A model for meta-analysis of correlated binary outcomes: the case of split-body interventions

Journal: Statistical Methods in Medical Research

Manuscript ID: SMM-17-0181.R1

Manuscript Type: Original Article

Keywords: correlated outcomes, rare events, safety outcomes, bivariate binomial, Sarmanov family

Abstract:

In several areas of clinical research it is common for trials to assign different sites of the participants’ bodies to different interventions. For example, a randomized controlled trial (RCT) comparing surgical techniques for correcting myopia may randomize each eye of a participant to a different operation. Under such bilateral ('split-body') interventions, the observations from each participant are correlated. It is challenging to account for these correlations at the meta-analysis level, especially when the outcome is rare. Here we present a meta-analysis model based on the bivariate binomial distribution. Our model can synthesize studies on patients who received one intervention at one body site, patients who received two interventions at different sites, or a mixture of these two groups. The model can analyze studies with zero events in one or both treatment arms and can handle the case of incomplete data reporting. We use simulations to assess the performance of our model and to compare it with the bivariate beta-binomial model. In the case of bilateral interventions our model performed well and outperformed the bivariate beta-binomial model in all scenarios explored. We illustrate our methods using two previously published meta-analyses from the fields of orthopaedics and ophthalmology. We conclude that our model constitutes a useful new tool for the meta-analysis of binary outcomes in the presence of split-body interventions.
A model for meta-analysis of correlated binary outcomes: the case of split-body interventions

Orestis Efthimiou¹, Dimitris Mavridis²,³, Adriani Nikolakopoulou¹, Gerta Rücker⁴,
Sven Trelle¹,⁵, Matthias Egger¹ and Georgia Salanti¹

¹ Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
² Department of Primary Education, University of Ioannina, Ioannina, Greece
³ Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité (CRESS-UMR1153), Inserm / Université Paris Descartes
⁴ Institute for Medical Biometry and Statistics, Medical Faculty and Medical Center - University of Freiburg, Germany
⁵ Clinical Trials Unit (CTU) Bern, University of Bern, Bern, Switzerland

Abstract

In several areas of clinical research it is common for trials to assign different sites of the participants’ bodies to different interventions. For example, a randomized controlled trial (RCT) comparing surgical techniques for correcting myopia may randomize each eye of a participant to a different operation. Under such bilateral (‘split-body’) interventions, the observations from each participant are correlated. It is challenging to account for these correlations at the meta-analysis level, especially when the outcome is rare. Here we present a meta-analysis model based on the bivariate binomial distribution. Our model can synthesize studies on patients who received one intervention at one body site, patients who received two
interventions at different sites, or a mixture of these two groups. The model can analyze studies with zero events in one or both treatment arms and can handle the case of incomplete data reporting. We use simulations to assess the performance of our model and to compare it with the bivariate beta-binomial model. In the case of bilateral interventions our model performed well and outperformed the bivariate beta-binomial model in all scenarios explored. We illustrate our methods using two previously published meta-analyses from the fields of orthopaedics and ophthalmology. We conclude that our model constitutes a useful new tool for the meta-analysis of binary outcomes in the presence of split-body interventions.

Keywords: correlated outcomes; rare events; safety outcomes; bivariate binomial; Sarmanov family

1 Introduction

Correlated outcome data can occur in trials when multiple related outcomes are measured, when a single outcome is measured at multiple time-points or when patients receive more than one intervention. Typical examples of the latter category are the cross-over trials and the within-person randomized trials\(^1\). For example, outcomes from patients receiving different surgical operations for myopia in each of their eyes or different carpal-tunnel release methods in each of their arms are correlated. We will refer to this kind of interventions as ‘split-body’ or ‘bilateral’ interventions. At the meta-analysis level, disregarding the correlation induced by bilateral interventions and analyzing observations as if they belong to different patients may affect the precision of the pooled estimates\(^2\text{-}^6\).

For the case of binary outcomes, if studies with bilateral interventions report the full, ‘cross-classified’ information (number of events by patients \textit{and} treatments) one can readily account for the correlations. This cross-classified information is however usually unavailable to the meta-analysts. If not available, an assumed value of the correlation between the arm-specific effects (e.g. the log-odds) can be imputed and the data can be synthesized using a bivariate normal likelihood.
While this approach might be reasonable for efficacy outcomes, it is problematic for safety outcomes. Safety outcomes are often rare and in this case the normal approximation performs poorly while the estimation of the between-study variance (heterogeneity) becomes challenging. Moreover, the normal approximation of the log-odds or log-risks cannot be readily used in the presence of studies with zero events in one (‘single-zero’ studies) or both arms (‘double-zero’ studies). Such studies are frequently encountered in meta-analyses; in a sample of 500 Cochrane reviews, 30% included at least one single-zero, while 34% of these reviews included at least one meta-analysis with a double-zero study. In such cases researchers usually resort to performing a ‘continuity correction’ i.e. they add a factor in each cell of the corresponding two-by-two table. The choice of the factor, however, may heavily affect the meta-analysis estimates and might lead to paradoxical results. Double-zero studies are typically excluded from meta-analysis.

Alternative meta-analysis models that do not require a continuity correction and do not exclude studies have been suggested. These include generalized linear mixed models, using the arcsine difference, a Poisson-Normal model, a zero-inflation model, a Poisson-Gamma model, Firth’s logistic regression model and a bivariate beta-binomial model (sometimes called Sarmanov beta-binomial model). Kuss reviewed and compared ten different methods for meta-analysing rare events in a simulation study and recommended the use of the beta-binomial model for the meta-analysis of rare events. None of the aforementioned methods, however, can account for correlations induced by bilateral interventions.

