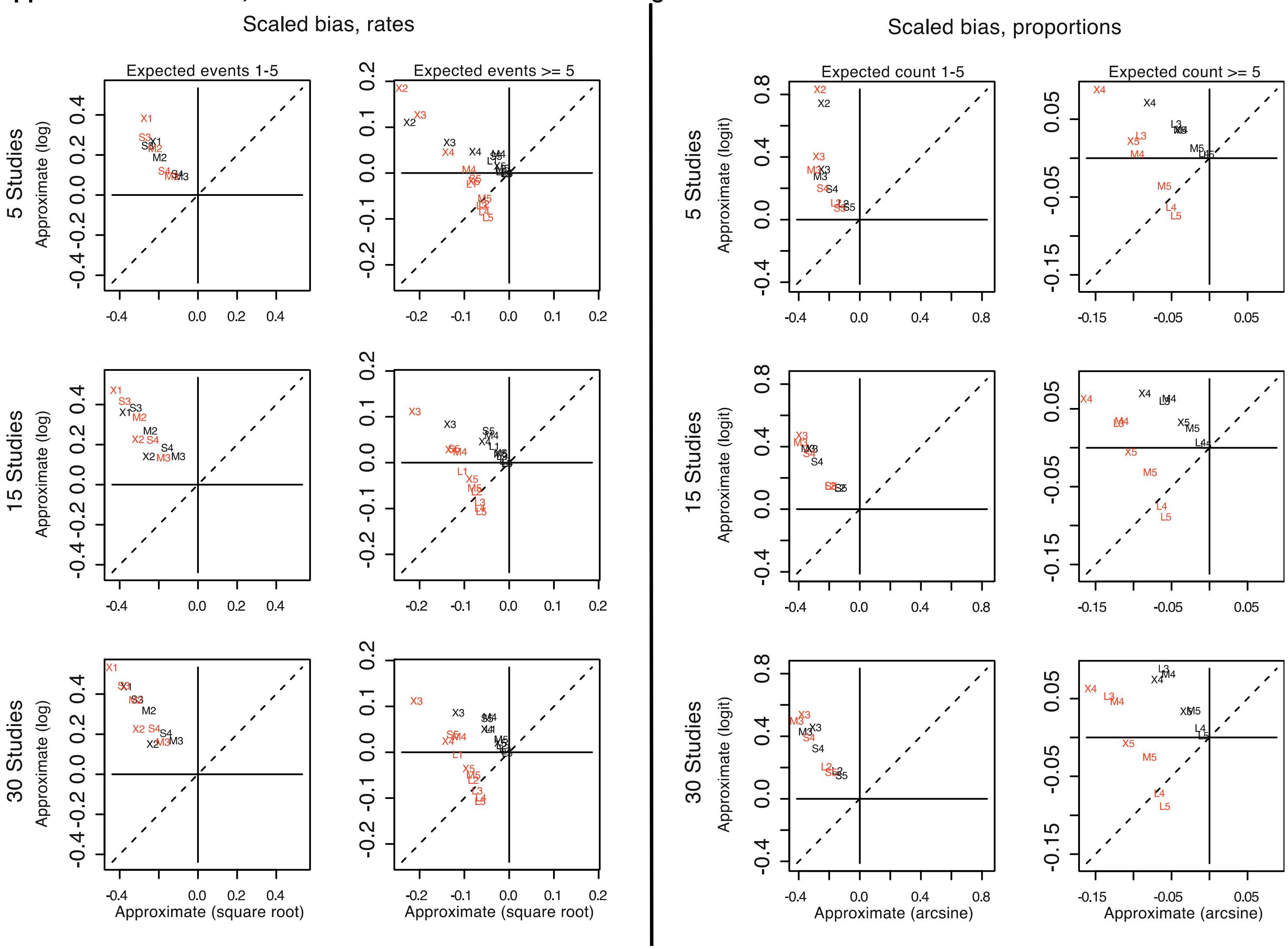


Rows correspond to number of studies. Columns correspond to ranges of expected count. Points on the dashed line have equal coverage for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 95 percent coverage.

Results for Rates

Simulation results for rates were very similar to those of proportions, and are thus not shown in detail. To convey the similarity, we show an example of the proportion bias for simulations of rates and of proportions for the comparison of approximate methods with the canonical versus the variance stabilizing transformation (Figure 30). The similarity in the results is obvious, and is theoretically expected, as the Poisson is a limiting case of the binomial distribution as the exposure increases.

Figure 30. Side by side comparison of results from simulations for proportions and rates: Comparison of proportion bias for approximate methods, canonical versus variance stabilizing transformations



Rows correspond to number of studies. Columns correspond to ranges of expected events (for rates, left side) and ranges of expected counts (for proportions, right side). Points on the dashed line have equal proportion bias for both methods. Points indicate simulation scenarios, and are coded with two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true rates or true proportions: 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1). Black color indicates scenarios with small heterogeneity, and red scenarios with large heterogeneity. Only a representative subset of scenarios is plotted. The solid black reference lines indicate 0 bias. Note the change in scale across columns, and from the left to the right panel.

Overview of Results Across All Methods

Table 10 shows the proportion bias with the random effects analysis methods for selected scenarios with high heterogeneity. Table 11 and Table 12 show the corresponding proportion RMSE and coverage. We make the following general observations for scenarios with expected counts 1 or less, between 1 and 5, and 5 or more:

- For expected counts ≤1, the hybrid method has proportion bias and RMSE that are closer to zero and coverage probability closer to 95 percent compared to other methods.
- For expected counts between 1 and 5 the random effects discrete likelihood method and the approximate method with the variance stabilizing transformation have comparable performance, and better than other methods.

> [For detailed descriptions and explanations refer to two sections: Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis and Pairwise Comparisons between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis. Analogous points apply to simulations of rates as well.]

- For expected counts of 5 or more, the differences between methods become less evident.

For numerical reasons, the random effects discrete likelihood method does not always converge. The following general comments can be made:

- For very small expected counts (<0.5) and for K=5 or K=15 the random effects discrete likelihood method reached convergence for fewer than 85 percent of the simulations (see column "Discrete (fraction converged)"). For expected counts above 1 or for K=30, the method converges practically for all simulations.
- For expected counts above 1 the random effect discrete likelihood method converged (almost) always, and thus the performance of the hybrid strategy is identical to that with random effects.

> The random effects discrete likelihood method will not converge when all studies in a meta-analysis have 0 events. This is more common in simulation scenarios with very low expected counts and when the number of studies is small.

We can make the following general comments on the preferred methods (hybrid strategy-discrete likelihood and variance-stabilizing transformation-approximate likelihood):

- For expected counts larger than 10, the hybrid strategy for meta-analysis using the discrete likelihood has larger absolute bias than the approximate method using the variance stabilizing transformation, and the reverse for smaller counts.

> [The explanations are analogous to those in two sections (Pairwise Comparisons Among Approximate Methods—Random Effects Meta-Analysis and Pairwise Comparisons Between Approximate and Discrete Likelihood Methods for Random Effects Meta-Analysis) for proportions; briefly, the discrete likelihood methods appear to have a constant relative bias of approximately -10 percent when heterogeneity is large, and when the expected number of counts is >5]

- For very large expected counts (for example when the true rate is 0.4) all compared methods converge in proportion bias and proportion RMSE.

This is congruent with what we expect theoretically: the normal approximation to the Poisson improves with increasing expected counts.

- The hybrid strategy using the discrete likelihood has better coverage probabilities than the other methods for the widest range of scenarios.

Finally, in sensitivity analyses that used numerical integration instead of a simple transformation to obtain the meta-analysis point estimates in the rate scale (see Methods section, paragraph on Approximate Methods), results were very comparable with those reported in Table 11 and Table 12 (not shown).

Table 10. Comparison of proportion bias across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	Small	0.1	2.689	3.207	-0.678	0.530	0.737	-0.079
M1	5	0.001	Medium	0.4	0.722	1.236	-0.561	0.836	0.080	-0.097
S2	5	0.005	Small	0.7	0.018	0.573	-0.430	0.972	-0.083	-0.109
S3	5	0.01	Small	1.5	-0.276	0.261	-0.289	1.000	-0.103	-0.103
M2	5	0.005	Medium	1.8	-0.313	0.211	-0.248	1.000	-0.100	-0.100
X1	5	0.001	Mixed	1.9	1.023	0.815	-0.158	0.994	0.066	0.059
M3	5	0.01	Medium	3.6	-0.359	0.073	-0.146	1.000	-0.094	-0.094
L1	5	0.001	Large	5.7	-0.337	0.014	-0.130	1.000	-0.100	-0.100
S4	5	0.05	Small	7.3	-0.296	0.012	-0.074	1.000	-0.074	-0.074
X2	5	0.005	Mixed	9.4	-0.065	0.161	-0.269	1.000	0.004	0.004
S5	5	0.1	Small	14.7	-0.243	-0.041	-0.060	1.000	-0.084	-0.084
M4	5	0.05	Medium	17.8	-0.240	-0.051	-0.059	1.000	-0.089	-0.089
X3	5	0.01	Mixed	18.8	-0.232	0.101	-0.192	1.000	-0.038	-0.038
L2	5	0.005	Large	28.4	-0.221	-0.058	-0.051	1.000	-0.085	-0.085
M5	5	0.1	Medium	35.6	-0.220	-0.081	-0.064	1.000	-0.101	-0.101
L3	5	0.01	Large	56.8	-0.206	-0.076	-0.052	1.000	-0.092	-0.092
S6	5	0.4	Small	58.8	-0.207	-0.088	-0.057	1.000	-0.099	-0.099
X4	5	0.05	Mixed	93.8	-0.275	-0.020	-0.078	1.000	-0.075	-0.075
M6	5	0.4	Medium	142.3	-0.185	-0.084	-0.044	1.000	-0.089	-0.089
X5	5	0.1	Mixed	187.6	-0.243	-0.053	-0.065	1.000	-0.087	-0.087
L4	5	0.05	Large	284	-0.200	-0.103	-0.060	1.000	-0.106	-0.106
L5	5	0.1	Large	567.9	-0.189	-0.098	-0.054	1.000	-0.099	-0.099
X6	5	0.4	Mixed	750.3	-0.210	-0.088	-0.063	1.000	-0.103	-0.103
L6	5	0.4	Large	2271.7	-0.166	-0.079	-0.033	1.000	-0.076	-0.076

