Received: 23 November 2018 | Revised: 11 April 2019 | Accepted: 12 April 2019

DOI: 10.1002/sim.8193

RESEARCH ARTICLE

WILEY Statistics in Medicine

Simulation and data-generation for random-effects network meta-analysis of binary outcome

Svenja E. Seide | Katrin Jensen | Meinhard Kieser

Institute of Medical Biometry and Informatics, Ruprecht-Karls University Heidelberg, Heidelberg, Germany

Correspondence
Svenja E. Seide, Institute of Medical Biometry and Informatics, Ruprecht-Karls University Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany.
Email: seide@imbi.uni-heidelberg.de

The performance of statistical methods is frequently evaluated by means of simulation studies. In case of network meta-analysis of binary data, however, available data-generating models (DGMs) are restricted to either inclusion of two-armed trials or the fixed-effect model. Based on data-generation in the pairwise case, we propose a framework for the simulation of random-effect network meta-analyses including multiarm trials with binary outcome. The only one of the common DGMs used in the pairwise case, which is directly applicable to a random-effects network setting uses strongly restrictive assumptions. To overcome these limitations, we modify this approach and derive a related simulation procedure using odds ratios as effect measure. The performance of this procedure is evaluated with synthetic data and in an empirical example.

KEYWORDS

binary data, data-generating model, multiarm trials, network meta-analysis, random-effects model, simulation

1 | BACKGROUND

Network meta-analysis,[1] or mixed treatment comparison,[2,3] allows to evaluate multiple treatments for the same condition, even though they might have not been directly compared in a head-to-head trial. As such, it is a generalization of pairwise meta-analysis. A considerable amount of methodological research of network meta-analysis has been published in recent years.[3-6] However, only a limited number of simulation studies have been conducted to assess the methodological advances. Unlike in other areas of biomedical research, methods are commonly evaluated using an empirical example,[7-10] resulting in the notion of differences between candidate methods rather than the evaluation against known theoretical values. Alternatively, if simulations are reported, the data-generating model (DGM) is restricted to a common-effect model[11,12] where the strong assumption that one true effect sized is shared by all included studies is needed. Simulation studies for network meta-analyses that relax this homogeneity assumption and allow between-study variance are currently solely handling two-armed trials.[13,14] Regarding empirical networks of studies, neither of these restrictions is common in practice.[7] The differences between the common-effect and the random-effects model in meta-analyses have been discussed in detail in, eg, the work of Borenstein et al,[15] along with potential sources of heterogeneity. In the following, we concentrate on the case where between-study heterogeneity, irrespective of its reason or source, as well as multiarm trials are present and should be included in the DGM for simulation. The empirical network of eight anti-tumor necrosis factor treatments with 6 two-armed, 5 three-armed, and 2 four-armed studies described by Warren et al[8] is quite typical in this context. In the original article, different network meta-analytical models were evaluated on this empirical data set. We chose to replicate "model D" (see page 20 in the work of Warren et al[8] for further information) with respect to the geometry of the network and the theoretical effect sizes. This network is illustrated in Figure 1. We will come back to this example in Section 3 where we illustrate our proposed method by means of this scenario.

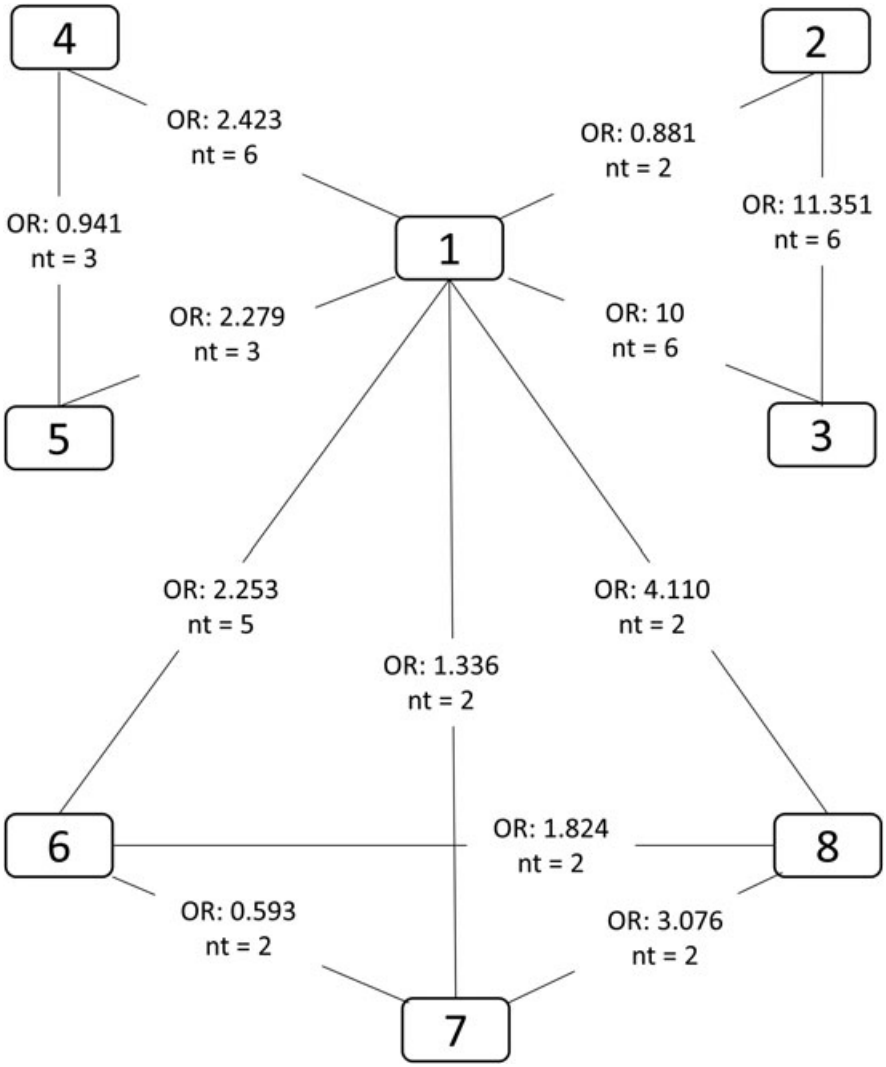

FIGURE 1 Network of trials for simulation (see the work of Warren et al[8]) with eight different treatments (nodes), and *nt* trials per relative effect (edges), 6 two-armed, 5 three-armed, and 2 four-armed studies

The standard model in random-effects network meta-analysis of binary outcome is the binomial-normal model, a contrast-based model with an arm-based likelihood. Network meta-analytical models require, beside the evaluation of potential between-study heterogeneity, the assumptions that indirect comparisons validly estimate the unobserved (missing) head-to-head comparison (transitivity) and should evaluate the closely related assumption of whether or not direct and indirect estimates are in agreement (consistency). An extensive discussion on these assumption and their implication can be found in the work of Salanti.[3] It is usually assumed that the number of events $x_{k,i}$ and the number of patients $n_{k,i}$ are available for treatment t in trial arm k of trial i. Commonly, a binomial likelihood is then placed on each trial-arm, such that $x_{k,i} \sim \text{Bin}(\pi_{k,i}, n_{k,i})$ and a logit transformation is used to obtain the relative treatment effect on a linear scale. Different relative effect measures are available for binary outcomes. We will concentrate exclusively on the odds-ratio (OR) as effect measure, where the log-odds ratio of treatment $k \neq 1$ relative to the reference treatment $k = 1$ in trial i will be denoted as $\theta_{1k,i}$. Heterogeneity between trials of a contrast may be taken into account by using (normally distributed) random-effects. However, unlike in the case of pairwise meta-analysis, relative effects in a network meta-analysis are not independent in the presence of multiarm trials, since each nonbaseline treatment is compared to the same baseline, such that treatment effects are commonly modeled by a multivariate normal distribution ($\boldsymbol{\theta}_i \sim \mathcal{MVN}(\boldsymbol{\theta}, \Sigma)$). A simple but convenient structure of the covariance matrix of the multivariate normal distribution is suggested by Higgins and Whitehead,[16] where Σ is symmetric and homogeneous, with τ^2 at its diagonal entries and $0.5\tau^2$ at the nondiagonal elements, assuming the heterogeneity variances to be a constant over treatment comparisons. Additionally, discrepancy between evidence from direct and from indirect comparisons is possible in network meta-analysis (inconsistency). We aim at simulating the simplified case, a consistent, transitive network with constant and common heterogeneity variances and multiarm trials within in the network we wish to generate.