Although there are methods available for meta-analysing correlated binary outcomes, it is not clear whether they are appropriate when the outcome is rare and when reporting is incomplete. In this paper we propose a new model fitted within a Bayesian framework that can be used to meta-analyse studies of unilateral design (each patient received one intervention in one body part), bilateral design (all patients received two interventions in different body parts) or a mixed design (where some patients received one
Our model uses a bivariate binomial likelihood that explicitly accounts for the correlations induced by the presence of bilateral interventions without employing the normal approximation. Our model can accommodate a meta-analysis in the presence of rare events, without excluding single- or double-zero studies from the dataset, and can handle the case of incomplete reporting in the original studies, i.e. when cross-classified information is not provided.

The paper is structured as follows. In Section 2 we describe the datasets from orthopaedics and ophthalmology we use to exemplify our methods. We chose these particular examples because they include studies of different designs and they cover both the case of rare and common outcomes. In Section 3 we present our model, and discuss how it can be applied for the case of studies of different designs. In Section 4 we present results from simulations that we conducted in order to assess the performance of our model and to compare it with the previously recommended beta-binomial model. In Section 5 we present results from the examples. Finally, in the Discussion section we highlight our main findings, summarize the strengths and weaknesses of our approach and discuss possible extensions.

2 Example datasets

2.1 Two surgical treatments for carpal tunnel syndrome

Our first example comes from a Cochrane review\textsuperscript{22} that compared the endoscopic release vs. any other open surgical intervention for carpal tunnel syndrome (CTS). CTS is a painful compression of a nerve at the root of the hand. One of the safety outcomes was the occurrence of minor complications (such as numbness around the incision or superficial infection) reported in 24 studies. Minor complications were rare: 5 out of 24 studies were single-zero and 4 were double-zero. In Section 1.1 of the Appendix we provide a detailed description of the data. We categorize the 24 studies in four different types, depending on their design and reporting:

- Unilateral design (studies 1-12): patients received one intervention, in one of their arms. Studies report number of patients and number of events per intervention group.
• Bilateral design (studies 13-15): each patient received both interventions, in different arms. In the CTS example these studies reported number of patients and number of events per intervention group. They did not report cross-classified information (e.g. number of patients with events in both of their arms, with events in none of their arms, with events in one of their arms).

• Mixed design (studies 16-22): some patients received one intervention in one of their arms, and some patients received both in different arms. In the CTS example, these studies reported the corresponding number of patients in each group (or we calculated them from the reported data). No cross-classified information was reported.

• Unknown design (studies 23-24): these studies only report number of events and number of treated arms, per intervention group. They do not provide information on how many patients (if any) received both interventions in both of their arms.

2.2 Two surgical techniques for correcting myopia

Our second example comes from a Cochrane review for the comparison of Laser-assisted in-situ keratomileusis (LASIK) versus photorefractive keratectomy (PRK) for myopia\textsuperscript{23}. The outcome of interest is the proportion of eyes with uncorrected visual acuity (UCVA) of 20/20 or better, at six months after treatment. Ten studies reported this outcome. Two studies had a unilateral design, five studies had a bilateral design and did not report any cross-classified information and three studies had an unknown design, i.e. they did not report the exact number of unilateral and bilateral interventions. We present the data for this example in Section 1.2 of the Appendix. Table 1 summarizes the type of data that were available for each example and introduces notation.

<TABLE 1 HERE>

3 Methods

In this section we present our random-effects meta-analysis model that can be fit within a Bayesian framework. Methods for estimating and summarizing treatment effects from studies of unilateral and
bilateral design are established and we start by revising them. However, the number of events per patient and intervention (the cross-classified information) is required for a proper analysis of studies of bilateral design. Such information is often not available. We introduce a model that bypasses this problem, and then we extend it for the analysis of studies of mixed and unknown design. In Section 3.6 we discuss approaches for formulating prior distributions for all model parameters. We only consider the case of a binary outcome, i.e. each patient can only have up to one event per intervention.

### 3.1 Unilateral design

Let’s assume that a study $i$ compared interventions A and B, and that it applied only unilateral interventions. Let us assume $N_{i,A}$ ($N_{i,B}$) patients received intervention A (B), and that we observed $r_{i,A}$ ($r_{i,B}$) events. Given that each patient only received one intervention, the events are independent and we can assume two univariate binomial distributions. The estimated probabilities of event in each intervention group of each study can be used to estimate study-specific intervention effects, with their corresponding standard errors. These study-specific estimates can be synthesized at a second stage, under a usual meta-analytical model. This corresponds to a two-stage meta-analysis model. The drawback of two-stage models is that in the second stage the study-specific intervention effects are assumed to follow a normal distribution, and that the variances are assumed to be exactly known$^{24}$. This implies that a two-steps approach is generally suboptimal for the case of binomially distributed data, and even more so when the data are sparse.

Alternatively, we can do the meta-analysis in a single step using a hierarchical model. A one-step model can be written as follows:
\[ r_{i,A} \sim \text{Bin}(p_{i,A}, N_{i,A}), \quad r_{i,B} \sim \text{Bin}(p_{i,B}, N_{i,B}) \]
\[ \logit(p_{i,A}) = u_i + \frac{\theta_i}{2}, \quad \logit(p_{i,B}) = u_i - \frac{\theta_i}{2} \]
\[ \theta_i \sim N(\mu, \tau^2) \quad u_i, \mu, \tau \sim ... \text{(prior distributions)} \]

where \( \theta_i \) is the study-specific log-odds ratio, \( \mu \) is the summary log-odds ratio and \( \tau^2 \) the heterogeneity variance. \( u_i \) is a nuisance parameter that corresponds to the log-odds of the mean event rate in both intervention groups (we will refer to this quantity as “average risk” throughout this paper). For \( u_i, \mu \) and \( \tau \) prior distributions are required. An alternative, one-stage model can be formulated by assuming exchangeability on the average risk:

\[ r_{i,A} \sim \text{Bin}(p_{i,A}, N_{i,A}), \quad r_{i,B} \sim \text{Bin}(p_{i,B}, N_{i,B}) \]
\[ \logit(p_{i,A}) = u_i + \frac{\theta_i}{2}, \quad \logit(p_{i,B}) = u_i - \frac{\theta_i}{2} \]
\[ u_i \sim N(\mu, \sigma_u^2), \quad \theta_i \sim N(\mu, \tau^2) \]
\[ \mu, u, \sigma_u, \theta_i \sim ... \text{(prior distributions)} \]