Table 10. Comparison of proportion bias across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) (continued)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	15	0.001	Small	0.1	3.330	4.002	-0.812	0.849	-0.161	-0.288
M1	15	0.001	Medium	0.3	0.812	1.400	-0.673	0.990	-0.144	-0.152
S2	15	0.005	Small	0.6	0.123	0.774	-0.543	1.000	-0.129	-0.129
X1	15	0.001	Mixed	0.8	0.904	0.982	-0.417	0.999	-0.049	-0.050
S3	15	0.01	Small	1.2	-0.252	0.420	-0.374	1.000	-0.108	-0.108
M2	15	0.005	Medium	1.6	-0.343	0.313	-0.303	1.000	-0.103	-0.103
M3	15	0.01	Medium	3.3	-0.415	0.159	-0.156	0.999	-0.081	-0.081
X2	15	0.005	Mixed	3.9	-0.198	0.227	-0.300	1.000	-0.065	-0.065
L1	15	0.001	Large	5.6	-0.371	0.053	-0.144	1.000	-0.103	-0.103
S4	15	0.05	Small	6.1	-0.414	0.028	-0.121	1.000	-0.109	-0.109
X3	15	0.01	Mixed	7.7	-0.355	0.110	-0.214	1.000	-0.093	-0.093
S5	15	0.1	Small	12.3	-0.342	-0.044	-0.089	1.000	-0.113	-0.113
M4	15	0.05	Medium	16.5	-0.307	-0.052	-0.072	1.000	-0.105	-0.105
L2	15	0.005	Large	27.9	-0.305	-0.070	-0.076	1.000	-0.111	-0.111
M5	15	0.1	Medium	32.9	-0.267	-0.078	-0.060	1.000	-0.106	-0.106
X4	15	0.05	Mixed	38.6	-0.348	-0.037	-0.085	1.000	-0.108	-0.108
S6	15	0.4	Small	49.2	-0.257	-0.093	-0.065	1.000	-0.113	-0.113
L3	15	0.01	Large	55.8	-0.277	-0.094	-0.072	1.000	-0.119	-0.119
X5	15	0.1	Mixed	77.2	-0.290	-0.063	-0.067	1.000	-0.106	-0.106
M6	15	0.4	Medium	131.6	-0.232	-0.099	-0.051	1.000	-0.106	-0.106
L4	15	0.05	Large	279	-0.231	-0.107	-0.059	1.000	-0.113	-0.113
X6	15	0.4	Mixed	309	-0.250	-0.099	-0.059	1.000	-0.113	-0.113
L5	15	0.1	Large	558	-0.230	-0.112	-0.060	1.000	-0.115	-0.115
L6	15	0.4	Large	2231.9	-0.231	-0.115	-0.059	1.000	-0.113	-0.113

Table 10. Comparison of proportion bias across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) (continued)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	30	0.001	Small	0.1	3.284	4.021	-0.840	0.980	-0.345	-0.358
M1	30	0.001	Medium	0.3	0.874	1.489	-0.708	1.000	-0.169	-0.169
S2	30	0.005	Small	0.6	0.109	0.779	-0.559	1.000	-0.134	-0.134
X1	30	0.001	Mixed	0.7	0.820	1.169	-0.441	1.000	-0.095	-0.095
S3	30	0.01	Small	1.2	-0.258	0.459	-0.375	0.999	-0.099	-0.099
M2	30	0.005	Medium	1.6	-0.338	0.381	-0.313	1.000	-0.095	-0.095
M3	30	0.01	Medium	3.1	-0.436	0.164	-0.185	0.999	-0.101	-0.101
X2	30	0.005	Mixed	3.3	-0.208	0.232	-0.304	1.000	-0.096	-0.096
L1	30	0.001	Large	4.6	-0.364	0.100	-0.163	1.000	-0.102	-0.102
S4	30	0.05	Small	6.2	-0.421	0.034	-0.115	1.000	-0.105	-0.105
X3	30	0.01	Mixed	6.5	-0.349	0.109	-0.207	1.000	-0.103	-0.103
S5	30	0.1	Small	12.4	-0.368	-0.035	-0.084	1.000	-0.110	-0.110
M4	30	0.05	Medium	15.7	-0.328	-0.046	-0.070	1.000	-0.104	-0.104
L2	30	0.005	Large	23.2	-0.343	-0.058	-0.083	1.000	-0.113	-0.113
M5	30	0.1	Medium	31.5	-0.294	-0.091	-0.075	1.000	-0.121	-0.121
X4	30	0.05	Mixed	32.5	-0.356	-0.032	-0.080	1.000	-0.103	-0.103
L3	30	0.01	Large	46.3	-0.303	-0.084	-0.070	1.000	-0.115	-0.115
S6	30	0.4	Small	49.8	-0.268	-0.091	-0.060	1.000	-0.111	-0.111
X5	30	0.1	Mixed	65	-0.324	-0.069	-0.074	1.000	-0.113	-0.113
M6	30	0.4	Medium	125.8	-0.252	-0.110	-0.061	1.000	-0.118	-0.118
L4	30	0.05	Large	231.7	-0.259	-0.115	-0.068	1.000	-0.124	-0.124
X6	30	0.4	Mixed	260.2	-0.260	-0.100	-0.060	1.000	-0.115	-0.115
L5	30	0.1	Large	463.4	-0.242	-0.110	-0.058	1.000	-0.115	-0.115
L6	30	0.4	Large	1853.4	-0.238	-0.114	-0.055	1.000	-0.114	-0.114

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true rates; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4).
The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Table 11. Comparison of proportion RMSE across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	Small	0.1	2.716	3.365	0.862	0.530	1.180	1.099
M1	5	0.001	Medium	0.4	0.772	1.450	0.755	0.836	0.693	0.752
S2	5	0.005	Small	0.7	0.240	0.815	0.630	0.972	0.542	0.560
S3	5	0.01	Small	1.5	0.363	0.500	0.498	1.000	0.424	0.424
M2	5	0.005	Medium	1.8	0.411	0.458	0.478	1.000	0.418	0.418
X1	5	0.001	Mixed	1.9	1.213	1.005	0.463	0.994	0.579	0.582
M3	5	0.01	Medium	3.6	0.457	0.319	0.355	1.000	0.328	0.328
L1	5	0.001	Large	5.7	0.444	0.295	0.335	1.000	0.317	0.317
S4	5	0.05	Small	7.3	0.415	0.273	0.288	1.000	0.283	0.283
X2	5	0.005	Mixed	9.4	0.294	0.497	0.480	1.000	0.460	0.460
S5	5	0.1	Small	14.7	0.365	0.258	0.267	1.000	0.270	0.270
M4	5	0.05	Medium	17.8	0.352	0.238	0.244	1.000	0.252	0.252
X3	5	0.01	Mixed	18.8	0.363	0.394	0.399	1.000	0.384	0.384
L2	5	0.005	Large	28.4	0.334	0.241	0.243	1.000	0.250	0.250
M5	5	0.1	Medium	35.6	0.328	0.241	0.241	1.000	0.251	0.251
L3	5	0.01	Large	56.8	0.311	0.232	0.229	1.000	0.239	0.239
S6	5	0.4	Small	58.8	0.305	0.229	0.223	1.000	0.235	0.235
X4	5	0.05	Mixed	93.8	0.394	0.274	0.283	1.000	0.289	0.289
M6	5	0.4	Medium	142.3	0.293	0.236	0.229	1.000	0.238	0.238
X5	5	0.1	Mixed	187.6	0.354	0.250	0.252	1.000	0.259	0.259
L4	5	0.05	Large	284	0.297	0.235	0.223	1.000	0.237	0.237
L5	5	0.1	Large	567.9	0.294	0.239	0.228	1.000	0.240	0.240
X6	5	0.4	Mixed	750.3	0.306	0.229	0.224	1.000	0.240	0.240
L6	5	0.4	Large	2271.7	0.276	0.229	0.221	1.000	0.231	0.231