The remainder of this paper is organized as follows. Section 2 evaluates DGMs used in the simulation of pairwise meta-analysis of binary outcome and their potential application and modification for a network of trials, which includes multiarm studies and between-study heterogeneity. In Section 3, we describe and discuss a simulation procedure replicating the empirical data set, and we close with a discussion in Section 4.

2 | DATA GENERATING MODELS FOR BINARY OUTCOME

Simulations of trial data in a common-effects meta-analysis of binary outcome would typically generate numbers of events and patients per trial arm, according to the binomial distribution with given theoretical event rates ($\pi_{k,i}$). An important difference exists between arm-based and contrast-based methods in meta-analysis. To evaluate an arm-based approach for network meta-analysis of a binary endpoint, synthetic data on arm-level is necessary. However, the same type of data is necessary in contrast-based approaches, which specify an arm-based likelihood, such as the model described in the work of Dias et al.[6] Contrast-based models that specify a contrast-based likelihood, eg, the model used in the work of Rücker,[17] do not require arm-level data but can be estimated using it by first calculating the effect measure. A data-generating mechanism resulting in arm-level data might therefore potentially be used in both types of methods. To reflect the random-effects model, the heterogeneity variance (τ^2) needs to be incorporated in the simulation of (relative) treatment effects previous to deriving event rates, resulting in a hierarchical structure. Pateras et al[18] recently evaluated three different DGMs commonly used in the literature for the simulation of pairwise binary meta-analyses and denoted these DGMs as "random," "average," and "fixed." The names of these data-generating mechanisms relate to the assumptions made about the event rate in the baseline arm and are not, despite their similarity, directly related to the notion of fixed- or random-effects meta-analyses. The three data-generating mechanisms are differing in the generation of event rates in the simulation of a random-effects meta-analysis. They are all equal in their second step, the generation of numbers of events per trial arm from a binomial distribution using these respective event rates, or in the absence of between-study heterogeneity. Pateras et al[18] find that statistical meta-analytical methods perform differently across the varying DGMs in pairwise meta-analysis, especially when simulating extreme scenarios, eg, few, unbalanced, and/or small trials, which are, however, common-case in practice. They are in line with Burton et al[19] in emphasizing the need of well-designed simulation studies and thorough reporting of the DGM in biomedical research to achieve a fair evaluation of potential methods. Pateras et al[18] conclude that, in simulation studies, interpretation of results and conclusions on properties or performance of methods, is impacted by the underlying DGM in simulation studies. When aiming at the simulation of a network of studies, all three of these DGMs may be used when solely two-armed trials are present in the network, as all simulated treatment effects are independent of each other. However, difficulties with the dependence structure of the data arise when multiarm trials are present. As each of the nonbaseline treatments in a trial is evaluated relative to the same baseline, treatment effects from the same trial may no longer be assumed independent in the simulation which is described in detail on page 35 in the work of Dias et al.[6] In the following, we therefore evaluate the potential application of these DGMs for multiarm trials under a random-effects model in network meta-analysis in more detail.

2.1 | DGM "random" applied to multiarm trials in a random-effects network meta-analysis

For this DGM, two initial event rates $\pi_{1,\text{init}}$ and $\pi_{2,\text{init}}$ need to be chosen in accordance with the theoretical treatment effect θ_{12} of treatment 2 relative to treatment 1. Using these initial event rates, log-odds are simulated independently for each trial i and arm k, as

$$\log\left(\frac{\pi_{k,i}}{1-\pi_{k,i}}\right) \sim \mathcal{N}\left(\log\left(\frac{\pi_{k,\text{init}}}{1-\pi_{k,\text{init}}}\right), \frac{\tau^2}{2}\right).$$

Arm- and trial-specific event rates are then back-calculated from the simulated log-odds. In this DGM, between-study variances are incorporated in both trial arms independently from each other, hence the name DGM "random," which we adapted from the work of Pateras et al.[18] Examples for the application of this DGM may, for instance, be found in the works of Hartung and Knapp,[20] Gonnermann et al,[21] Mathes and Kuss,[22] or Jackson et al.[23] However, this model is not applicable when multiarm trials are part of the simulation. In this case, an additional dependent structure between the relative treatment effects is introduced as the trial-specific baseline treatment is part of each simulated contrast from that trial.[24] This dependence contradicts the independence assumed in the DGM "random" and cannot be taken into account when simulating log-odds independently. As dependence and independence of treatment effects cannot be simultaneously incorporated but multiarm trials should be part of the simulated data, we will not consider this DGM any further.

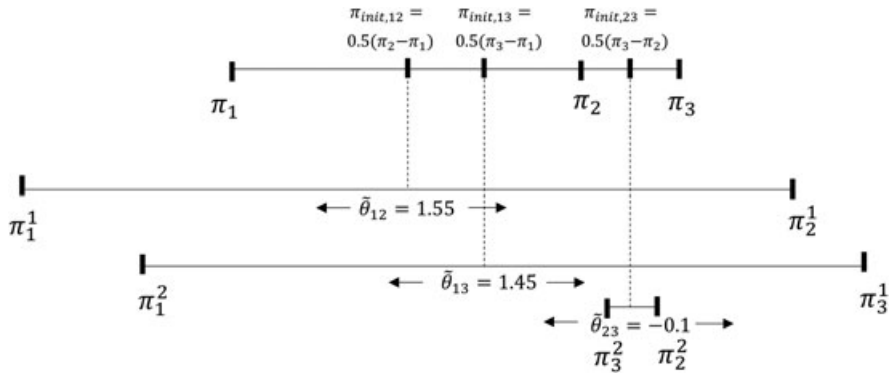

FIGURE 2 Initial situation in a three-armed trial (theoretical values: $\theta_{12} = 0.7, \theta_{13} = 0.9, \theta_{23} = 0.2, \tau^2 = 4$), and situation after one simulation draw $\tilde{\theta}_{12} = 1.55, \tilde{\theta}_{13} = 1.45, \tilde{\theta}_{23} = -0.1$

2.2 | DGM "average" applied to multiarm trials in a random-effects network meta-analysis

In the DGM "average," a theoretical average trial event rate (π_{init}) is assumed, often according to practical aspects. In the random-effects model, trial-specific treatment effects $\theta_{12,i}$ are then simulated around their theoretical value θ_{12} as

$$\log\left(\frac{\pi_{2,i}(1-\pi_{1,i})}{(1-\pi_{2,i})\pi_{1,i}}\right) = \theta_{12,i} \sim \mathcal{N}\left(\theta_{12}, \tau^2\right), \tag{1}$$

from which, the arm- and trial-specific event rates $\pi_{k,i}$ are then back-calculated using the initial (theoretical) value

$$\pi_{k,i} = \frac{\pi_{\text{init}} \exp(\pm 0.5\theta_{12,i})}{1 - \pi_{\text{init}} + \pi_{\text{init}} \exp(\pm 0.5\theta_{12,i})}.$$

This DGM is commonly applied in practice for pairwise meta-analysis.[25-28] However, it cannot be designed for the random-effects setting in the presence of multiarm trials. Depending on the contrast considered, different potential initial values could be chosen in each multiarm trial, resulting in different possibilities for the same event rate within one trial. This situation is illustrated in Figure 2, where a three-armed trial is generated with theoretical values $\theta_{12} = 0.7$, $\theta_{13} = 0.9$, and $\theta_{23} = 0.2$ and the (constant, but for exemplary reasons deliberately high) heterogeneity variance $\tau^2 = 4$. Taking one draw from the multivariate normal distribution, we arrive at $\tilde{\theta}_{12} = 1.55$, $\tilde{\theta}_{13} = 1.45$, and $\tilde{\theta}_{23} = -0.1$. Starting from the centered event rates between any two treatment arms ($\pi_{\text{init},12}$, $\pi_{\text{init},13}$, or $\pi_{\text{init},23}$), we obtain two different event rates for each treatment arm for the same simulated trial ($\pi_1^1, \pi_1^2, \pi_2^1, \pi_2^2, \pi_3^1, \pi_3^2$ in our example). As the number of possibilities for each event rate increases linearly with the number of trial-arms (one for a two-armed trial, two for a three-armed trial, three for a four-armed trial, etc), this DGM is not suitable for multiarm trials in a random-effects setting, as it is impossible to construct a data set that is, on average, in agreement with the theoretical values when multiarm trials are present. This DGM is therefore not considered any further.