The advantage of model (2) is that in the case of rare events it facilitates the estimation of the parameters by “borrowing strength” for the average risk across studies. The price to pay is that the model makes an additional assumption, i.e. the exchangeability of average risk across studies. Note that one can write alternative one-stage models, e.g. by assigning a prior to \( p_{i,A} \) or \( p_{i,B} \) instead of \( u_i \). We discuss this assumption more in the Discussion.

### 3.2 Bilateral design: studies providing cross-classified information

Let us assume that study \( i \) compared interventions A and B and that each patient received both interventions. Let us also assume that the study reports the full cross-classified information, as shown in Table 2. For simplicity, in what follows we drop the study index.
From this table we can estimate the probability of an event in each treatment. E.g. by maximizing the likelihood we estimate \( \hat{p}_A = r_A/N \), \( \hat{p}_B = r_B/N \), and also \( \hat{p}_k = n_k/N \), where \( k = 1,2,3,4 \). If we choose odds-ratio as an effect size, we have \( \log \text{OR} = \text{logit}(\hat{p}_A) - \text{logit}(\hat{p}_B) \). Observations \( (r_A \text{ and } r_B) \) belong to the same patients, which means that the estimates for the two probabilities \( \hat{p}_A \) and \( \hat{p}_B \) are correlated within each study. The estimated \( \text{var}(\log \text{OR}) \) should thus include a covariance term:

\[
\text{var}(\log \text{OR}) = \text{var}(\text{logit}(\hat{p}_A)) + \text{var}(\text{logit}(\hat{p}_B)) - 2 \text{Cov} (\text{logit}(\hat{p}_A), \text{logit}(\hat{p}_B))
\]

This covariance term can be written as a function of Pearson’s \( \phi \) coefficient\(^{25} \), which is a measure of the correlation between two binary variables of a 2 × 2 table\(^{26} \):

\[
\hat{\phi} = \frac{\hat{p}_A \hat{p}_B - \hat{p}_3 \hat{p}_4}{\sqrt{\hat{p}_A (1 - \hat{p}_A) \hat{p}_B (1 - \hat{p}_B)}}
\]

If we ignore the cross-classifications \( (n_1, n_2, n_3 \text{ and } n_4) \) and use only the margins of Table 2, we treat the two intervention groups as if they were independent. In that case the variance of the effects will not account for any covariance. This, in turn will lead to the precision of the estimated relative effects being either inflated or deflated, depending on the underlying correlation between \( r_A \) and \( r_B \). In most clinical examples we would expect a positive correlation, so ignoring this correlation is generally expected to lead to less precise results.

In order to correctly account for this covariance we need to use the full information and to employ a multinomial likelihood:

\[
(n_1, n_2, n_3, n_4) \sim \text{Multinomial}(p_1, p_2, p_3, p_4; N)
\]

We can then estimate \( \hat{p}_1, \hat{p}_2, \hat{p}_3 \) and \( \hat{p}_4 \), and we can identify \( \hat{p}_A = \hat{p}_1 + \hat{p}_2 \) and \( \hat{p}_B = \hat{p}_3 + \hat{p}_4 \), and calculate their covariance. At the meta-analysis level, we can use model (2) with a multinomial instead of a binomial likelihood.

\[\text{TABLE 2 HERE}\]
For this analysis one would need to have the cross-classified information of Table 2. Alternatively, if the study has performed an appropriate analysis, one could use the reported results (e.g. standard error or p-value) to reverse-engineer Equation (3) so as to calculate the entries of Table 2 and then use the correct multinomial distribution of Equation (5). To our experience, published studies rarely report the information needed for such back-calculation. To make matters worse, some of the studies may include a mixture of unilateral and bilateral interventions, often without reporting the corresponding numbers. In the following sections we show how to correctly analyze the data in such circumstances. We start by the simplest cases of data availability and move on to more complex situations.

3.3 Bilateral design: studies that do not provide cross-classified information

Let us assume a bilateral study where only the margins of Table 2 are reported (i.e. \(r_A, r_B\) and \(N\)). The exact likelihood of the marginal data (i.e. \(r_A, r_B\) and \(N\)) is given by Gonin and Aitken in the form of a bivariate binomial distribution\(^{27-29}\). More specifically, the probability of having \(r_A\) events in intervention group A and \(r_B\) events in B, in a total of \(N\) patients that received both interventions is given by:

\[
P(r_A, r_B) = \sum_{n_1=\max(0, r_A + r_B - N)}^{\min(r_A, r_B)} \frac{n_1! (r_A - n_1)! (r_B - n_1)! (N - r_A - r_B + n_1)!}{n_1! (r_A - n_1)! (r_B - n_1)! (N - r_A - r_B + n_1)!} \times p_1^{n_1} p_2^{r_A-n_1} p_3^{r_B-n_1} p_4^{N-r_A-r_B+n_1} \tag{6}
\]

where \(p_1, p_2, p_3\) and \(p_4 = 1 - p_1 - p_2 - p_3\) are the probabilities corresponding to each cell of Table 1.