Table 11. Comparison of proportion RMSE across random effects methods for meta-analysis of rates (selected scenarios with with high heterogeneity) (continued)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	15	0.001	Small	0.1	3.339	4.054	0.849	0.849	0.810	0.842
M1	15	0.001	Medium	0.3	0.827	1.487	0.722	0.990	0.520	0.527
S2	15	0.005	Small	0.6	0.183	0.877	0.605	1.000	0.396	0.396
X1	15	0.001	Mixed	0.8	0.974	1.065	0.499	0.999	0.439	0.440
S3	15	0.01	Small	1.2	0.282	0.520	0.451	1.000	0.292	0.292
M2	15	0.005	Medium	1.6	0.365	0.405	0.384	1.000	0.261	0.261
M3	15	0.01	Medium	3.3	0.440	0.248	0.253	0.999	0.208	0.208
X2	15	0.005	Mixed	3.9	0.250	0.392	0.380	1.000	0.270	0.270
L1	15	0.001	Large	5.6	0.405	0.178	0.228	1.000	0.200	0.200
S4	15	0.05	Small	6.1	0.446	0.162	0.205	1.000	0.193	0.193
X3	15	0.01	Mixed	7.7	0.382	0.253	0.295	1.000	0.225	0.225
S5	15	0.1	Small	12.3	0.384	0.150	0.172	1.000	0.185	0.185
M4	15	0.05	Medium	16.5	0.351	0.147	0.157	1.000	0.173	0.173
L2	15	0.005	Large	27.9	0.344	0.152	0.155	1.000	0.174	0.174
M5	15	0.1	Medium	32.9	0.309	0.150	0.144	1.000	0.167	0.167
X4	15	0.05	Mixed	38.6	0.388	0.155	0.169	1.000	0.180	0.180
S6	15	0.4	Small	49.2	0.292	0.156	0.144	1.000	0.169	0.169
L3	15	0.01	Large	55.8	0.315	0.159	0.147	1.000	0.174	0.174
X5	15	0.1	Mixed	77.2	0.333	0.149	0.151	1.000	0.171	0.171
M6	15	0.4	Medium	131.6	0.269	0.160	0.138	1.000	0.165	0.165
L4	15	0.05	Large	279	0.270	0.165	0.140	1.000	0.169	0.169
X6	15	0.4	Mixed	309	0.289	0.159	0.140	1.000	0.169	0.169
L5	15	0.1	Large	558	0.266	0.167	0.139	1.000	0.169	0.169
L6	15	0.4	Large	2231.9	0.265	0.167	0.136	1.000	0.168	0.168

Table 11. Comparison of proportion RMSE across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) (continued)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	30	0.001	Small	0.1	3.288	4.051	0.853	0.980	0.704	0.711
M1	30	0.001	Medium	0.3	0.881	1.534	0.728	1.000	0.418	0.418
S2	30	0.005	Small	0.6	0.144	0.835	0.588	1.000	0.300	0.300
X1	30	0.001	Mixed	0.7	0.873	1.206	0.490	1.000	0.332	0.332
S3	30	0.01	Small	1.2	0.273	0.515	0.416	0.999	0.222	0.222
M2	30	0.005	Medium	1.6	0.349	0.431	0.358	1.000	0.206	0.206
M3	30	0.01	Medium	3.1	0.448	0.214	0.235	0.999	0.173	0.173
X2	30	0.005	Mixed	3.3	0.235	0.314	0.344	1.000	0.200	0.200
L1	30	0.001	Large	4.6	0.385	0.163	0.210	1.000	0.166	0.166
S4	30	0.05	Small	6.2	0.438	0.122	0.166	1.000	0.157	0.157
X3	30	0.01	Mixed	6.5	0.366	0.192	0.253	1.000	0.178	0.178
S5	30	0.1	Small	12.4	0.391	0.104	0.131	1.000	0.148	0.148
M4	30	0.05	Medium	15.7	0.354	0.112	0.126	1.000	0.146	0.146
L2	30	0.005	Large	23.2	0.363	0.113	0.129	1.000	0.148	0.148
M5	30	0.1	Medium	31.5	0.316	0.128	0.119	1.000	0.152	0.152
X4	30	0.05	Mixed	32.5	0.376	0.109	0.130	1.000	0.143	0.143
L3	30	0.01	Large	46.3	0.325	0.124	0.117	1.000	0.148	0.148
S6	30	0.4	Small	49.8	0.289	0.125	0.106	1.000	0.141	0.141
X5	30	0.1	Mixed	65	0.346	0.120	0.123	1.000	0.149	0.149
M6	30	0.4	Medium	125.8	0.270	0.140	0.107	1.000	0.146	0.146
L4	30	0.05	Large	231.7	0.277	0.146	0.115	1.000	0.153	0.153
X6	30	0.4	Mixed	260.2	0.280	0.131	0.104	1.000	0.143	0.143
L5	30	0.1	Large	463.4	0.260	0.139	0.105	1.000	0.143	0.143
L6	30	0.4	Large	1853.4	0.256	0.142	0.102	1.000	0.143	0.143

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true rates; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4).

The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Table 12. Comparison of coverage across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	Small	0.1	1.000	0.000	0.999	0.530	0.947	0.972
M1	5	0.001	Medium	0.4	1.000	0.737	0.836	0.836	0.963	0.969
S2	5	0.005	Small	0.7	1.000	0.849	0.824	0.972	0.971	0.972
S3	5	0.01	Small	1.5	1.000	0.912	0.891	1.000	0.964	0.964
M2	5	0.005	Medium	1.8	1.000	0.907	0.894	1.000	0.946	0.946
X1	5	0.001	Mixed	1.9	1.000	0.737	0.796	0.994	0.788	0.783
M3	5	0.01	Medium	3.6	1.000	0.929	0.899	1.000	0.942	0.942
L1	5	0.001	Large	5.7	1.000	0.904	0.865	1.000	0.879	0.879
S4	5	0.05	small	7.3	1.000	0.897	0.868	1.000	0.876	0.876
X2	5	0.005	mixed	9.4	1.000	0.643	0.789	1.000	0.557	0.557
S5	5	0.1	small	14.7	1.000	0.864	0.842	1.000	0.826	0.826
M4	5	0.05	medium	17.8	1.000	0.886	0.867	1.000	0.853	0.853
X3	5	0.01	mixed	18.8	1.000	0.625	0.789	1.000	0.545	0.545
L2	5	0.005	large	28.4	1.000	0.877	0.870	1.000	0.842	0.842
M5	5	0.1	medium	35.6	0.998	0.851	0.840	1.000	0.825	0.825
L3	5	0.01	large	56.8	1.000	0.856	0.869	1.000	0.845	0.845
S6	5	0.4	small	58.8	0.995	0.869	0.875	1.000	0.846	0.846
X4	5	0.05	mixed	93.8	1.000	0.817	0.809	1.000	0.718	0.718
M6	5	0.4	medium	142.3	0.998	0.828	0.849	1.000	0.813	0.813
X5	5	0.1	mixed	187.6	0.999	0.827	0.815	1.000	0.776	0.776
L4	5	0.05	large	284	1.000	0.827	0.851	1.000	0.809	0.809
L5	5	0.1	large	567.9	1.000	0.810	0.828	1.000	0.789	0.789
X6	5	0.4	mixed	750.3	0.993	0.823	0.820	1.000	0.802	0.802
L6	5	0.4	large	2271.7	0.999	0.831	0.851	1.000	0.784	0.784

Table 12. Comparison of coverage across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) (continued)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	15	0.001	small	0.1	1.000	0.000	0.542	0.849	0.967	0.972
M1	15	0.001	medium	0.3	1.000	0.168	0.444	0.990	0.980	0.970
S2	15	0.005	small	0.6	1.000	0.503	0.545	1.000	0.980	0.980
X1	15	0.001	mixed	0.8	1.000	0.232	0.681	0.999	0.912	0.911
S3	15	0.01	small	1.2	1.000	0.624	0.710	1.000	0.965	0.965
M2	15	0.005	medium	1.6	1.000	0.713	0.774	1.000	0.949	0.949
M3	15	0.01	medium	3.3	1.000	0.833	0.875	0.999	0.930	0.930
X2	15	0.005	mixed	3.9	1.000	0.661	0.730	1.000	0.837	0.837
L1	15	0.001	large	5.6	1.000	0.928	0.856	1.000	0.897	0.897
S4	15	0.05	small	6.1	1.000	0.921	0.869	1.000	0.890	0.890
X3	15	0.01	mixed	7.7	1.000	0.773	0.780	1.000	0.851	0.851
S5	15	0.1	small	12.3	0.999	0.917	0.874	1.000	0.854	0.854
M4	15	0.05	medium	16.5	1.000	0.901	0.880	1.000	0.842	0.842
L2	15	0.005	large	27.9	1.000	0.887	0.893	1.000	0.865	0.865
M5	15	0.1	medium	32.9	0.996	0.882	0.897	1.000	0.856	0.856
X4	15	0.05	mixed	38.6	1.000	0.906	0.865	1.000	0.859	0.859
S6	15	0.4	small	49.2	0.978	0.825	0.870	1.000	0.821	0.821
L3	15	0.01	large	55.8	1.000	0.836	0.868	1.000	0.822	0.822
X5	15	0.1	mixed	77.2	0.997	0.884	0.863	1.000	0.847	0.847
M6	15	0.4	medium	131.6	0.990	0.816	0.895	1.000	0.835	0.835
L4	15	0.05	large	279	1.000	0.787	0.885	1.000	0.823	0.823
X6	15	0.4	mixed	309	0.979	0.804	0.874	1.000	0.827	0.827
L5	15	0.1	large	558	1.000	0.783	0.877	1.000	0.814	0.814
L6	15	0.4	large	2231.9	0.993	0.778	0.893	1.000	0.810	0.810

Table 12. Comparison of coverage across random effects methods for meta-analysis of rates (selected scenarios with high heterogeneity) (continued)