2.3 | DGM "fixed" applied to multiarm trials in a random-effects network meta-analysis

The third DGM, requires in the pairwise case to pre-specify a range for the event rates in one of the two arms. Then, trial-specific event rates in this arm are drawn from a random distribution in this range (eg, uniform) and the trial-specific treatment effect is simulated as in Equation (1). The second event rate is then determined as

$$\pi_{2,i} = \frac{\pi_{1,i} \exp(\theta_i)}{1 - \pi_{1,i} + \pi_{1,i} \exp(\theta_i)}.$$

Examples for the application of this DGM in the pairwise case may be found in, eg, the works of Lambert et al[29] and Novianti et al.[30] It can be directly transferred to the simulation of multiarm trials, arbitrarily choosing one of the arms as starting point and pre-specifying a range in which the respective event rates lie. However, two strong assumptions are part of this DGM. By pre-specifying the range of potential values, it is (a) implicitly assumed that event rates in this arm are independent of the treatment effect, while those of the other arms are not, and that (b) the range of potential values is directly restricted in one arm, while all values between 0 and 1 are possible in the second arm. This has been observed to inflate the treatment effect, especially in presence of between-trial heterogeneity and small studies.[18] Additionally, the use

of a random distribution (eg, the uniform) to determine the exact trial-specific event rate for each repetition introduces an additional level of randomness that is not accounted for in meta-analytical methods.

The just described assumptions are even more restrictive when multiarm trials and ultimately network meta-analyses should be considered. When a trial includes more arms, contrasts between at least some of them tend to be larger than in pairwise meta-analysis. This would increase the inflation, described by Pateras et al,[18] of a potential simulated treatment effect as compared to the pairwise case. Additionally, it is not clear why and when it is suitable to assume one arm to be independent of the treatment effect and several others not. This may be of special interest in simulations of different network geometries, as in, eg, the work of Kibret et al.[11]

When considering larger networks with more complicated geometries, pre-defining ranges of event rates for multiple reference treatments might become deliberately complex. The assumptions underlying the DGM "fixed" might become even more problematic when unbalanced trials should be considered in the simulation study. Due to the dependence of the variance of a binomially distributed variable on its event rate and the group size, these choices influence the variation within a trial. As this variation is used in meta-analytical weights, a small inaccuracy in the simulation design might heavily influence the evaluation results of estimation methods.

2.4 | Modification of DGM "fixed"

For the reasons just described, we see the need of a modified version of the DGM "fixed" that avoids restricting any of the event rates and refrains from the additional level of variability introduced by simulating one event rate. As it is necessary to choose one system of event rates out of infinitely many possible ones, this choice may depend on the treatment effect(s). In addition, it is not necessary to restrict one event rate to a certain range of values. Due to the nonlinearity of the relation between event rates under constant ORs, it is possible to choose the combination of event rates that jointly fulfill an optimization criterion (and take the whole potential range for all event rates into account). We chose as optimization criterion the minimization of the sum of quadratic differences to 0.5

$$\min_{\tilde{\pi}_1} \sum_{j=1}^{K} (\tilde{\pi}_j - 0.5)^2 = \min_{\tilde{\pi}_1} \sum_{j=2}^{K} \left(\frac{\tilde{\pi}_1 \exp(\theta_{1j,i})}{1 - \tilde{\pi}_1 + \tilde{\pi}_1 \exp(\theta_{1j,i})} - 0.5 \right)^2 + (\tilde{\pi}_1 - 0.5)^2. \quad (2)$$

Pragmatic reasons led to this criterion, as we initially wished to obtain a simulated data set that minimizes imbalances in the trial-arm variances, as well as potential interference with unbalanced study sizes. We therefore use that the binomial distribution is symmetric around 0.5. Additionally, we want to prevent extreme event rates, limiting the amount of trials where our effect measure, the OR, is not defined and a continuity correction might be necessary. Hence, we chose a quadratic penalty term. Furthermore, we wanted to prevent additional (unaccounted) sources of variation (eg, an additional random draw from the uniform distribution) to be added into the synthetic data set. Lastly, we wanted to allow for variation in all arms of our simulated trials, without restricting one arm to a pre-defined range. As the above described criterion (2) is additive with respect to the number of trial-arms, it is easy and fast to optimize when more than three arms should be considered. Different choices for such an optimization criterion would also be possible and should depend on the characteristics the synthetic data should reflect. If, for example, methods for sparse data in network meta-analysis should be investigated, minimizing the sum of trial-arm variances could be a choice to obtain a suitable data set. For the proposed modification, we want to emphasize the increased flexibility of the DGM "fixed" modified in simultaneously computing event rates for all arms in a trial as compared to the DGM "fixed."

2.5 | Performance of DGM "fixed" and its modification in two-arm and multiarm trials

To compare the implications of the DGM "fixed" to its modified version, we conducted a simulation study including at first a two-armed trial and, subsequently, a three-armed trial. For sake of comparability in case of two-armed trials, we followed the work of Pateras et al[18] and varied the between-trial heterogeneity standard deviation ($\tau = \{0.001, 0.5, 1, 2\}$) and the number of patients per trial arm ($n_{k,i} = \{small \sim U(20, 30), large \sim U(230, 240)\}$). Two different log-odds ratios are simulated as theoretical treatment effects ($\theta_{12} = \{0, 1\}$) for the scenarios in the pairwise case, while we used the same theoretical treatment effects ($\theta_{12} = 0.7, \theta_{13} = 0.9, \theta_{23} = 0.2$) as before for the simulation of the three-armed trial. The event rate for the reference arm is again drawn from $U(0.4, 0.6)$ in all cases for the DGM "fixed." All computations were performed using `R`[31] and its extensions `tidyverse`,[32] `mvtnorm2018`,[33] and `ggplot2`.[34] Optimization for DGM "fixed" modified by means of `optimize` in R, which uses a combination of golden section search and successive parabolic