Equation (6) runs through the total number of different set-ups of the contingency table that give rise to the same margins \(r_A\) and \(r_B\). One could in principle assume an uninformative prior for \(p_1, p_2, p_3\) and \(p_4\) (e.g. through a uniform Dirichlet distribution) and then use the likelihood of Equation (6) to obtain the posterior distribution for these four probabilities. As discussed in the previous section, the estimation of the relative effects between the two interventions would use \(\hat{p}_A = \hat{p}_1 + \hat{p}_2\) and \(\hat{p}_B = \hat{p}_1 + \hat{p}_3\). This would induce a correlation between the estimates of \(p_A\) and \(p_B\).
However, this approach would result to an average of all possible values for this correlation. It might be the case, however, that having an event in A is highly correlated with having an event in B. This could happen when for instance events are induced by the presence of a certain characteristic in the patient, irrespective of intervention received. In the CTS example, it might be the case that patients with manual labour professions are more prone to having a minor complication in both arms. In such cases, there would be a strong underlying correlation, which would not show in this approach. A hypothetical example to clarify this point is presented in Figure 1, where a study of 100 bilateral interventions reports 60 events in intervention A and 70 in B, but no information on the cross-classification. The likelihood of Equation (6) sums across all possible contingency tables: from those that correspond to a high correlation, like the one with \( n_1 = 60 \) in Figure 1 (\( \varphi = 0.80 \)), up to those that correspond to a strong negative correlation like the bottom configuration in Figure 1 (where \( \varphi = -0.53 \)). I.e. Equation (6) incorporates a correlation of \( p_A \) and \( p_B \) based on the average of all possible values of \( \varphi \). Essentially, the estimates for \( p_A \) and \( p_B \) are correlated in a study of a bilateral design, but one cannot estimate this correlation using only the margins of the 2x2 table.

In order to use an informed account of the correlations while employing the correct likelihood of the data described in Equation (6) we instead utilize external information for \( \varphi \). In Section 3.6 we discuss how to formulate prior distributions for this quantity based on external data or expert opinion. Given \( p_A \), \( p_B \) and \( \varphi \) we calculate \( p_1 = p_A p_B + \varphi \sqrt{p_A(1-p_A)p_B(1-p_B)} \), \( p_2 = p_A - p_1 \), \( p_3 = p_B - p_1 \) and \( p_4 = 1 - p_1 - p_2 - p_3 \). This method allows for an informed reconstruction of the full contingency table. This is performed in a stochastic way: more plausible configurations (according to the prior) are given a higher probability. Note that given a set of values for \( p_A \) and \( p_B \), \( \varphi \) is bounded between a minimum and a maximum value. Thus, after drawing values for the model’s parameters from the corresponding prior distributions, values the prior distribution for \( \varphi \) need to be appropriately truncated. Finally, note that
when one (or more) of the margins is 0, then the contingency table can be reconstructed with certainty.

For the reader’s convenience we provide the full likelihood of the model in the Appendix, Section 2.

An alternative method for reconstructing the 2x2 table uses the odds-ratios of the cross-classified
information instead of $\phi$. The advantage of this method is that it does not require any truncation. The
disadvantage is that it is less easy to formulate clinically meaningful priors for the parameters. Details on
this alternative approach are presented in Section 3 of the Appendix.

<FIGURE 1 HERE>

3.4 Mixed design

We now move on to the situation where a study includes a mixture of bilateral and unilateral
interventions (mixed-design studies). For including this type of studies in the analysis we will make one
extra assumption. We will assume that in each such study the probability of having an event when
receiving a specific intervention is the same in patients receiving unilateral interventions (patients
receiving only that intervention) and bilateral interventions (patients receiving both interventions):

\[ p_{A^{\text{uni}}}^{\text{uni}} = p_{B^{\text{bilat}}}^{\text{bilat}}(A^{+}|B^{+}) + p_{B^{\text{bilat}}}^{\text{bilat}}(A^{+}|B^{-}) p_{B^{\text{bilat}}}^{\text{bilat}}(B^{-}) \equiv p_{A^{+}} \]

This assumption might not hold if there is an interaction between the two interventions, for patients that received both. In such case additional assumptions would be needed. Note that for the case of zero correlation we would have

\[ p_{B^{\text{bilat}}}^{\text{bilat}}(A^{+}|B^{+}) = p_{B^{\text{bilat}}}^{\text{bilat}}(A^{+}|B^{-}) \]

In what follows we suppress the study index. Let us start by assuming that such a study provides
all relative information, i.e. the number of patients who only received intervention A, the number of
patients who only received B, as well as the number of patients that received both A and B; also, the
number of events for each of these three patients groups, and the full cross-classifications for the bilateral
interventions. We can estimate relative effects from such a study by employing the methods described in
the previous sections, namely two independent binomial likelihoods for the unilateral interventions in A
and B, and a multinomial likelihood as in Equation (5) for the bilateral ones. If no information on the
cross-classifications is available, likelihood (6) needs to be used instead, together with some external information on the correlation coefficient. According to our assumptions, these three groups inform only two probability parameters, \( p_A \) and \( p_B \).

Let us now focus on an even more complicated scenario, which also corresponds to the seven mixed-design studies in the CTS example. More precisely, let us assume that such a study provides information on the events on each intervention \( (r_A, r_B) \), the total number of patients receiving each intervention \( (N_A, N_B) \), and the total number of patients in the study \( (N) \). The number of events and the number of patients with a unilateral intervention \( (A \) and \( B) \), and the number events and patients with a bilateral intervention are not reported.