Scenario	K	Rate	Exposure size	Expected count	Approximate untransformed	Approximate log	Approximate square root	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	30	0.001	small	0.1	1.000	0.000	0.223	0.980	0.978	0.958
M1	30	0.001	medium	0.3	1.000	0.000	0.192	1.000	0.980	0.980
S2	30	0.005	small	0.6	1.000	0.174	0.306	1.000	0.970	0.970
X1	30	0.001	mixed	0.7	1.000	0.000	0.470	1.000	0.931	0.931
S3	30	0.01	small	1.2	1.000	0.362	0.508	0.999	0.963	0.963
M2	30	0.005	medium	1.6	1.000	0.398	0.620	1.000	0.934	0.934
M3	30	0.01	medium	3.1	1.000	0.754	0.747	0.999	0.887	0.887
X2	30	0.005	mixed	3.3	1.000	0.598	0.569	1.000	0.886	0.886
L1	30	0.001	large	4.6	1.000	0.829	0.764	1.000	0.884	0.884
S4	30	0.05	small	6.2	1.000	0.926	0.815	1.000	0.848	0.848
X3	30	0.01	mixed	6.5	1.000	0.796	0.676	1.000	0.866	0.866
S5	30	0.1	small	12.4	0.989	0.935	0.848	1.000	0.803	0.803
M4	30	0.05	medium	15.7	1.000	0.884	0.862	1.000	0.784	0.784
L2	30	0.005	large	23.2	1.000	0.886	0.838	1.000	0.788	0.788
M5	30	0.1	medium	31.5	0.994	0.809	0.859	1.000	0.739	0.739
X4	30	0.05	mixed	32.5	1.000	0.927	0.860	1.000	0.834	0.834
L3	30	0.01	large	46.3	1.000	0.836	0.865	1.000	0.762	0.762
S6	30	0.4	small	49.8	0.935	0.796	0.883	1.000	0.770	0.770
X5	30	0.1	mixed	65	0.997	0.853	0.840	1.000	0.760	0.760
M6	30	0.4	medium	125.8	0.961	0.710	0.877	1.000	0.728	0.728
L4	30	0.05	large	231.7	1.000	0.677	0.838	1.000	0.704	0.704
X6	30	0.4	mixed	260.2	0.946	0.737	0.870	1.000	0.749	0.749
L5	30	0.1	large	463.4	1.000	0.692	0.873	1.000	0.741	0.741
L6	30	0.4	large	1853.4	0.980	0.666	0.886	1.000	0.730	0.730

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true rates; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4).
The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Practical Recommendations for Meta-Analysis of Proportions or Rates

The following are practical recommendations for meta-analysts, and probably apply generally. For most meta-analyses we would think that a random effects analysis will be adopted, because clinical and methodological diversity are more often present than not.

Recommendation 1. Use meta-analysis methods that model within study data using the binomial likelihood (for proportions) or the Poisson likelihood (for rates).

If the number of events is 0 for the vast majority of studies, the random effects methods may not converge. It is reasonable to use a hybrid strategy where if the random effects analysis fails to converge, one performs a meta-analysis with the fixed effects discrete likelihood method.

Recommendation 2. If recommendation 1 cannot be followed,[i] use meta-analysis models that rely on a variance stabilizing transformation (arcsine for proportions and square root for rates along with the normal approximation to the binomial (for proportions) or the Poisson (for rates).

Recommendations 1 and 2 both favor methods that do not suffer from the correlation between the estimate of the proportion and its variance (or the estimate of the rate and its variance), which is a problem with approximate methods on untransformed or canonically transformed proportions (rates). Further, using methods other than those in recommendations 1 and 2 may necessitate continuity corrections in applications with rare events. Continuity corrections can yield very biased summary proportions or rates, and thus should be avoided. Therefore, using methods other than the ones above recommended above can affect the meta-analysis summary results and conclusions.

If the estimated proportions and sample sizes in all studies are such that the expected number of events and the expected number of no events are relatively large (e.g., at least 5) the examined methods converge. Thus the choice of method is arguably less important. Similarly for rates, methods converge when the expected numbers of events are relatively large for all studies (e.g., more than 5).

[i] For example, if someone does not have access to software that can perform the analyses suggested in Recommendation 1.

Discussion

We describe a comprehensive simulation study of a wide range of methods for meta-analysis of proportions and rates. Based on the proportion bias, proportion RMSE and coverage probabilities of the compared methods we make concrete and clear recommendations of immediate relevance to meta-analytic practice. Specifically, we recommend the discrete likelihood methods that model within study data using the binomial likelihood (for proportions) or the Poisson likelihood (for rates) over methods that use normal distributions to approximate within-study variability. If approximate methods must be used (e.g., because there is no access to specialized statistical packages) we recommend using a variance stabilizing transformation, that is, the arcsine for proportions and the square root for rates.

Based on the observations gleaned from this study we conjecture that three major sources of statistical bias have to be dealt with in the meta-analysis of proportions or rates. The first two can be dealt with in a straightforward manner by following the aforementioned recommendations. The third source of statistical bias cannot be dealt with, but we believe that it does not change the recommendations, because it leaves the relative ranking of the available methods unchanged. We expand below.

Bias Stemming From Continuity Corrections

We have shown clearly that the use of correction factors can introduce arbitrary bias, and degrade RMSE and coverage. Methods necessitating continuity corrections should therefore be avoided when events are rare (or, for proportions, when all units experience an event). The discrete likelihood method (recommendation #1) and the approximate method using the variance stabilizing transformation (recommendation #2) are not susceptible to this limitation.

Bias Stemming From the Correlation Between Estimate and Its Variance

A second source of bias stems from the correlation that is induced between the point estimate and its variance for untransformed or canonically transformed data. Intuitively, this correlation results in a “distortion” of the weights assigned to studies in inverse variance meta-analysis, and this results in a statistical bias. The discrete likelihood method (recommendation #1) is not susceptible to this problem. For approximate methods, the variance stabilizing transformation (recommendation #2) removes the correlation between estimates and their variance, and corrects the statistical bias.

Note that this bias is not limited to rare event situations, but applies throughout the range of proportions or rates. Indeed, we found clear differences between the three approximate methods in the proportion bias, proportion RMSE and coverage probabilities, with the variance stabilizing transformations being preferable. In a limited simulation study comparing the discrete likelihood method with the canonical transformation for the meta-analysis of proportions, Hamza et al. conjectured that all approximate methods would have comparably large bias, because “there will always be a correlation between the estimate and the within-study variance, as they are determined by the same parameter […]”.[6] Hamza et al. did not recognize the arcsine transformation as a variance stabilizing one, i.e., one that does not suffer from the bias induced by the correlation between the estimates of the proportion (or rate) and the variance.

Bias Stemming From a Misspecification of the Random Effects Distribution

Finally, the third source of bias stems from the fact that the true random effects distribution of the proportions or rates is unknown. Meta-analysis models that assume a random effects distribution other than the true will introduce some statistical bias. The more extreme the departure of the assumed random effects distribution from the true one, the greater the bias. Because the true distribution is unknowable, it is impossible to eliminate this bias. Instead, one can examine the relative performance of meta-analysis methods in data simulated using alternative random effects distributions. General recommendations are feasible when the relative ranking of the methods is stable across the simulated random effects distributions. We simulated beta and uniform distributions in the proportion scale and (gamma and uniform distributions for rates). Excluding the approximate methods that use untransformed data, the other meta-analysis methods make calculations in a transformed scale (i.e., using the logit() or the arcsin(sqrt()) transformation for proportions, and the log() and sqrt() transformation for rates).[j]

All these functions are concave for proportions between 0, 0.50, and therefore introduce a negative bias: The mean in the transformed scale will be smaller than the transformation of the mean in the proportion scale. For concreteness, consider that the true proportion is 0.10 and that the square root of the heterogeneity parameter is 0.05 (large heterogeneity). These values correspond to a Beta(3.5, 31.5). The mean of the logit transformation of this Beta is -2.331 (by numerical Monte Carlo integration), which corresponds to 0.089 in the proportion scale, a -11 percent proportion bias. Using a uniform distribution instead of a beta yields a similar bias. The specific distribution would be a Uniform(0.0134, 0.1866), and the mean of its logit-transformed values is -2.368, which corresponds to 0.086 in the proportion scale, and a proportion bias of -14 percent. If one were using the variance stabilizing transformation, which is less “drastic”[k], the statistical bias is smaller: -2.6 percent for the beta, and -3.4 percent for the uniform.

The canonical and variance stabilizing transformations for proportions are point-symmetric around 0.50. Thus, they have no bias at exactly 0.50, and have a positive bias above 0.50. In our simulations we observed that for true proportions equal to 0.50, the bias with the canonical and variance stabilizing transformations is almost zero.

The analogous applies to rates. However, the canonical and variance-stabilizing transformations for rates are strictly concave, and therefore have a negative bias throughout the positive domain, which diminishes as the expected number of events rises.

As discussed above, the bias introduced by the potential misspecification of the true random effects distribution in the meta-analysis affects meta-analysis methods using canonical and variance-stabilizing transformations, as well as the discrete likelihood methods in a similar manner. We conjecture that for many forms of sampling distributions the relative ordering of the methods with respect to proportion bias and RMSE and coverage will remain relatively stable. Therefore, the recommendations #1 and #2 above are reasonable in general. For example, Hamza

---

jNote that this holds for the discrete likelihood methods as well: for proportions, the link function is the logit() and for rates the log().