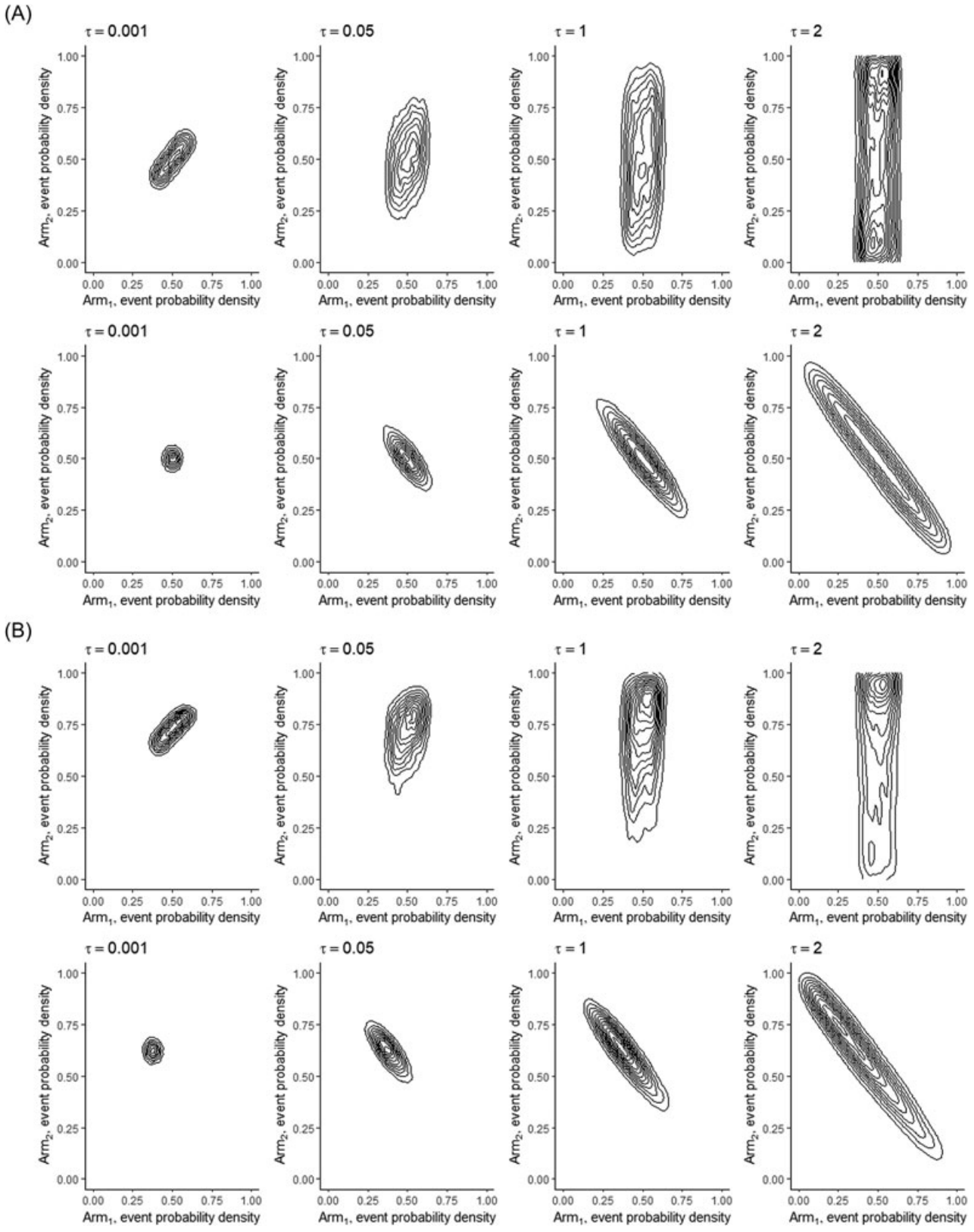

FIGURE 3 Joint densities for two treatment arms in a two-armed trial of binary data. (A) No treatment effect ($\theta_{12} = 0$). (B) With treatment effect ($\theta_{12} = 1$). The upper row of each pair corresponds to data-generating model (DGM) "fixed" and the lower row to DGM "fixed" modified. Shown are all values of τ, $\theta_{12} = \{0, 1\}$ for small studies with $n \sim U(20, 30)$, and $\pi_1 \sim U(0.4, 0.6)$ for DGM "fixed"

interpolation. Results for all investigated values of between-study heterogeneity for $n_{k,i} = small$ are shown in text, while those of $n_{k,i} = large$ are added as supplementary file.

The joint densities from the simulated events for the two-armed trial are illustrated in Figure 3 for the situation of no treatment effect (upper two rows) and an existing treatment effect (lower two rows) for heterogeneity variance increasing from left to right. The results from the two DGMs can be directly compared as the first row of a pair illustrates the result from DGM "fixed" and the second those of its modification. In this Figure, the diversity between these two DGMs with respect to the joint distribution of event rates $\tilde{\pi}_1$ and $\tilde{\pi}_2$ can be observed. The previously discussed restriction to a pre-specified range of potential values for the first event rate ($\tilde{\pi}_1$) can be clearly seen, especially when τ^2 is large. The simulated event rates stem from a $U(0.4, 0.6)$ for one of the arms in DGM "fixed." Only a small area of potential values for $\pi_{1,i}$ is covered. This is observed even though additional variation is introduced by estimating $\pi_{1,i}$ for the joint density through the number of events $x_{1,i}$ and patients $n_{1,i}$. The values of event rates for the DGM "fixed" modified, however, are observed to be simulated over the whole range of potential rates in both arms with more mass placed in areas where neither of the arms' event rates are extreme. On the other hand, the restriction of the range for the reference arms' event rate can be observed to be even stronger in the presence of a treatment effect in the DGM "fixed." It may be a suitable assumptions that event rates are centered around a true underlying event rate in cases where heterogeneity is small, which, in this situation, is in fact observed for both DGMs. However, it becomes implausible in scenarios with high variation between studies, where it is difficult to motivate why such heterogeneity may not be mirrored in the simulation strategy of the reference arm. It can be observed that, for the modified version of the DGM "fixed," both event rates can take on values over the whole possible range, with higher density being placed at combinations where neither of the event rates takes extreme values. It should be noted that, when only two arms are simulated, due to the symmetry of the criterion chosen, this DGM corresponds to the case where $p_{0,\text{init}} = 0.5$ for the DGM "average." Results from 10 000 simulation repetitions are reported in Table 1 for the extreme case of $\tau^2 = 4$ and $n_{k,i} \sim U(20, 30)$, and tables showing results for the respective other combinations can be found in the Appendix (see Tables A1, A3, and A5). As can be observed, the event rates obtained in both cases correspond to the theoretical ones. When the estimated event rates ($\frac{x_{k,i}}{n_{k,i}}$) are used, however, the treatment effects in the DGM "fixed" deviates further from the theoretical ones than those in its modified version, which might be due to the restricted flexibility to account for the varying $\theta_{12,i}$ and the higher number of simulation draws in which continuity corrections are necessary. The average empirical variance ($\bar{\hat{\tau}}^2$) corresponds to the theoretical value (4) in both DGMs, but the differences between the variances of the two arms are smaller in the modified version of the DGM "fixed." These results can also be observed (in less extreme form) in the simulated scenarios with larger trial-arm sizes (results shown in the supplementary files).

For the simulation of three-armed trials in network meta-analysis, we evaluate the same properties as in the two-armed case using the same parameters as for the two-armed trial, except for the treatment effect for which the values from the above discussed example (In Figure 2, values chosen are $\boldsymbol{\theta} = (\theta_{12}, \theta_{13})' = (0.7, 0.9)'$). In Figure 4, we illustrate the joint densities for each of the possible combinations of two of the three arms such that the first two rows show the results for arm 2 relative to arm 1, the third and fourth rows show results for arm 3 relative to arm 1, and the last two rows illustrate arm 3 relative to arm 2 with increasing heterogeneity from left to right. As in Figure 3, the first row of a pair shows results of the DGM "fixed" while the second shows the results of its modification. As in the two-armed case, one event rate is restricted in its range, while now, all variation is incorporated in two nonreference arms instead of one. Even though the event rates of these second and third nonreference arms can take on all possible values, the highest mass for this comparison is also

TABLE 1 Results for the two data-generating models (DGMs) for a two-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, and $\tau = 2$. ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"		DGM "Fixed" Modified	
	$\theta = 0$	$\theta = 1$	$\theta = 0$	$\theta = 1$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.0016	1.0134	0.0016	1.0134
$\bar{\tilde{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	−0.0421	0.8761	−0.0189	1.0738
No. of continuity corrections	1082	1377	160	217
$\bar{\tilde{\tau}}^2$	4.0756	3.9271	4.0756	3.9271
Mean variance of arm 1* (reference arm)	0.2369	0.2369	0.1912	0.1978
Mean variance of arm 2*	0.1473	0.1399	0.1978	0.1920