We first need to calculate the number of patients that received unilateral and bilateral interventions, using the reported data. Let us keep in mind that each patient can have a maximum of one event per intervention. Let us denote \( N_A^u \) \( (N_B^u) \) the number of patients that received only intervention A \( (B) \). Let \( N^b \) denote the bilateral interventions, i.e. the number of patients that received both A and B. Obviously \( N_A^u + N_B^u + N^b = N \). It also holds that \( N_A^u + N_B^u + N^b = N_A \) since the \( N_A^u + N_B^u + N^b = N_B \). By solving this set of equations we find the number of patients in each group, i.e. \( N_A^u = N - N_B \), \( N_B^u = N - N_A \) and \( N^b = N_A + N_B - N \).

We do not have information on how the \( r_A \) events in intervention A are divided among the \( N_A^u \) patients who received only A and the \( N^b \) patients who received both A and B. Let us assume that \( x \) out of these \( r_A \) events were in the \( N_A^u \) patients; the rest \( r_A - x \) were in the \( N^b \) bilateral cases. Obviously it should hold that \( x < N_A^u \). Also \( r_A - x < N^b \), i.e. the number of events in the bilateral interventions group must be smaller than the total number of patients in that group. Of course \( x \geq 0 \) and also \( x \leq r_A \). All these constraints are summarized in the following double inequality: \( \max(0, r_A - N^b) \leq x \leq \min(N_A^u, r_A) \).

Likewise, if we denote by \( y \) the number of patients who only received intervention B and had an event, it holds that \( \max(0, r_B - N^b) \leq y \leq \min(N_B^u, r_B) \).
We also do not know how the \( r_A - x \) and \( r_B - y \) events observed in patients who received bilateral interventions are cross-classified in a \( 2 \times 2 \) contingency table, like the one in Table 1. In Figure 2 we give a graphical analysis of the possible ways that the observed events may be distributed in each category of patients.

Quantities \( x, y \) and \( n_1 \) of Figure 2 are not reported. In order to write the exact likelihood of the reported data \( (r_A, r_B, N_A, N_B, N) \), we will combine the likelihoods of all possible configurations that lead to these observations, as presented in Figure 2. Since \( x \) and \( y \) correspond to events in different groups of patients, they follow independent binomial distributions i.e. \( x \sim Bin(p_A, N_A) \) and \( y \sim Bin(p_B, N_B) \). For the events in the group of patients with bilateral interventions we follow the analysis of Section 3.3 and we can write the likelihood after carefully adjusting Equation (6). The full, exact likelihood of the observed data \( (r_A, r_B) \) can be written as:

\[
P(r_A, r_B) = \sum_{x=\max(0,r_A-N_B)}^{\min(N_A, r_A)} \sum_{y=\max(0,r_B-N_A)}^{\min(N_B, r_B)} \frac{N_A^x N_B^y (1-p_A)^{N_A-x} (1-p_B)^{N_B-y}}{n_1! (r_A - x - n_1)! (r_B - y - n_1)! (N_A - r_A - x + y + n_1)!}
\times p_1^{n_1} p_2^{r_A-x-n_1} p_3^{r_B-y-n_1} p_4^{N_A-x-r_B+y+n_1}.
\]

This likelihood is a generalization of the cases described in Section 3.1 (unilateral interventions only) and 3.3 (bilateral interventions only). It is easy to see that this likelihood decomposes into a simple product of two independent binomials if we set \( N_B = 0 \) (in which case \( x = r_A, y = r_B \) and \( n_1 = 0 \)): this corresponds to the case where there are no bilateral interventions. It is also straightforward to see that if we set \( N_B = N \) then \( N_A^x = N_B^y = x = y = 0 \) and the likelihood of Equation (7) reduces to the one of Equation (6), as expected. A more detailed description of this model can be found in Section 4 of the Appendix.
The joint posterior distribution for \( p_A \) and \( p_B \) includes their correlation. The extent of this correlation depends on (i) the number of patients who received bilateral interventions in the studies, where zero bilateral interventions will mean zero correlation, and (ii) the prior distribution used for \( \varphi \).

<FIGURE 2 HERE>

3.5 Unknown design

A study of this type reports \( r_A, r_B, N_A \) and \( N_B \), but does not provide any information on the total number of patients. This implies that there is no (exact) information on how many of the included patients received unilateral and how many bilateral interventions; but meta-analysts might have reasons to suspect that at least some of the patients in these studies received both interventions.

The lack of more information from these studies means that researchers cannot use the methods described in the previous sections to take into account correlations in the estimation of relative intervention effects, unless some strong assumption is employed, i.e. on the percent of bilateral interventions in the study. The simplest approach for the analysis is to assume two independent binomials. As we discussed, this may lead to over- or under-precision in the estimated relative intervention effects if the correlation is non-zero.

The two extreme cases for such a study are (i) there are no bilateral interventions, and (ii) there are a maximum number of bilateral interventions (= \( \max(N_A, N_B) \)). Given that information on this number is unavailable, when analyzing such a study researchers may want to use the most conservative estimate. This means that if \( p_A \) and \( p_B \) are expected to be positively correlated in bilateral interventions, the analyst should analyze a study of unknown design as if all patients received only one intervention, i.e. as if the study was unilateral. If instead researchers are instead interested in obtaining the most liberal estimates, they should assume that a large number of interventions might be bilateral (maximum value is \( \max(N_A, N_B) \)). These considerations need to be reversed for the (perhaps less relevant) case of having a
negative correlations. In any case, a sensitivity analysis is warranted to explore the effect of the
(unknown) study design in the overall, pooled effect.

In Section 5 of the Appendix we provide a summary table of how to choose the likelihood for each
study, depending on its design and on the information it provides.