[k]For proportions between 0 and 0.50, the first derivatives of the logit() and arcsine(sqrt()) transformations are both negative, but that of the logit is strictly smaller (more extreme). Because the statistical bias is determined by the first derivative of the transformations, it is more extreme for the canonical (former) compared to the variance stabilizing (latter) transformation. Because 0.50 is an inflection point for both transformations, there is no bias for mean proportions of 0.50.

et al.[6] assumed a log normal distribution for the random effects, which removes the third type of bias, as it matches the analysis methods exactly. Hamza et al. showed that the discrete likelihood method outperforms the approximate method using the canonical link (presumably because of the two other biases),[6] which is in accordance to our first recommendation.

We find it encouraging that our results (which focus on the summary proportion) did not seem to depend on the form of the distribution for proportions simulated: results were very consistent between scenarios using the uniform or the beta (gamma) distribution. This is not surprising given the above discussion. Further, it may be that the number of studies in our simulation study was small enough that —so long as the standard deviation is the same—it does not make a difference whether they were generated by a uniform or by a beta (or gamma). The shape of the distribution may have a large impact on the statistical properties of the heterogeneity estimator with each method. However, we did not focus on this aspect, which is reserved for future work.

Other Considerations

The recommended discrete likelihood methods are straightforward to fit in the familiar generalized linear mixed models (GLMM) framework. Meta-analysts who are not statistically sophisticated and are not familiar with statistical programming environments such as R, Stata or SAS will probably be challenged to follow our primary recommendation. First, most of these environments do not have an elaborate point-and-click interface, and some programming may be needed. Second, GLMM software routines can be unstable, and the default options may need adjustment to ensure convergence of the fitting algorithm. Specifically, to evaluate the log likelihood of a GLMM is it necessary to perform a numerical integration step.[22] The numerical integration is achieved by quadrature or adaptive quadrature algorithms, which become more precise (and more computationally expensive, i.e., slow) when they use more integration points. It is good practice to examine the robustness of the results using increasing number of integration points (e.g., 8, 12, 16, 20, 25).[10-12] The estimate of the between-study variance rather than the estimate of the summary proportion is more sensitive to the number of integration points.[11]

So what should one do if the GLMM algorithm does not converge? In our experience, non-convergence can be an issue if all or almost all studies have 0 in the numerators of the proportions or rates. Increasing the number of integration points does not necessarily fix the problem. When this happens, we propose to perform a meta-analysis using the discrete likelihood and fixed effects, or the approximate method with the variance stabilizing transformation as a secondary approach.

As of this writing, none of the standalone meta-analysis packages performs GLMM-based meta-analysis of proportions or rates,[13,14,23] despite the fact that it can be very easily added in those with a modular open source architecture such as OpenMeta-Analyst.[14] Therefore, the second recommendation to use a variance stabilizing transformation is important, as it is very easy to implement even in a spreadsheet program such as Microsoft Excel ™.

In sum, we believe that enough information exists to provide strong guidance on the performance characteristics of alternative methods for the meta-analysis of proportions and rates. While differences between the approximate and the discrete likelihood methods attenuate when the expected count is larger than 5, i.e., when the sample sizes or the exposures are large and the proportions/rates are not close to zero, it is probably best to avoid analyses of untransformed or logit proportions. Analysts experienced with programming and GLMMs should prefer the discrete likelihood methods. When these fail or when the analyst is not experienced enough to use them, the variance stabilizing transformations will usually offer a safe alternative since many meta-analyses will have large enough expected counts and rates. Only if sufficient data are available and other options are not practical, should untransformed or logit proportions be used.

References

1. Kent DM, Trikalinos TA. Are “treatment” bare metal stents superior to “control” bare metal stents? A meta-analytic approach. Am Heart J. Apr 2008;155(4):624-9.
2. Alsheikh-Ali AA, Trikalinos TA, Kent DM, Karas RH. Statins, low-density lipoprotein cholesterol, and risk of cancer. J Am Coll Cardiol. Sep 30 2008;52(14):1141-7.
3. Chappell FM, Raab GM, Wardlaw JM. When are summary ROC curves appropriate for diagnostic meta-analyses? Stat med. Sep 20 2009;28(21):2653-68.
4. Trikalinos TA, Balion CM, Coleman CI, et al. Chapter 8: Meta-analysis of test performance when there is a “gold standard”. J Gen Intern Med. 2012 Jun;27 Suppl 1:S56-66.
5. Chang BH, Waternaux C, Lipsitz S. Meta-analysis of binary data: which within study variance estimate to use? Stat med. Jul 15 2001;20(13):1947-56.
6. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. J Clin Epidemiol. Jan 2008;61(1):41-51.
7. Platt RW, Leroux BG, Breslow N. Generalized linear mixed models for meta-analysis. Stat med. Mar 30 1999;18(6):643-54.
8. Dahabreh IJ, Chung M, Kitsios GD, et al. Comprehensive Overview of Methods and Reporting of Meta-Analyses of Test Accuracy. AHRQ Methods for Effective Health Care. AHRQ Publication No. 12-EHC044-EF. Rockville, MD: Agency for Healthcare Research and Quality (US); 2012 Mar.
9. Kulinskaya E, Morgenthaler S, Staudte RG. Meta-analysis: a guide to calibrating and combining statistical evidence. Sussex, England: Wiley; 2008.
10. Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Boca Raton, FL: Chapman & Hall/CRC; 2004.
11. Stata Corp. Stata longitudinal-data/panel-data reference manual. Release 11. College Station, TX: Stata Press; 2009.
12. Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using Stata. 2 ed. College Station, TX: Stata Press; 2008.
13. Wallace BC, Schmid CH, Lau J, Trikalinos TA. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. BMC medical research methodology. 2009;9:80.
14. Wallace BC, Dahabreh IJ, Trikalinos TA, et al. Closing the Gap Between Methodologists and End-Users: R as a Computational Back-end. J Stat Software. 2013;49:1-15.
15. Brown LD, Cai TT, DasGupta A. Confidence intervals for a binomial proportion and asymptotic expansions. Ann Stat. Feb 2002;30(1):160-201.
16. Brown LD, Cai TT, DasGupta A, et al. Interval estimation for a binomial proportion—Comment—Rejoinder. Stat Sci. May 2001;16(2):101-33.
17. Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. Am Stat. May 1998;52(2):119-26.
18. Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. Am Stat. Nov 2000;54(4):280-8.
19. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. Stat med. Dec 20 2010;29(29):3046-67.
20. Hedges LV, Olkin I. Statistical methods for meta-analysis. London, UK: Academic Press; 1985.
21. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled clinical trials. Sep 1986;7(3):177-88.
22. Bishop CM. Pattern recognition and machine learning. Singapore: Springer; 2006.

23. Bax L, Yu LM, Ikeda N, Moons KG. A systematic comparison of software dedicated to meta-analysis of causal studies. BMC medical research methodology. 2007;7:40.

Abbreviations

AHRQ	Agency for Healthcare Research and Quality
CER	Comparative Effectiveness Review
EPC	Evidence based Practice Center
GLMM	generalized linear mixed models
RCT	Randomized controlled trial

Appendix A. Additional Descriptions of Methods and Results

Table A-1 Sample sizes used in simulations for proportions

	Small	*Heterogeneity Medium*	*Large*	*Mixed*
5 Studies	41 50 33 26 21	120 69 161 154 138	339 974 619 675 576	26 9 109 63 887
15 Studies	32 28 15 30 6 46 45 33 21 29 26 20 19 44 11	176 78 152 85 60 92 73 55 126 164 113 126 138 165 118	655 326 815 889 991 373 262 565 956 245 701 282 881 810 676	46 42 21 16 35 26 32 83 197 119 198 100 58 241 794
30 Studies	20 23 22 15 8 13 27 11 38 48 44 31 50 40 33 45 43 16 8 34 42 35 24 29 31 30 5 20 21 13	60 71 153 184 183 105 99 145 112 59 153 59 90 70 154 134 124 176 59 66 104 94 75 66 117 62 144 113 132 62	270 626 212 267 760 217 659 946 547 884 829 342 680 261 272 206 259 440 906 717 234 421 772 714 477 512 529 640 900 958	43 36 49 11 27 5 18 14 31 27 50 47 9 50 12 157 199 178 91 121 173 107 181 155 169 120 186 441 324 782

These vectors were drawn randomly and then kept fixed throughout the simulations.

Table A-2 Sample sizes used in simulations for rates

Number of studies	*Small*	*Heterogeneity Medium*	*Large*	*Mixed*
K=5	170 199 142 120 104	339 237 421 407 375	2143 9686 5469 6137 4961	120 64 317 226 8652
K=15	138 125 84 132 53 186 181 142 102 130 120 100 99 180 72	452 256 403 269 220 284 245 210 352 427 325 352 375 429 336	5893 1994 7792 8674 9889 2552 1232 4823 9467 1025 6443 1463 8575 7733 6143	187 171 102 86 149 119 139 265 494 338 495 299 216 977 7549
K=30	101 110 107 85 62 79 123 71 161 193 178 138 198 165 142 184 176 87 61 145 173 148 114 129 136 132 52 100 105 77	220 242 405 467 466 309 297 389 323 218 406 217 280 240 407 367 348 452 218 231 308 288 249 231 334 223 388 326 363 223	1330 5551 633 1284 7146 691 5941 9358 4609 8613 7966 2180 6198 1220 1353 561 1194 3349 8876 6638 903 3120 7281 6598 3785 4205 4399 5724 8807 9493	175 151 197 70 125 50 95 82 138 124 199 189 64 197 73 414 497 455 281 342 446 314 461 409 437 339 471 3353 1967 7399

These vectors were drawn randomly and then kept fixed throughout the simulations.