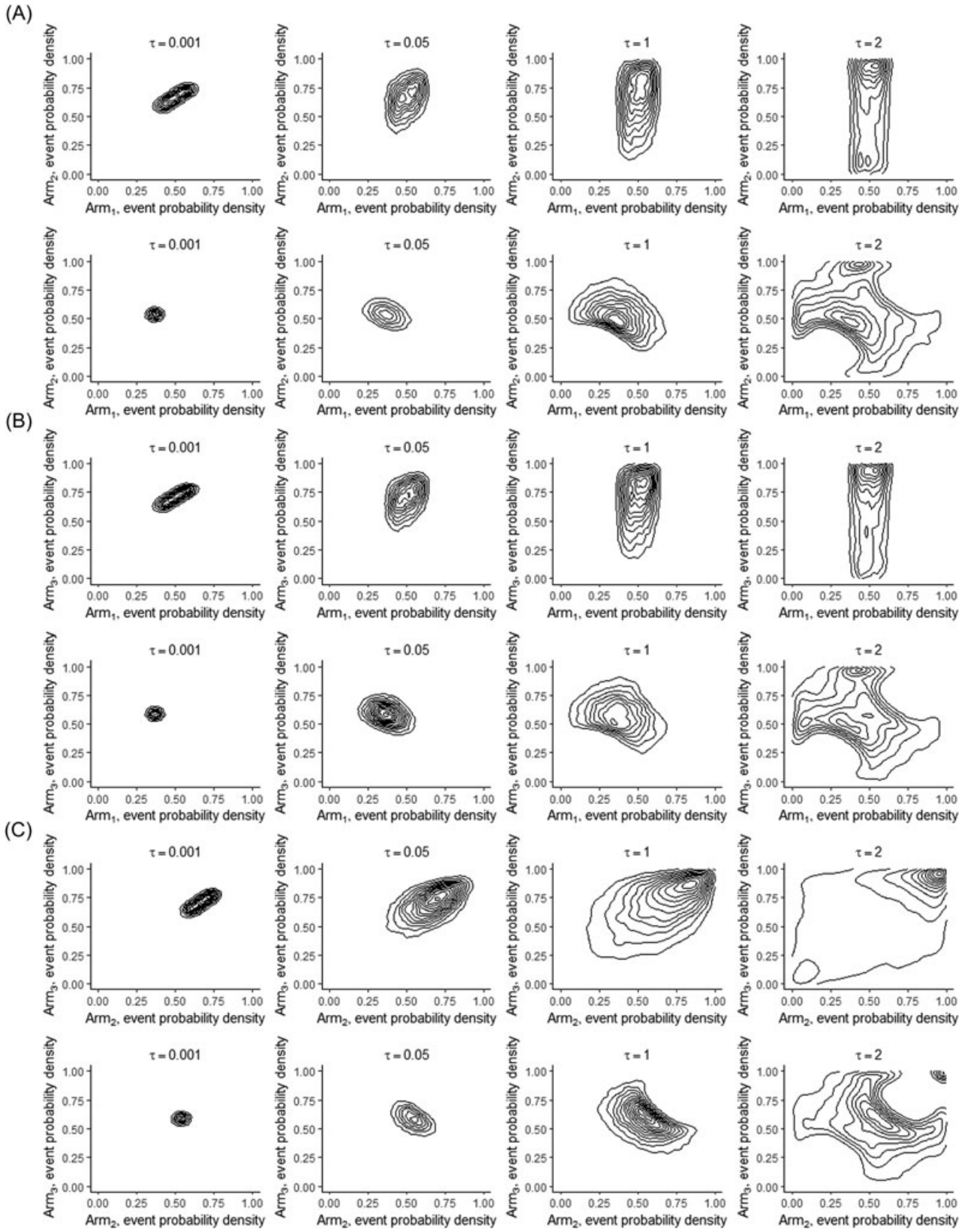

FIGURE 4 Joint densities for two treatment arms in a three-armed trial of binary data. (A) Arm 2 relative to arm 1 ($\theta_{12} = 0.7$). (B) Arm 3 relative to arm 1 ($\theta_{13} = 0.9$). (C) Arm 3 relative to arm 2 ($\theta_{23} = 0.2$). The upper row of each pair corresponds to data-generating model (DGM) "fixed" the lower row to DGM "fixed" modified. Shown are all values of τ for small studies with $n \sim U(20, 30)$, and $\pi_1 \sim U(0.4, 0.6)$ for DGM "fixed"

TABLE 2 Results for the two data-generating models (DGMs) for a three-armed trial, $n_{k,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, $\tau = 2$, ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"			DGM "Fixed" Modified		
	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.6979	0.8790	0.1811	0.6979	0.8790	0.1811
$\bar{\tilde{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	0.5968	0.7459	0.1491	0.6594	0.8314	0.1720
No. of continuity corrections	1245	1324	2248	841	936	1039
$\bar{\tilde{\tau}}^2$	3.9994	4.1129	4.0426	3.9994	4.1129	4.0426
Mean variance of arm 1* (reference arm)		0.2365			0.1949	
Mean variance of arm 2*		0.1453			0.1925	
Mean variance of arm 3*		0.1418			0.1882	

placed in areas where both event rates are jointly very extreme (Figure 4C). The modified version of the DGM "fixed" places more mass in less extreme values of event rates for all three arms and the distribution of the joint event rates lies more central in the possible area. However, as a symmetric solution is not possible when more than two arms are present, the shape of the joint density is slightly dented, the more so, when heterogeneity is increasing. Results from the simulation study for $\tau^2 = 4$ and $n_{ki} \sim U(20, 30)$ are reported in Table 2, while tables for the respective other combinations are shown in the Appendix (see Tables A2, A4, and A6).

As in the case of two-arms, the average empirical treatment effect calculated using the simulated event rates ($\pi_{k,i}$) corresponds to the theoretical values, while deviation from the theoretical values is larger when events and number of patients are used to estimate these event rates. Again, this might be due to the restricted flexibility inherent in the construction of event rates in DGM "fixed," as well as the high number of extreme values in either arm that makes continuity corrections necessary. As in the case of two-armed trials, these results can also be observed (in less extreme form) in simulated scenarios with larger trial-arm sizes (results shown in supplementary file).

3 | SIMULATION PROCEDURE FOR A NETWORK OF STUDIES

In this section, we propose a simulation procedure for network meta-analyses including multiarm trials. As the above described simulation of two- and three-armed trials included solely single studies due to illustrative purposes, we developed a simulation procedure to combine several different independent two-arm and multiarm trials into the desired network of studies. For better comparability between the DGM "fixed" and its modification, the network used is of "star-shaped" geometry, with a common reference treatment in all trials of the network (treatment 1 in our case). This assumption facilitates the simulation and presentation of results, as the range for one event rate has to be pre-specified in the DGM "fixed." However, it is not necessary for the generation of a data set using either data-generating mechanism as long as the network is connected.

The simulation procedure works similarly for both DGMs. We demonstrate the approach with the previously presented example from the work of Warren et al.[8] To construct the network of eight treatments, a vector of theoretical (pairwise) treatment effects ($\theta_{1k\neq1}$), all relative to the common reference treatment 1, are determined. These are, together with two vectors of treatment IDs (one for each treatment in the contrast), needed to construct the theoretical network (compare Figure 1). We further need $nt_{1k\neq1}$, the number of trials per respective contrast, $n_{k,i}$, the number of patients per trial-arm, the contrast-specific between-trial variance τ^2_{1k}, and the study ID with which the presence of multiarm trials may be specified in the theoretical network of trials. We use the values empirically observed in the work of Warren et al[8] as underlying theoretical values that are summarized in Table 3. The trial-arm–specific numbers of "patients" could also be simulated (as for the DGMs) if, eg, the impact of imbalances in trial-arms should be evaluated.