3.6 Formulating informative prior distributions

In what follows we discuss how meta-analysts can augment the information included in the studies
using external information in the form of informative prior distributions for the model parameters. We
focus on one-step meta-analysis models. We discuss the following parameters: the heterogeneity
variance, the mean and variance of the average risk, and the correlation ($\rho$ coefficient).

Heterogeneity

In a recent meta-epidemiological study by Turner et al.\textsuperscript{32}, the authors re-analysed data from
around 15,000 binary outcome meta-analyses from the Cochrane Database of Systematic Reviews.
Results were used to create informative predictive distributions for $\tau^2$\textsuperscript{33}. These prior distributions cover
80 different settings, categorized by the outcome being assessed, the types of interventions being
compared, etc.

Average risk

The optimal source of information for this parameter would be to use eligible RCTs that were
excluded from the original meta-analysis due e.g. to inadequate reporting of effect sizes. This would
allow researchers to obtain a realistic estimate of the mean and variance of the (logit-transformed)
average risk in order to formulate priors that can be included in the model. Alternatively, one could use
observational studies (such as registries or large cohort studies). Note however that in order to formulate a
prior distribution for the variance (parameter $\sigma^2$ in Equation (2)), one would need to include multiple
studies. If no external source of information is readily available, then a prior distribution can be elicited
from the experts, following a method described in\textsuperscript{25,34}. In Section 6 of the Appendix we discuss the
details of this method, i.e. how researchers can elicit information and how they can synthesize information obtained from multiple experts (after possibly weighting each expert’s opinion, e.g. for the years of experience on the field, or the number of studies he/she has been involved in).

**Correlation**

If some of the studies report full, cross-classified information (as described in Section 3.2), then the corresponding estimates of the correlation coefficients can be used to formulate an informative prior distribution for the analysis of the rest of the studies. Alternatively, one can formulate priors by eliciting information from expert clinicians. In a previous work we describe how to elicit information on correlation indirectly, using a conditional probability. Here, we propose a more straightforward method. As shown empirically by Clemen et al., directly asking for the correlations is a reasonable approach, often more accurate than indirect methods. More detail on how to synthesize information from multiple experts in order to construct a prior distribution for $\varphi$ can be found in Section 7 of the Appendix. The method uses the Fisher transformation to synthesize information, $\zeta = \frac{1}{2} \log \left( \frac{1 + \varphi}{1 - \varphi} \right)$. If there is no usable external information for $\varphi$, researchers can use a minimally informative prior $\varphi \sim U(0,1)$ if the correlation is generally expected to be positive (and conversely $\varphi \sim U(-1,0)$ if it is expected to be negative).

It is unclear whether the choice of the prior distribution for the different parameters of our model is equally important, and whether researchers should focus on obtaining informative priors for some of the parameters rather than others. It is also unclear what would be the impact of using misspecified prior distributions to our model estimates. Lambert et al. showed that in general the choice of scale parameters (such as standard deviations) is more important than the choice of location parameters in Bayesian analyses with sparse data. As we will see in the next section, this agrees with the findings of our simulations. There we discuss that heterogeneity is the most influential parameter in our model. Using an informative prior for heterogeneity can greatly increase precision and enhance the power of the model. As
we will also see, using minimally informative (or even misspecified) priors for the rest of the parameters may have a smaller impact on the model’s performance.

4 Simulations

4.1 Data generation procedure and description of the scenarios

We performed a small-scale simulation study to assess the performance of the model we proposed in this paper. We limited the simulations to the case of studies with bilateral design where only the marginal numbers of events are reported, and among which there is variability (heterogeneity). We explored a range of different scenarios. For each scenario we simulated 100 datasets, with each dataset corresponding to an independent meta-analysis.

Scenarios 1-20 aimed to compare our model with the bivariate beta-binomial model. This model has been previously recommended for meta-analyzing rare events. For these scenarios we simulated 20 studies for each meta-analysis. Two additional scenarios (21 and 22) aimed to explore the importance of specifying informative priors for the parameters of our bivariate binomial model. For these scenarios we simulated 10 studies for each meta-analysis, aiming to increase the impact of the prior distributions.

To generate data, for each scenario we first defined the true relative treatment effect on the log-odds ratio scale ($\mu$), the log-odds of the average risk of event and its standard deviation ($\mu_u$, $\sigma_u$) and also the correlation between the events ($\varphi$). For each independent meta-analysis we sampled the variance of random effects from a log-normal distribution: $\tau^2 \sim \text{LN}(-2, 1.5)$. Suppressing the study index for clarity, for each study we sampled the study-specific treatment effects, $\theta \sim N(\mu, \tau^2)$ and the log-odds of the average risk, $u \sim N(\mu_u, \sigma_u^2)$. We simulated the number of patients ($N$) in each study by drawing from a uniform distribution, $U(50, 100)$.

Using $\theta$, $u$ and $\varphi$ we calculated the probabilities $p_1$, $p_2$, $p_3$ and $p_4$ (corresponding to $n_1$, $n_2$, $n_3$ and $n_4$ of Table 2). Using these four probabilities and the number of patients per study we generated the number of events of the 2x2 cross-classified table by drawing from a multinomial distribution. We then
calculated $r_a, r_b$; together with $N$, these are the only data that were supposed to be available from the study.

We explored scenarios using the following values for the simulation parameters:

- $\phi = 0.30$ (\( \zeta = 0.31 \), low correlation) and $\phi = 0.60$ (\( \zeta = 0.69 \), high correlation).
- $\mu_u = -1.5$, $\sigma_u = 0.5$: in these scenarios there were almost no single- and double-zero studies (<1%). $\mu_u = -3.0$, $\sigma_u = 1.0$: these scenarios have some single- and double-zero studies (0 – 30% in total). $\mu_u = -4.0$, $\sigma_u = 1.0$ and $\mu_u = -3.0$, $\sigma_u = 3.0$: these scenarios have many single- and double-zero studies (more than 30% in total).