Table A-3. Formulas for setting the parameters of beta, uniform and gamma distributions

Parameters for the distribution of the true effects in simulations	***Explanations for simulations of proportions and rates***	***Formulas for the first parameter***	***Formulas for the second parameter***
Beta: parameters α and β	For proportions, set $\mu = \pi_j$, and $\sigma = \tau_j \pi_j$ For rates: Not applicable	$\alpha = (1-\mu)\left(\frac{\mu}{\sigma}\right)^2 - \mu$	$\beta = \frac{1-\mu}{\mu}\alpha$
Uniform: bounds [a, b]	For proportions, set $\mu = \pi_j$, and $\sigma = \tau_j \pi_j$ For rates, set $\mu = \lambda_j$, and $\sigma = \tau_j \lambda_j$	$a = \mu - \sqrt{3}\sigma$	$b = \mu + \sqrt{3}\sigma$
Gamma: parameters α and β	For proportions: Not applicable For rates, set $\mu = \lambda_j$, and $\sigma = \tau_j \lambda_j$	$\alpha = \left(\frac{\mu}{\sigma}\right)^2$	$\beta = \frac{\sigma^2}{\mu}$

In the above, *j* indexes the simulation scenario. As described in the methods, the standard deviation of the true effects is parameterized as a multiplier on the magnitude of the true proportion or rate. When $\tau_j = 0$ the above distributions are not used and all simulated studies have the same true effect (see Methods).

Table A-4. Comparison of proportion bias across random effects methods (selected scenarios with zero heterogeneity)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	small	<0.1	12.422	14.531	-0.733	0.168	5.039	0.015
M1	5	0.001	medium	0.1	2.946	3.700	-0.680	0.482	0.947	-0.061
S2	5	0.005	small	0.2	1.949	2.698	-0.635	0.569	0.646	-0.064
X1	5	0.001	mixed	0.2	0.421	6.042	-0.258	0.652	0.467	-0.043
S3	5	0.01	small	0.3	0.656	1.233	-0.557	0.818	0.109	-0.093
L1	5	0.001	large	0.6	0.118	0.663	-0.401	0.956	-0.001	-0.045
M2	5	0.005	medium	0.6	0.080	0.605	-0.436	0.956	-0.049	-0.091
X2	5	0.005	mixed	1.1	0.061	0.743	-0.202	0.994	0.022	0.016
M3	5	0.01	medium	1.3	-0.189	0.288	-0.262	0.997	-0.043	-0.046
S4	5	0.05	small	1.7	-0.191	0.200	-0.173	1.000	-0.026	-0.026
X3	5	0.01	mixed	2.2	-0.005	0.318	-0.223	1.000	0.016	0.016
L2	5	0.005	large	3.2	-0.184	0.094	-0.108	1.000	-0.032	-0.032
S5	5	0.1	small	3.4	-0.132	0.083	-0.065	1.000	-0.011	-0.011
L3	5	0.01	large	6.4	-0.113	0.026	-0.058	1.000	-0.029	-0.029
M4	5	0.05	medium	6.4	-0.083	0.043	-0.032	1.000	-0.007	-0.007
X4	5	0.05	mixed	10.9	-0.056	0.063	-0.080	1.000	0.003	0.003
M5	5	0.1	medium	12.8	-0.032	0.021	-0.007	1.000	0.002	0.002
S6	5	0.4	small	13.7	-0.009	0.004	-0.003	1.000	-0.001	-0.001
S7	5	0.5	small	17.1	-0.001	-0.001	-0.001	1.000	-0.001	-0.001
X5	5	0.1	mixed	21.9	-0.045	0.025	-0.043	1.000	-0.001	-0.001
L4	5	0.05	large	31.8	-0.017	0.007	-0.005	0.998	-0.001	-0.001
M6	5	0.4	medium	51.4	-0.002	0.001	-0.001	1.000	0.000	0.000
L5	5	0.1	large	63.7	-0.007	0.003	-0.002	1.000	0.000	0.000
M7	5	0.5	medium	64.2	-0.001	-0.001	-0.001	1.000	-0.001	-0.001
X6	5	0.4	mixed	87.5	-0.006	0.002	-0.002	1.000	-0.001	-0.001
X7	5	0.5	mixed	109.4	-0.001	-0.001	-0.001	1.000	-0.001	-0.001
L6	5	0.4	large	254.6	0.001	0.001	0.001	0.999	0.001	0.001
L7	5	0.5	large	318.3	0.001	0.001	0.001	0.998	0.001	0.001
S1	15	0.001	small	<0.1	14.550	19.530	-0.887	0.336	0.456	-0.511
M1	15	0.001	medium	0.1	3.192	4.308	-0.815	0.808	-0.106	-0.278
X1	15	0.001	mixed	0.1	0.986	5.984	-0.634	0.864	-0.039	-0.169
S2	15	0.005	small	0.1	2.366	3.687	-0.784	0.875	-0.120	-0.230
S3	15	0.01	small	0.3	0.851	1.747	-0.678	0.982	-0.131	-0.146
M2	15	0.005	medium	0.6	0.117	0.775	-0.520	0.999	-0.072	-0.072
L1	15	0.001	large	0.6	0.017	0.741	-0.505	1.000	-0.084	-0.084
X2	15	0.005	mixed	0.7	0.001	0.886	-0.402	1.000	-0.037	-0.037
M3	15	0.01	medium	1.1	-0.215	0.390	-0.321	0.999	-0.033	-0.032

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
X3	15	0.01	mixed	1.3	-0.134	0.388	-0.295	1.000	-0.016	-0.016
S4	15	0.05	small	1.4	-0.250	0.302	-0.272	1.000	-0.036	-0.036
S5	15	0.1	small	2.7	-0.205	0.137	-0.109	1.000	-0.014	-0.014
L2	15	0.005	large	3.1	-0.228	0.137	-0.119	0.999	-0.021	-0.021
M4	15	0.05	medium	5.7	-0.127	0.070	-0.045	1.000	-0.006	-0.006
L3	15	0.01	large	6.3	-0.128	0.067	-0.046	0.999	-0.006	-0.006
X4	15	0.05	mixed	6.7	-0.118	0.061	-0.086	0.999	-0.010	-0.010
S6	15	0.4	small	10.8	-0.014	0.011	-0.003	0.998	0.002	0.002
M5	15	0.1	medium	11.5	-0.057	0.028	-0.018	0.999	-0.003	-0.003
X5	15	0.1	mixed	13.4	-0.069	0.033	-0.026	1.000	0.001	0.001
S7	15	0.5	small	13.5	0.000	0.000	0.000	1.000	0.000	0.000
L4	15	0.05	large	31.4	-0.024	0.011	-0.006	0.994	-0.001	-0.001
M6	15	0.4	medium	45.9	-0.003	0.002	-0.001	0.997	0.000	0.000
X6	15	0.4	mixed	53.5	-0.005	0.002	-0.001	0.998	0.000	0.000
M7	15	0.5	medium	57.4	0.001	0.001	0.001	0.996	0.001	0.001
L5	15	0.1	large	62.8	-0.010	0.006	-0.002	0.993	0.000	0.000
X7	15	0.5	mixed	66.9	0.000	0.000	0.000	0.999	0.000	0.000
L6	15	0.4	large	251.4	0.000	0.000	0.000	0.996	0.000	0.000
L7	15	0.5	large	314.2	0.000	0.000	0.000	0.985	0.000	0.000
S1	30	0.001	small	<0.1	14.013	19.620	-0.937	0.537	-0.684	-0.830
M1	30	0.001	medium	0.1	3.346	4.654	-0.854	0.949	-0.333	-0.367
X1	30	0.001	mixed	0.1	1.114	6.593	-0.733	0.979	-0.200	-0.217
S2	30	0.005	small	0.1	2.248	3.701	-0.823	0.983	-0.278	-0.291
S3	30	0.01	small	0.3	0.793	1.771	-0.696	1.000	-0.130	-0.130
M2	30	0.005	medium	0.5	0.141	0.846	-0.561	1.000	-0.071	-0.071
L1	30	0.001	large	0.5	0.101	0.864	-0.542	1.000	-0.059	-0.059
X2	30	0.005	mixed	0.6	-0.081	1.021	-0.418	1.000	-0.043	-0.043
M3	30	0.01	medium	1.1	-0.210	0.419	-0.360	0.998	-0.037	-0.037
X3	30	0.01	mixed	1.3	-0.193	0.456	-0.274	0.998	-0.009	-0.010
S4	30	0.05	small	1.4	-0.263	0.310	-0.275	1.000	-0.027	-0.027
S5	30	0.1	small	2.7	-0.221	0.150	-0.115	0.999	-0.008	-0.008
L2	30	0.005	large	2.7	-0.264	0.182	-0.138	1.000	-0.012	-0.012
M4	30	0.05	medium	5.4	-0.149	0.082	-0.051	0.999	-0.006	-0.006
L3	30	0.01	large	5.5	-0.162	0.081	-0.063	0.998	-0.011	-0.011
X4	30	0.05	mixed	6.4	-0.110	0.065	-0.069	1.000	-0.007	-0.007
M5	30	0.1	medium	10.8	-0.069	0.032	-0.021	1.000	-0.004	-0.004
S6	30	0.4	small	10.9	-0.017	0.009	-0.005	0.998	0.000	0.000
X5	30	0.1	mixed	12.7	-0.063	0.030	-0.027	0.998	-0.001	-0.001