For sake of simplicity, we set the contrast-specific heterogeneity variances to a constant, τ within one simulated network, but vary this constant over our scenarios using the same values as above ($\tau = \{0.001, 0.5, 1, 2\}$). This simplification is not necessary for the simulation procedure and may be relaxed when necessary. Setting the heterogeneity to a constant, we obtain the nondiagonal elements of the covariance matrix Σ as $0.5\tau^2$ to be $\{0.0000005, 0.125, 0.5, 2\}$.[16] This simplification seems suitable in our case, as we are focusing mainly on the generation of (armwise) event rates from treatment effects. If necessary, more complex covariance structures could be easily implemented in the simulation of treatment effects

TABLE 3 Parameters used in the simulation illustrating the differences between data-generating model (DGM) "fixed" and DGM "fixed" modified correspond to characteristics reported in the work of Warren et al[8]

Study ID (i)	Reference ($k = 1$)	Treatment ($k \neq 1$)	θ_{1k}	$n_{1,i}$	$n_{k,i}$
Maini (2004)	1	2	−0.13	88	86
St. Clair (2004)	1	2	−0.13	291	372
Maini (2004)	1	3	2.30	88	254
St. Clair (2004)	1	3	2.30	291	377
Ericson (1999)	1	4	0.89	105	111
Moreland (1999)	1	4	0.89	80	78
Genovese (2002)	1	4	0.89	217	207
Combe (2006)	1	4	0.89	50	204
Van der Heijde (2006)	1	4	0.89	228	454
Weisman (2007)/Baumgartner (2004)	1	4	0.89	269	266
Ericson (1999)	1	5	0.82	105	343
Moreland (1999)	1	5	0.82	80	76
Genovese (2002)	1	5	0.82	217	208
Furst (2003)	1	6	0.81	318	318
Weinblatt (2003)	1	6	0.81	62	67
Keystone (2004)	1	6	0.81	200	419
Van de Putte (2004)	1	6	0.81	110	225
Breedveld (2006)	1	6	0.81	257	542
Weinblatt (2003)	1	7	0.29	62	73
Van de Putte (2004)	1	7	0.29	110	103
Weinblatt (2003)	1	8	1.41	62	69
Van de Putte (2004)	1	8	1.41	110	106

without affecting the generation of event rates from those treatment effects. We simulated $r = 10\,000$ repetitions for each scenario. Starting with these parameter values, we used the following steps to simulate a network of studies.

1. Drawing of treatment effects as contrast relative to a trial-specific baseline treatment $k = 1$ trialwise as $\theta_{1\mathbf{k},\mathbf{i}} \sim \mathcal{MVN}(\theta_{1\mathbf{k}}, \Sigma)$ for all nonbaseline arms $k \neq 1$ in trial i.
2. Depending on the DGM used, for each trial i, either
   - "fixed": drawing of event rates for treatment 1 from a uniform distribution $\pi_{1,i} \sim U(0.4, 0.6)$ and using $\pi_{k,i} = \frac{\pi_{1,i}\exp(\theta_{1k,i})}{1-\pi_{1,i}+\pi_{1,i}\exp(\theta_{1k,i})}$ to obtain the event rate $\pi_{k,i}$ for all nonbaseline arms $k \neq 1$ in trial i.
   - "fixed" modified: expressing the event rates of nonreference arms $\pi_{k,i}$ in dependence of $\pi_{1,i}$ as $\pi_{k,i}(\pi_{1,i}) = \frac{\pi_{1,i}\exp(\theta_{1,i})}{1-\pi_{1,i}+\pi_{1,i}\exp(\theta_{1k,i})}$ for all nonbaseline arms $k \neq 1$ in trial i and use $\min_{\tilde{\pi}_1} \sum_{k=1}^{k} (\tilde{\pi}_j - 0.5)^2$ to find all $\pi_{k,i}$ simultaneously.
3. Simulation of trial-arm–specific number of events as $x_{k,i} \sim \text{Bin}(\pi_{k,i}, n_{k,i})$.
4. Repeat Steps 1 to 3 the needed number of times to obtain r repetitions of the data set.

Code for the simulation of the described network meta-analysis is provided in the Appendix and the results for $\tau^2 = 4$ are summarized in Table 4. Results are reported for each trial and treatment effect combination separately to avoid naive pooling with strongly varying trial-arm sizes. As can be observed, averages and variances from the higher level of hierarchy in the simulation (before drawing arm-specific event rates from the binomial distribution) are reasonably close to the theoretical ones. The number of simulation draws where a continuity correction is needed in any of the trial treatment effect combinations is still higher in the DGM "fixed" than in its modified version, but when taking the increased number of patients into consideration, not as large as in the assessment of single two- and three-armed trials. This might be due to the partly larger trial-arm sizes in the empirical data set with which the extreme cases of no event or all events in arms are avoided, even when event rates are close to the extreme values. Differences between the two DGMs can be observed when considering the mean event rates for the trial-arms, which are by construction more symmetric around 0.5 in case of the DGM "fixed." Providentially, the average event rate for the reference arm (treatment 1) in the modified version of the DGM "fixed" falls into similar areas in all trial-treatment combinations, but varies on average in both arms, while

TABLE 4 Results of the two data-generating model (DGMs) for a network meta-analysis including three-armed trials, * = using continuity correction, † = results from network meta-analysis using Bayesian estimation and the model described in the work of Dias et al[6]

	Results Applicable to Both DGMs Higher Level of Hierarchy				DGM "Fixed" No. of Continuity Corrections = 5650					DGM "Fixed" Modified No. of Continuity Corrections = 4935				
	$\bar{\tilde{\theta}}$ from $\tilde{\pi}_{k,i}$	n_1	n_2	$\tilde{\tau}^2$ from $\tilde{\pi}_{k,i}$	$\bar{\pi}_1; \bar{\pi}_2$	$\bar{\tilde{\theta}}$ from $\frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}$*	$\hat{\tilde{\tau}}^2$	$\hat{\theta}^\dagger$	$\lvert\theta - \hat{\theta}^\dagger\rvert$	$\bar{\pi}_1; \bar{\pi}_2$	$\bar{\tilde{\theta}}$ from $\frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}$*	$\hat{\tilde{\tau}}^2$	$\hat{\theta}^\dagger$	$\lvert\theta - \hat{\theta}^\dagger\rvert$
$\theta_{12} = -0.1267$	−0.1225	88	86	3.9956	0.50; 0.48	−0.1328	4.1770	−0.1433	1.1437	0.43; 0.41	−0.1241	4.1273	−0.1340	1.1315
	−0.1489	291	372	3.9355	0.50; 0.47	−0.1548	4.0840			0.43; 0.41	−0.1515	4.0572		
$\theta_{13} = 2.3026$	2.2982	88	254	3.9771	0.5; 0.81	2.2676	3.7950	2.3755	1.2172	0.43; 0.77	2.2818	3.9421	2.3359	1.1617
	2.3003	291	377	4.0011	0.50; 0.81	2.2910	3.8776			0.43; 0.77	2.2891	3.8977		
$\theta_{14} = 0.8850$	0.9056	105	111	3.9158	0.50; 0.63	0.886	3.9781	0.8737	0.6444	0.41; 0.57	0.8960	4.0142	0.8741	0.6446
	0.8850	80	78	4.0034	0.50; 0.63	0.8594	3.9993			0.41; 0.56	0.8739	4.0308		
	0.8498	217	207	3.9432	0.50; 0.63	0.8512	4.0611			0.42; 0.56	0.8495	4.0424		
	0.8603	50	204	3.9273	0.50; 0.63	0.8604	4.0892			0.45; 0.60	0.8601	4.0641		
	0.9147	228	454	3.8745	0.50; 0.64	0.9164	3.9582			0.45; 0.60	0.9197	3.9588		
	0.8477	269	266	3.9113	0.45; 0.63	0.8501	3.9831			0.45; 0.60	0.8458	3.9450		
$\theta_{15} = 0.8237$	0.8213	105	343	3.9925	0.50; 0.62	0.8310	4.1369	0.8154	0.8925	0.41; 0.56	0.8355	4.2005	0.8145	0.8744
	0.8251	80	76	3.9241	0.50; 0.63	0.8002	3.8989			0.41; 0.56	0.8167	4.0209		
	0.7857	217	208	3.9339	0.50; 0.61	0.7845	4.0514			0.42; 0.55	0.7850	3.9339		
$\theta_{16} = 0.8123$	0.8168	318	318	3.9916	0.50; 0.62	0.8169	4.1072	0.8110	0.7912	0.45; 0.60	0.8152	4.0595	0.8090	0.7184
	0.7811	62	67	3.9035	0.50; 0.61	0.7465	3.9203			0.41; 0.54	0.7756	4.1218		
	0.8306	200	419	3.9446	0.50; 0.62	0.8341	4.0293			0.45; 0.60	0.8307	4.0191		
	0.81048	110	225	4.1805	0.50; 0.62	0.8086	4.3153			0.40; 0.54	0.8108	4.3839		
	0.7865	257	542	4.0207	0.50; 0.63	0.8309	4.0956			0.46; 0.60	0.8267	4.0794		
$\theta_{17} = 0.2897$	0.2763	62	73	3.9531	0.50; 0.54	0.2625	4.1113	0.3102	1.1253	0.41; 0.45	0.2711	4.3020	0.2943	1.0801
	0.2972	110	103	4.1223	0.50; 0.54	0.2886	4.2470			0.40; 0.45	0.3050	4.4092		
$\theta_{18} = 1.4134$	1.3933	62	69	3.9532	0.50; 0.70	1.3433	3.7585	1.5540	1.2223	0.41; 0.64	1.3816	4.0405	1.4551	1.0981
	1.4158	110	106	4.0009	0.50; 0.70	1.3842	3.8647			0.40; 0.64	1.4060	4.0394		