We generated all data in R\textsuperscript{37}, code is provided in the Appendix, Section 8.

### 4.2 Analyses of the simulated datasets

The simulated datasets were analyzed using the bivariate binomial model discussed in Section 3.2. The model is also presented in more detail in Section 2 of the Appendix. We performed our analyses in OpenBUGS\textsuperscript{39}, the code is available in Section 9 of the Appendix. For each analysis we simulated a single chain of 20,000 samples, and we discarded the first 5,000 samples. This was deemed to be sufficient based on some initial runs and also after visually inspecting the posterior distributions\textsuperscript{40}.

For the analysis of scenarios 1-20 we chose between using weakly informative prior distributions centered near the true value of the parameters, vague priors and misspecified priors for the correlation. These choices are described in detail in Table 4 of the Appendix (Section 10.1). In addition, these scenarios were analyzed using the bivariate (Sarmanov) beta-binomial model, which accounts for correlations in the 2x2 tables\textsuperscript{21}. For this we used a routine in R\textsuperscript{37} provided by Chen et al.\textsuperscript{21}. We also checked results by comparing with results obtained from the mmeta package in R, which also implements the beta-binomial model\textsuperscript{41}. We compared our model with beta-binomial in terms of (a) mean bias; (b)
mean squared error (MSE) (the mean of the squared bias); (c) empirical coverage (the percent of meta-analyses that included the true effect in their 95% credible interval (CrI)); (d) empirical power (percent of meta-analyses that rejected the null hypothesis of no treatment effect, when the true treatment effect was non-zero); and (e) the gain in precision (bivariate binomial model over beta-binomial model). This was calculated as the percentage reduction of the mean width of the 95% Credible Interval obtained from our model as compared to the 95% Confidence Interval from the beta-binomial model. The datasets generated under scenarios 21 and 22 were analyzed multiple times, using only the model presented in this paper. Each dataset was analyzed using first vague priors for all parameters, and then using vague priors for all model parameters except for one at a time. Details can be found in Table 5, Section 10.2 of the Appendix.

4.3 Results from the simulations

The results from the analyses of scenarios 1-20 are presented in Table 3. Regarding the interpretation of results, and taking scenario 1 as an example, we can see that our model had a slightly larger mean bias for the logOR than the bivariate beta-binomial (0.012 vs 0.001), and also a slightly larger MSE (0.024 vs. 0.018). However, the CrIs obtained from our model were narrower than the CIs obtained from the bivariate beta-binomial model. Thus, coverage for the former model was 95% while for the latter it was 99%. By comparing the precision of the two models (quantified by the width of the corresponding credible/confidence intervals) we calculated a 29% increased precision of our model as compared to the bivariate-binomial.

Simulations showed that our model performs markedly better than the beta-binomial model in most scenarios we explored. In almost all cases our approach led to smaller mean bias and similar MSE. Coverage probability was closer to the nominal 95% in our approach for most scenarios. Our model led to an increased precision of the estimates. This increase in precision, quantified as the percentage reduction of the width of the 95% CrI, ranged from 29 to 44% in all scenarios. Our approach performed much better in scenarios with many single- and double-zero studies, when we assumed non-zero log-odds ratios. In
these settings the beta-binomial model had very low power to detect a relative treatment effect. Our simulations showed that in the presence of bilateral interventions, even when the prior distributions are only mildly informative or misspecified, the bivariate binomial model performs better than the bivariate beta-binomial (Scenarios 13-16).

Another interesting finding in our simulations was that for the case of non-zero relative treatment effects and large heterogeneity in the average risk (Scenarios 17-20), the beta-binomial model was very inefficient. In these scenarios this model showed excessive bias, and in all cases the bias was towards zero treatment effects. Moreover, there was large MSE, insufficient coverage and minimal power. This was the case even under zero correlation scenario, i.e. for the usual unilateral design (Scenario 20). This finding comes in contrast with the recommendations by Kuss\(^8\), i.e. to use the beta-binomial model for all meta-analyses of rare events. But, in his simulations\(^8\), Kuss did not explore the scenario of large heterogeneity in the average risk of an event. In a somewhat different context (meta-analysis of proportions, not relative effects) Ma et al. showed that the beta-binomial model performs worse when the event rates are relatively large (e.g. \(>5\%\))\(^{42}\).

In Table 4 we present the results from the analysis of data simulated under scenarios 21 and 22, using different priors for the model’s parameters. Results suggested that the most influential prior is the one for heterogeneity. Using an informative prior distribution for \(\tau^2\) instead of a vague one greatly enhanced the performance of the model in terms of bias, precision, coverage and power. Using informative priors for the other model parameters had a smaller impact in all scenarios explored.

<TABLE 3 HERE>
<TABLE 4 HERE>

5 Applications

In this section we show results from applying our methods to the two examples presented in Section 2. We used the following models for the analysis:
i. Univariate fixed-effects meta-analysis (UNI-FE), i.e. we used model in Equation (2), but omitting the random effects distribution. This analysis ignores the correlations due to bilateral interventions. We fitted this model in a Bayesian as well as a frequentist setting, using maximum likelihood estimation (MLE).

ii. Univariate random-effects meta-analysis (UNI-RE), using the model in Equation (2). This analysis ignores the correlations due to bilateral interventions. We fitted this model in a Bayesian as well as a frequentist setting, using MLE.

iii. Bivariate binomial, fixed-effects meta-analysis (BB-FE): we used the model described in this paper accounting for the correlations due to bilateral interventions, as discussed in Sections 3.3 and 3.4. We omitted random effects.

iv. Bivariate binomial, random effects meta-analysis (BB-RE): the same as in model iii, but also including random effects.

v. The bivariate beta-binomial model. Note that this is by definition a random effects model.