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S7	30	0.5	small	13.6	-0.002	-0.002	-0.002	1.000	-0.002	-0.002
L4	30	0.05	large	27.4	-0.030	0.014	-0.008	0.993	-0.001	-0.001
M6	30	0.4	medium	43.0	-0.002	0.004	0.001	0.997	0.002	0.002
X6	30	0.4	mixed	50.8	-0.004	0.003	0.000	0.997	0.001	0.001
M7	30	0.5	medium	53.8	0.000	0.000	0.000	0.999	0.000	0.000
L5	30	0.1	large	54.9	-0.013	0.006	-0.003	0.988	0.000	0.000
X7	30	0.5	mixed	63.5	0.000	0.000	0.000	0.997	0.000	0.000
L6	30	0.4	large	219.4	0.000	0.001	0.000	0.992	0.001	0.001
L7	30	0.5	large	274.3	0.000	0.000	0.000	0.986	0.000	0.000

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4, 7=0.5).
The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Table A-5. Comparison of proportion RMSE across random effects methods (selected scenarios with zero heterogeneity)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	small	<0.1	12.454	14.629	1.034	0.168	5.336	2.370
M1	5	0.001	medium	0.1	2.978	3.813	0.882	0.482	1.350	1.182
S2	5	0.005	small	0.2	1.999	2.847	0.850	0.569	1.111	1.064
X1	5	0.001	mixed	0.2	1.020	6.099	0.846	0.652	0.985	0.990
S3	5	0.01	small	0.3	0.732	1.406	0.743	0.818	0.661	0.734
L1	5	0.001	large	0.6	0.329	0.873	0.644	0.956	0.562	0.588
M2	5	0.005	medium	0.6	0.286	0.827	0.649	0.956	0.544	0.572
X2	5	0.005	mixed	1.1	0.436	0.898	0.450	0.994	0.438	0.444
M3	5	0.01	medium	1.3	0.329	0.500	0.477	0.997	0.400	0.403
S4	5	0.05	small	1.7	0.342	0.400	0.413	1.000	0.355	0.355
X3	5	0.01	mixed	2.2	0.290	0.478	0.418	1.000	0.313	0.313
L2	5	0.005	large	3.2	0.309	0.271	0.297	1.000	0.260	0.260
S5	5	0.1	small	3.4	0.269	0.241	0.257	1.000	0.234	0.234
L3	5	0.01	large	6.4	0.212	0.176	0.193	1.000	0.180	0.180
M4	5	0.05	medium	6.4	0.197	0.178	0.181	1.000	0.175	0.175
X4	5	0.05	mixed	10.9	0.167	0.179	0.220	1.000	0.153	0.153
M5	5	0.1	medium	12.8	0.128	0.121	0.122	1.000	0.121	0.121
S6	5	0.4	small	13.7	0.100	0.094	0.097	1.000	0.096	0.096
S7	5	0.5	small	17.1	0.080	0.076	0.078	1.000	0.077	0.077
X5	5	0.1	mixed	21.9	0.125	0.109	0.146	1.000	0.103	0.103
L4	5	0.05	large	31.8	0.079	0.077	0.077	0.998	0.077	0.077
M6	5	0.4	medium	51.4	0.049	0.048	0.049	1.000	0.048	0.048
L5	5	0.1	large	63.7	0.053	0.053	0.053	1.000	0.053	0.053
M7	5	0.5	medium	64.2	0.039	0.039	0.039	1.000	0.039	0.039
X6	5	0.4	mixed	87.5	0.052	0.043	0.047	1.000	0.041	0.041
X7	5	0.5	mixed	109.4	0.044	0.036	0.038	1.000	0.034	0.034
L6	5	0.4	large	254.6	0.023	0.023	0.023	0.999	0.022	0.022
L7	5	0.5	large	318.3	0.018	0.018	0.018	0.998	0.017	0.017
S1	15	0.001	small	<0.1	14.561	19.564	0.931	0.336	1.877	1.359
M1	15	0.001	medium	0.1	3.203	4.356	0.849	0.808	0.823	0.860
X1	15	0.001	mixed	0.1	1.272	6.021	0.747	0.864	0.728	0.771
S2	15	0.005	small	0.1	2.381	3.733	0.824	0.875	0.740	0.777
S3	15	0.01	small	0.3	0.877	1.811	0.734	0.982	0.560	0.571
M2	15	0.005	medium	0.6	0.194	0.856	0.586	0.999	0.377	0.377
L1	15	0.001	large	0.6	0.153	0.812	0.568	1.000	0.349	0.349
X2	15	0.005	mixed	0.7	0.288	0.934	0.481	1.000	0.332	0.332
M3	15	0.01	medium	1.1	0.261	0.471	0.405	0.999	0.253	0.253
X3	15	0.01	mixed	1.3	0.246	0.454	0.376	1.000	0.235	0.235

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S4	15	0.05	small	1.4	0.297	0.378	0.361	1.000	0.229	0.229
S5	15	0.1	small	2.7	0.255	0.200	0.199	1.000	0.154	0.154
L2	15	0.005	large	3.1	0.270	0.205	0.197	0.999	0.151	0.151
M4	15	0.05	medium	5.7	0.170	0.128	0.122	1.000	0.109	0.109
L3	15	0.01	large	6.3	0.164	0.120	0.114	0.999	0.100	0.100
X4	15	0.05	mixed	6.7	0.155	0.120	0.150	0.999	0.101	0.101
S6	15	0.4	small	10.8	0.069	0.060	0.063	0.998	0.061	0.061
M5	15	0.1	medium	11.5	0.095	0.077	0.077	0.999	0.074	0.074
X5	15	0.1	mixed	13.4	0.105	0.078	0.084	1.000	0.071	0.071
S7	15	0.5	small	13.5	0.054	0.049	0.051	1.000	0.050	0.050
L4	15	0.05	large	31.4	0.052	0.047	0.046	0.994	0.045	0.045
M6	15	0.4	medium	45.9	0.030	0.029	0.029	0.997	0.029	0.029
X6	15	0.4	mixed	53.5	0.031	0.028	0.029	0.998	0.028	0.028
M7	15	0.5	medium	57.4	0.024	0.024	0.024	0.996	0.024	0.024
L5	15	0.1	large	62.8	0.033	0.032	0.031	0.993	0.031	0.031
X7	15	0.5	mixed	66.9	0.023	0.022	0.023	0.999	0.022	0.022
L6	15	0.4	large	251.4	0.013	0.012	0.012	0.996	0.012	0.012
L7	15	0.5	large	314.2	0.010	0.010	0.010	0.985	0.010	0.010
S1	30	0.001	small	<0.1	14.017	19.637	0.944	0.537	0.993	0.996
M1	30	0.001	medium	0.1	3.352	4.677	0.867	0.949	0.759	0.773
X1	30	0.001	mixed	0.1	1.212	6.610	0.766	0.979	0.588	0.599
S2	30	0.005	small	0.1	2.254	3.723	0.836	0.983	0.633	0.641
S3	30	0.01	small	0.3	0.806	1.799	0.722	1.000	0.432	0.432
M2	30	0.005	medium	0.5	0.181	0.888	0.591	1.000	0.280	0.280
L1	30	0.001	large	0.5	0.158	0.903	0.574	1.000	0.278	0.278
X2	30	0.005	mixed	0.6	0.227	1.047	0.468	1.000	0.259	0.259
M3	30	0.01	medium	1.1	0.236	0.466	0.404	0.998	0.198	0.198
X3	30	0.01	mixed	1.3	0.244	0.484	0.325	0.998	0.168	0.169
S4	30	0.05	small	1.4	0.287	0.352	0.321	1.000	0.165	0.165
S5	30	0.1	small	2.7	0.245	0.182	0.163	0.999	0.105	0.105
L2	30	0.005	large	2.7	0.286	0.216	0.185	1.000	0.114	0.114
M4	30	0.05	medium	5.4	0.169	0.113	0.096	0.999	0.079	0.079
L3	30	0.01	large	5.5	0.180	0.113	0.103	0.998	0.080	0.080
X4	30	0.05	mixed	6.4	0.133	0.097	0.107	1.000	0.073	0.073
M5	30	0.1	medium	10.8	0.088	0.062	0.059	1.000	0.055	0.055
S6	30	0.4	small	10.9	0.051	0.043	0.045	0.998	0.044	0.044
X5	30	0.1	mixed	12.7	0.082	0.057	0.059	0.998	0.048	0.048
S7	30	0.5	small	13.6	0.038	0.034	0.036	1.000	0.035	0.035
L4	30	0.05	large	27.4	0.046	0.037	0.035	0.993	0.034	0.034

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
M6	30	0.4	medium	43.0	0.022	0.021	0.021	0.997	0.021	0.021
X6	30	0.4	mixed	50.8	0.022	0.020	0.020	0.997	0.020	0.020
M7	30	0.5	medium	53.8	0.018	0.017	0.017	0.999	0.017	0.017
L5	30	0.1	large	54.9	0.027	0.025	0.024	0.988	0.024	0.024
X7	30	0.5	mixed	63.5	0.017	0.016	0.016	0.997	0.016	0.016
L6	30	0.4	large	219.4	0.010	0.010	0.010	0.992	0.010	0.010
L7	30	0.5	large	274.3	0.008	0.008	0.008	0.986	0.008	0.008

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (K), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4, 7=0.5).

The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.