variation is only observed in the event rates of the nonreference arm in DGM "fixed." Empirical averages of simulated treatment effects (on the log-odds ratio scale), estimated through numbers of events and patients ($x_{k,i}$ and $n_{k,i}$), are on average closer to the theoretical values for the modified version than for the DGM "fixed" even though not as clearly as in the previous simulations. This might also be due to the larger trial-arm sizes. The average empirical variances are in most cases slightly smaller using the DGM "fixed" modified, which can also be observed (in less extreme form) in simulated scenarios with smaller heterogeneity variance (results not shown). To illustrate the differences in potential simulation studies, we report the combined treatment effects estimated in a network meta-analysis using the standard binomial-normal consistency model and Bayesian estimation in the default setting as reported in the work of Dias et al.[6] The estimation was based on 3 chains, a chain length of 25 000 observations, a burn-in of 5000 and a thinning factor of 10. Even though the same theoretical treatment effects and exactly the same draws from a multivariate-normal distribution are used, and the estimation method is similar in both cases, the average estimated treatment effect differs slightly between the two data-generating mechanisms. The estimated treatment effects are comparable between the two DGM mechanisms for some contrasts ($\theta_{14}, \theta_{15}, \theta_{15}$). However, they are on average slightly closer to the theoretical values in the other contrast on the data set simulated with DGM "fixed" modified. The mean absolute deviation from the theoretical values are smaller in the data set generated with DGM "fixed" modified, except for one contrast, in which it starts to deviate in the fourth decimal point.

4 | DISCUSSION

The evaluation and comparison of statistical methods by simulation studies is state of the art in many fields of research.[35] To our knowledge, however, no disseminated and generally accepted simulation procedure exists for random-effects network meta-analysis with binary outcome when multiarm trials should be part of the simulated data. As this is a demanded scenario in many meta-analytical applications, we evaluated whether existing DGMs for pairwise meta-analysis can be extended to our case. Starting with the three DGMs evaluated in the work of Pateras et al,[18] we found that only the DGM "fixed" may be directly applied in the simulation of multiarm trials. This DGM, however, assumes the event rate of one treatment to be independent of the treatment effect and restricted to a pre-defined range. These assumptions are already restrictive in the pairwise case, but become even more problematic when considering a network of studies where the choice of the independent treatment and its range are not always obvious. Especially when the geometry of the simulated network is not star-shaped (ie, when there is no common reference treatment evaluated in each trial), several different choices of the reference treatments in different trials, as well as different ranges of event rates for those reference treatments might be possible. Despite the, somehow, strange assumption that the event rate for the same treatment may be assumed to be independent of treatment effect in some cases (when it is considered to be the reference treatment of a trial) and independent in other trials, it is not clear how the choice of the said range and the additional random draw could influence results in simulations.

To avoid the above discussed assumptions and choices, we proposed a modification to the DGM "fixed" in which all event rates of a (multiarm) trial are simultaneously evaluated. Resulting event rates in the modification to DGM "fixed" do, therefore, not depend on the choice of the reference treatment in the trial and do neither require to restrict any of the event rates in their range. To obtain event rates with preferable properties, we used the fact that the relation between two event rates that result in a constant OR is not linear. Depending on the application of the simulation study, different optimization criteria could possibly be used. The choice of the exact criterion presented is rather a pragmatic one, which is driven by the wish to lessen the extent of unbalancedness of event rates (which influence the variances in the individual trial arms and, therefore, potentially, the meta-analytic weights) and at the same time avoid the need of continuity corrections. Additionally, it is additive with respect to the number of arms and therefore fast and easy to optimize, irrespective of the number of arms in the trial.

Using DGM "fixed" and its presented modification, we developed a simulation procedure for a random-effects network of binary data that includes multiarm trials. We used the simplifying assumption of common and constant between-study heterogeneity here, which results in a simple covariance structure. There are examples where such a simplification might be not suitable. Hong et al,[36] for example, discuss network meta-analytical models under differing misingness patterns, for which a more complicated covariance structure in the simulation of treatment effects is necessary. Such differing covariance structures could be easily implemented into the simulation of treatment effects without affecting the determination of event rates in the second step.

Taking an empirical network meta-analysis as motivating example, we assess the simulated data set with respect to comparability to the theoretical values, number of draws where a continuity correction is needed, and symmetry of trial arm variances. Our simulation is not extensive and does not cover effects in other settings, eg, when networks are only sparsely connected or when the assumption of constant between-trial variance over all contrast is violated. Nonetheless, we observed that the modified version of the DGM "fixed" shows good results with respect to the evaluated properties.

The elaboration of the simulation procedure and the potentially underlying DGM articulates one of the crucial conceptual difficulties in the simulation of hierarchical random-effects models with correlated effects. Even though both approaches are designed to result in the same theoretical effect, the properties of the modeled joint empirical distributions can differ drastically and potentially result in differing conclusions about meta-analytical methods. We therefore agree with Pateras et al[18] in the sense that the exact DGM should be reported in detail as part of methodological reviews. We believe that the common practice of using simulation studies to assess and compare properties of established and newly developed statistical methods should be extended to the field of random-effects network meta-analysis of binary data and are confident that the proposed modification of the DGM "fixed" and the proposed simulation procedure might serve as suitable starting points. We believe that this procedure is a solid foundation (and can be adapted for other specific meta-analytical questions) for further simulation studies critically investigating network meta-analytical methods for binary endpoints.

ACKNOWLEDGEMENTS

The authors would like to thank Gerta Rücker (Medical Center - University of Freiburg) for the helpful discussion on the simulation of binary random-effects meta-analyses in general, which encouraged us to dive deeper into this topic. In addition, the authors want to thank two anonymous reviewers, who committed their time and efforts to significantly improving and clarifying this manuscript, for their very helpful comments and questions.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this published article (and its supplementary information files) by providing the R Code to generate the synthetic data.

ORCID

Svenja E. Seide https://orcid.org/0000-0002-9113-7373
Katrin Jensen https://orcid.org/0000-0002-5088-5279
Meinhard Kieser https://orcid.org/0000-0003-2402-4333

REFERENCES

1. Lumley T. Network meta-analysis for indirect treatment comparisons. *Statist Med*. 2002;21(16):2313-2324.
2. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101(474):447-459.
3. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Syn Methods*. 2012;3(2):80-97.
4. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol*. 2014;67(2):138-143.
5. Greco T, Biondi-Zoccai G, Saleh O, et al. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. *Heart Lung Vessels*. 2015;2(7):133-142.
6. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network Meta-Analysis for Decision-Making*. Oxford, UK: Wiley; 2018. *Statistics in Practice*.
7. Franchini AJ, Dias S, Ades AE, Jansen JP, Welton NJ. Accounting for correlation in network meta-analysis with multi-arm trials. *Res Syn Methods*. 2012;3(2):142-160.
8. Warren FC, Abrams KR, Sutton AJ. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. *Statist Med*. 2014;33(14):2449-2466.

9. Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value Health*. 2015;18(1):116-126.
10. Günhan B, Friede T, Held L. A design-by-treatment interaction model for network meta-analysis and meta-regression with integrated nested Laplace approximations. *Res Syn Methods*. 2018;9(2):179-194.
11. Kibret T, Richer D, Beyene J. Bias in identification of the best treatment in a Bayesian network meta-analysis for binary outcome: a simulation study. *Clinical Epidemiology*. 2014;6:451-460.
12. Uhlmann L, Jensen K, Kieser M. Bayesian network meta-analysis for cluster randomized trials with binary outcomes. *Res Syn Methods*. 2017;8(2):236-250.
13. Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med Res Methodol*. 2012;138(12).
14. Mills EJ, Ghement I, O'Regan C, Thorlund K. Estimating the power of indirect comparisons: a simulation study. *PLoS ONE*. 2011;6(1).
15. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Chichester, UK: Wiley; 2011.
16. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statist Med*. 1996;15(24):2733-2749.
17. Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Syn Methods*. 2012;3(4):312-324.
18. Pateras K, Nikolakopoulos S, Roes K. Data-generating models of dichotomous outcomes: heterogeneity in simulation studies for a random-effects meta-analysis. *Statist Med*. 2018;37(7):1115-1124.
19. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Statist Med*. 2006;25(24):4279-4292.
20. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statist Med*. 2001;20(24):3875-3889.
21. Gonnermann A, Framke T, Großhennig A, Koch A. No solution yet for combining two independent studies in the presence of heterogeneity. *Statist Med*. 2015;34(16):2476-2480.
22. Mathes T, Kuss O. A comparison of methods for meta-analysis of a small number of studies with binary outcomes. *Res Syn Methods*. 2018.
23. Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Statist Med*. 2018;37(7):1059-1085.
24. Rücker G, Cates CJ, Schwarzer G. Methods for including information from multi-arm trials in pairwise meta-analysis. *Res Syn Methods*. 2017;8(4):392-403.
25. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 2014;14(1):1-12.
26. Veroniki AA, Mavridis D, Higgins JPT, Salanti G. Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. *BMC Med Res Methodol*. 2014;14(1):1-12.
27. Friede T, Röver C, Wandel S, Neuenschwander B. Meta-analysis of few small studies in orphan diseases. *Res Syn Methods*. 2017;8(1):79-91.
28. Seide SE, Röver C, Friede T. Likelihood-based random-effects meta-analysis with few studies: empirical and simulation studies. *BMC Med Res Methodol*. 2019;19(1):16.
29. Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statist Med*. 2005;24(15):2401-2428.
30. Novianti PW, Roes KCB, van der Tweel I. Estimation of between-trial variance in sequential meta-analyses: a simulation study. *Contemp Clin Trials*. 2014;37(1):129-138.
31. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org/
32. Wickham H. tidyverse: Easily Install and Load the 'tidyverse'. R package version 1.2.1. 2017. https://CRAN.R-project.org/package=tidyverse
33. Genz A, Bretz F, Miwa T, et al. mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-8. 2018. https://CRAN.R-project.org/package=mvtnorm
34. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer; 2009. http://ggplot2.org
35. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. arXiv:1712.03198. 2017.
36. Hong H, Chu H, Zhang J, Carlin BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Syn Methods*. 2016;7(1):6-22.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Seide SE, Jensen K, Kieser M. Simulation and data-generation for random-effects network meta-analysis of binary outcome. *Statistics in Medicine*. 2019;38:3288–3303. https://doi.org/10.1002/sim.8193

APPENDIX

TABLE A1 Results for the two data-generating models (DGMs) for a two-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, and $\tau = 0.001$. ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"		DGM "Fixed" Modified	
	$\theta = 0$	$\theta = 1$	$\theta = 0$	$\theta = 1$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.0000	1.0000	0.000	1.0000
$\bar{\hat{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	0.0052	1.0474	−0.0018	1.0348
No. of continuity corrections	0	12	0	0
$\bar{\tilde{\tau}}^2$	0.0000	0.0000	0.0000	0.0000
Mean variance of arm 1* (reference arm)	0.2370	0.2370	0.2399	0.2256
Mean variance of arm 2*	0.2366	0.1882	0.2401	0.2261

TABLE A2 Results for the two data-generating models (DGMs) for a three-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, $\tau = 0.001$, ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"			DGM "Fixed" Modified		
	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.7000	0.9000	0.2000	0.7000	0.9000	0.2000
$\bar{\hat{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	0.7429	0.9466	0.2017	0.7377	0.9475	0.2098
No. of continuity corrections	0	3	3	0	0	0
$\bar{\tilde{\tau}}^2$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Mean variance of arm 1* (reference arm)		0.2365			0.2228	
Mean variance of arm 2*		0.2107			0.2384	
Mean variance of arm 3*		0.1963			0.2324	

TABLE A3 Results for the two data-generating models (DGMs) for a two-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, and $\tau = 0.5$. ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"		DGM "Fixed" Modified	
	$\theta = 0$	$\theta = 1$	$\theta = 0$	$\theta = 1$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.0004	1.0033	0.0004	1.0033
$\bar{\hat{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	−0.0100	1.0477	0.0015	1.0044
No. of continuity corrections	4	56	0	2
$\bar{\tilde{\tau}}^2$	0.2547	0.2454	0.2547	0.2454
Mean variance of arm 1* (reference arm)	0.2366	0.2238	0.2362	0.2226
Mean variance of arm 2*	0.2369	0.1840	0.2365	0.2231

TABLE A4 Results for the two data-generating models (DGMs) for a three-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, $\tau = 0.5$, ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"			DGM "Fixed" Modified		
	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.7006	0.8906	0.1901	0.7006	0.8906	0.1901
$\bar{\hat{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	0.7527	0.9498	0.1971	0.7455	0.9423	0.1968
No. of continuity corrections	26	47	73	3	3	0
$\bar{\tilde{\tau}}^2$	0.2514	0.2515	0.2459	0.2514	0.2515	0.2459
Mean variance of arm 1* (reference arm)		0.2367			0.2186	
Mean variance of arm 2*		0.2022			0.2345	
Mean variance of arm 3*		0.1960			0.2288	

TABLE A5 Results for the two data-generating models (DGMs) for a two-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, and $\tau = 1$. ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"		DGM "Fixed" Modified	
	$\theta = 0$	$\theta = 1$	$\theta = 0$	$\theta = 1$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.0008	1.0067	0.0008	1.0067
$\bar{\tilde{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	−0.0059	1.0393	−0.0025	1.0056
No. of continuity corrections	86	331	4	10
$\bar{\tilde{\tau}}^2$	1.0189	0.9818	1.0189	0.9818
Mean variance of arm 1* (reference arm)	0.2370	0.2370	0.2263	0.2153
Mean variance of arm 2*	0.1965	0.1711	0.2262	0.2150

TABLE A6 Results for the two data-generating models (DGMs) for a three-armed trial, $n_{j,i} \sim U(20, 30)$, $\pi_1 \sim U(0.4, 0.6)$, $\tau = 1$, ˜ = simulated value, ^ = estimated value, ¯ = average value, * = using continuity correction

	DGM "Fixed"			DGM "Fixed" Modified		
	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$	$\theta_{12} = 0.7$	$\theta_{13} = 0.9$	$\theta_{23} = 0.2$
$\bar{\tilde{\theta}}$ using $\tilde{\pi}_{1,i}$ and $\tilde{\pi}_{2,i}$	0.6929	0.8959	0.2030	0.6929	0.8959	0.2030
$\bar{\tilde{\theta}}$ using $\hat{\tilde{\pi}}_{k,i} = \frac{\tilde{x}_{k,i}}{\tilde{n}_{k,i}}^*$	0.7161	0.9281	0.2120	0.6994	0.8914	0.1920
No. of continuity corrections	207	248	440	77	88	35
$\bar{\tilde{\tau}}^2$	1.0013	1.0007	1.0037	1.0013	1.0007	1.0037
Mean variance of arm 1* (reference arm)		0.2367			0.2088	
Mean variance of arm 2*		0.1850			0.2237	
Mean variance of arm 3*		0.1758			0.2173	