Table A-6. Comparison of coverage across random effects methods (selected scenarios with zero heterogeneity)

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S1	5	0.001	small	<0.1	1.000	0.000	1.000	0.168	0.988	1.000
M1	5	0.001	medium	0.1	1.000	0.000	1.000	0.482	0.977	1.000
S2	5	0.005	small	0.2	1.000	0.431	1.000	0.569	0.952	1.000
X1	5	0.001	mixed	0.2	0.998	0.000	0.650	0.652	0.972	0.998
S3	5	0.01	small	0.3	1.000	0.774	0.818	0.818	0.985	1.000
L1	5	0.001	large	0.6	1.000	0.847	0.823	0.956	0.964	1.000
M2	5	0.005	medium	0.6	0.999	0.863	0.815	0.956	0.965	0.999
X2	5	0.005	mixed	1.1	0.936	0.849	0.913	0.994	0.970	0.936
M3	5	0.01	medium	1.3	0.946	0.917	0.888	0.997	0.971	0.946
S4	5	0.05	small	1.7	0.852	0.931	0.926	1.000	0.965	0.852
X3	5	0.01	mixed	2.2	0.932	0.919	0.946	1.000	0.950	0.932
L2	5	0.005	large	3.2	0.795	0.954	0.949	1.000	0.975	0.795
S5	5	0.1	small	3.4	0.852	0.962	0.968	1.000	0.959	0.852
L3	5	0.01	large	6.4	0.875	0.955	0.950	1.000	0.960	0.875
M4	5	0.05	medium	6.4	0.916	0.956	0.964	1.000	0.961	0.916
X4	5	0.05	mixed	10.9	0.930	0.950	0.969	1.000	0.946	0.930
M5	5	0.1	medium	12.8	0.939	0.954	0.960	1.000	0.955	0.939
S6	5	0.4	small	13.7	0.956	0.968	0.960	1.000	0.965	0.956
S7	5	0.5	small	17.1	0.960	0.971	0.971	1.000	0.970	0.960
X5	5	0.1	mixed	21.9	0.935	0.965	0.975	1.000	0.949	0.935
L4	5	0.05	large	31.8	0.951	0.963	0.962	0.998	0.965	0.951
M6	5	0.4	medium	51.4	0.953	0.959	0.957	1.000	0.957	0.953
L5	5	0.1	large	63.7	0.954	0.952	0.954	1.000	0.952	0.954
M7	5	0.5	medium	64.2	0.964	0.964	0.964	1.000	0.961	0.964
X6	5	0.4	mixed	87.5	0.965	0.963	0.964	1.000	0.950	0.965
X7	5	0.5	mixed	109.4	0.971	0.972	0.972	1.000	0.958	0.971
L6	5	0.4	large	254.6	0.961	0.959	0.960	0.999	0.956	0.961
L7	5	0.5	large	318.3	0.973	0.973	0.973	0.998	0.968	0.973
S1	15	0.001	small	<0.1	0.000	0.000	1.000	0.336	0.954	0.000
M1	15	0.001	medium	0.1	0.230	0.000	0.521	0.808	0.975	0.230
X1	15	0.001	mixed	0.1	0.954	0.000	0.665	0.864	0.949	0.954
S2	15	0.005	small	0.1	0.761	0.000	0.585	0.875	0.955	0.761
S3	15	0.01	small	0.3	0.988	0.018	0.481	0.982	0.965	0.988
M2	15	0.005	medium	0.6	0.999	0.490	0.606	0.999	0.975	0.999
L1	15	0.001	large	0.6	0.999	0.498	0.592	1.000	0.978	0.999
X2	15	0.005	mixed	0.7	0.909	0.301	0.725	1.000	0.970	0.909
M3	15	0.01	medium	1.1	0.869	0.726	0.772	0.999	0.968	0.869

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
X3	15	0.01	mixed	1.3	0.851	0.688	0.802	1.000	0.950	0.851
S4	15	0.05	small	1.4	0.686	0.782	0.823	1.000	0.955	0.686
S5	15	0.1	small	2.7	0.642	0.890	0.925	1.000	0.966	0.642
L2	15	0.005	large	3.1	0.578	0.875	0.907	0.999	0.964	0.578
M4	15	0.05	medium	5.7	0.746	0.912	0.942	1.000	0.957	0.746
L3	15	0.01	large	6.3	0.743	0.924	0.952	0.999	0.971	0.743
X4	15	0.05	mixed	6.7	0.760	0.927	0.928	0.999	0.960	0.760
S6	15	0.4	small	10.8	0.948	0.960	0.955	0.998	0.957	0.948
M5	15	0.1	medium	11.5	0.882	0.945	0.950	0.999	0.954	0.882
X5	15	0.1	mixed	13.4	0.861	0.941	0.958	1.000	0.956	0.861
S7	15	0.5	small	13.5	0.943	0.964	0.960	1.000	0.957	0.943
L4	15	0.05	large	31.4	0.924	0.957	0.950	0.994	0.956	0.924
M6	15	0.4	medium	45.9	0.972	0.969	0.968	0.997	0.967	0.972
X6	15	0.4	mixed	53.5	0.950	0.957	0.955	0.998	0.953	0.950
M7	15	0.5	medium	57.4	0.958	0.960	0.959	0.996	0.958	0.958
L5	15	0.1	large	62.8	0.949	0.963	0.964	0.993	0.965	0.949
X7	15	0.5	mixed	66.9	0.966	0.965	0.963	0.999	0.959	0.966
L6	15	0.4	large	251.4	0.958	0.955	0.958	0.996	0.955	0.958
L7	15	0.5	large	314.2	0.963	0.964	0.963	0.985	0.962	0.963
S1	30	0.001	small	<0.1	0.000	0.000	1.000	0.537	0.989	0.000
M1	30	0.001	medium	0.1	0.000	0.000	0.254	0.949	0.966	0.000
X1	30	0.001	mixed	0.1	0.846	0.000	0.429	0.979	0.959	0.846
S2	30	0.005	small	0.1	0.000	0.000	0.227	0.983	0.977	0.000
S3	30	0.01	small	0.3	0.822	0.000	0.226	1.000	0.964	0.822
M2	30	0.005	medium	0.5	1.000	0.106	0.308	1.000	0.975	1.000
L1	30	0.001	large	0.5	1.000	0.092	0.331	1.000	0.972	1.000
X2	30	0.005	mixed	0.6	0.906	0.003	0.479	1.000	0.957	0.906
M3	30	0.01	medium	1.1	0.754	0.445	0.544	0.998	0.953	0.754
X3	30	0.01	mixed	1.3	0.701	0.278	0.659	0.998	0.962	0.701
S4	30	0.05	small	1.4	0.449	0.556	0.661	1.000	0.963	0.449
S5	30	0.1	small	2.7	0.381	0.754	0.881	0.999	0.956	0.381
L2	30	0.005	large	2.7	0.274	0.686	0.840	1.000	0.968	0.274
M4	30	0.05	medium	5.4	0.512	0.839	0.916	0.999	0.959	0.512
L3	30	0.01	large	5.5	0.458	0.857	0.914	0.998	0.957	0.458
X4	30	0.05	mixed	6.4	0.652	0.881	0.901	1.000	0.950	0.652
M5	30	0.1	medium	10.8	0.750	0.911	0.944	1.000	0.951	0.750
S6	30	0.4	small	10.9	0.926	0.958	0.957	0.998	0.957	0.926
X5	30	0.1	mixed	12.7	0.785	0.920	0.960	0.998	0.971	0.785

Scenario	K	Proportion	Sample size	Exp count	Approximate untransformed	Approximate logit	Approximate arcsine	Discrete (fraction converged)	Discrete (random)	Discrete (hybrid)
S7	30	0.5	small	13.6	0.949	0.964	0.961	1.000	0.960	0.949
L4	30	0.05	large	27.4	0.885	0.938	0.956	0.993	0.954	0.885
M6	30	0.4	medium	43.0	0.958	0.968	0.967	0.997	0.965	0.958
X6	30	0.4	mixed	50.8	0.958	0.959	0.962	0.997	0.959	0.958
M7	30	0.5	medium	53.8	0.950	0.956	0.951	0.999	0.952	0.950
L5	30	0.1	large	54.9	0.916	0.953	0.958	0.988	0.955	0.916
X7	30	0.5	mixed	63.5	0.964	0.958	0.963	0.997	0.959	0.964
L6	30	0.4	large	219.4	0.954	0.954	0.956	0.992	0.954	0.954
L7	30	0.5	large	274.3	0.948	0.948	0.948	0.986	0.949	0.948

"Discrete" stands for discrete likelihood methods. Scenarios are ordered by number of studies (*K*), and then by expected count. Bold horizontal lines separate scenarios by number of studies. White and grey shading separates scenarios with expected counts ≤1, between 1 and 5, and ≥5. We code scenarios using two characters, a letter (indicating sample size scenarios; S=small, M=medium, L=large, X=mixed); and a number (indicating the true proportions; 1=0.001, 2=0.005, 3=0.01, 4=0.05, 5=0.1, 6=0.4, 7=0.5). The column "Discrete (fraction converged)" shows the proportion of simulations for which random effects methods converged successfully. Values of 1.000 mean that random effects methods converged successfully in all 1000 simulations in a scenario, and a value of e.g., 0.335 means that they converged in 335 out of 1000 simulations in a scenario. The columns "Discrete (random)" and "Discrete (hybrid)" are identical when the fraction converged is 1.000, because the random effects method was used in all simulations.