We provide the OpenBUGS code in Section 11 of the Appendix. In Section 12 of the Appendix we provide codes for fitting the model in R using the R2WinBUGS package, and for fitting the model of Equation (2) in a frequentist setting. We performed 20,000 iterations, and we discarded the first 5,000 samples. We fitted the bivariate beta-binomial model in R. We fitted all models using a conventional laptop computer. The run-time required for fitting our bivariate binomial model was around 23 minutes for the CTS example and less than a minute for the myopia example.

5.1 Surgical operations for CTS

Two studies in this dataset had unclear design. We analyzed these studies as if they were of unilateral design following the arguments presented in Section 3.5 as we expect a positive correlation between complications in two arms of the same patient.
For models i to iv we assume a vague prior $\mu_u \sim U(-6, 0)$ for the log-odds of the mean event rate, $\sigma_u \sim U(0, 3)$ for the corresponding variance, and $\mu \sim U(-2, 2)$ for the log-odds ratio. For the random-effects models (models ii and iv) we use informative priors from the empirical distributions for a safety outcome for non-pharmacological interventions $\log(r^2) \sim N(-1.43, 1.45^2)$ provided in 33. For the bivariate models (iii and iv), we perform two analyses: one assuming a low correlation between the outcomes, $\zeta \sim N(0.2, 0.1^2)$, and one assuming a moderate-high correlation, $\zeta \sim N(0.5, 0.1^2)$. Results are in Figure 3. The BB-RE model with moderate-high correlation was assumed to be our primary model, the rest of the models can be seen as sensitivity analyses. For the primary model, the posterior median estimate for $\mu_u$ was -3.41 (-4.35, 2.56) and for $\sigma_u$ was 1.93 (1.33, 2.80).

The endoscopic surgical operation was found to be more safe than the open in all analyses. When switching from the univariate to the bivariate model the increase in precision was rather small, especially when a low correlation was assumed. This should come as no surprise, as only 13% of the patients in this meta-analysis received bilateral interventions. One interesting observation was that when we increase the assumed correlation in the BB-RE scenario the estimate and CrI. for the odds ratio remained unchanged. This is because the increased precision in study-estimates is accompanied by an increase in the estimated heterogeneity. All implementations of the bivariate binomial model were more precise than the beta-binomial model. Comparing the Bayesian and frequentist implementations of the UNI-RE model, we see some differences, which reflect the impact of the prior distributions. E.g. the MLE estimate for $\tau$ in the UNI-RE (F) model was 0.83, while for UNI-RE (B) the median of the posterior median for $\tau$ was 0.75. This difference was due to the impact of the informative prior distribution, which has median 0.48, thus pulling the MLE estimate towards lower values. The estimate for the odds ratios was also slightly different (0.50 for the Bayesian implementation vs. 0.52 for the frequentist). This difference can be attributed to the prior used for logOR, which was centered at 0, thus pulling the MLE estimate to lower values.
In order to assess the impact of the prior distributions in our results, we did a sensitivity analysis for model IV (BB-RE) with low correlation. Assuming vague priors for $\mu$, $\mu_u$ and $\sigma_u$ had an immaterial impact on the results. However, a less informative prior for the heterogeneity ($\tau^2 \sim U(0, 2)$) considerably increased imprecision in the treatment effects. This highlights once more the importance of using informative prior distributions specifically for the random effects variance parameter.

Finally, as we argued in Section 3.3 the only information the data carries for the correlation coefficient $\zeta$ is the range of the allowed values within each study determined by the marginal total counts (as also depicted in Figure 1). As a result, the posterior estimates for $\zeta$ are entirely determined by their prior distributions.

5.2 Surgical techniques for correcting myopia

Two studies in this dataset had unclear design. We analyzed these studies as if they were of unilateral design assuming a positive correlation. We used the following prior distributions for the model parameters: $\mu_u \sim U(-5, 5)$; $\sigma_u \sim U(0, 5)$ and $\mu \sim U(-2, 2)$. For heterogeneity we use an informative prior distribution $\log(\tau^2) \sim N(-1.69, 1.68^2)$ based on the empirical evidence. We explore two options for correlation, $\zeta \sim N(0.2, 0.1^2)$ (low correlation), and $\zeta \sim N(0.5, 0.1^2)$ (moderate-high correlation). Results are presented in Figure 4. For our primary model, the posterior median estimate for $\mu_u$ was 1.35 (-0.16, 2.99) and for $\sigma_u$ was 2.20 (1.34, 3.90).

Accounting for correlations using our approach led to an increase in precision of the pooled estimates. For the case of random-effects model, some of the increase in precision was again counterbalanced by an increase in the estimates for $\tau^2$, as in the CTS example. There are some differences between Bayesian and frequentist implementations of the UNI-RE model, particularly for $\tau$. Heterogeneity was estimated to be 0 in the frequentist implementation, while in the Bayesian approach
the median posterior $\tau$ was 0.29. This difference is again due to the effect of the informative prior distribution (which had a median of 0.43).

The odds ratio estimated from the beta-binomial model was 1.00 (0.94, 1.06). This finding was markedly different to the results from all bivariate binomial models. It was also in disagreement with the frequentist analysis performed in the original publication; the authors performed a Mantel-Hanszel fixed-effects meta-analysis resulting into $\text{OR}=1.40(1.00, 2.00)$, which is in broad agreement with bivariate binomial models.

One explanation for this important discrepancy between beta-binomial and the other approaches lies in the distribution of the average risk in the included studies. The average event risk ranged from 8% to 97% in the included studies, despite the fact the relative treatment effect was not very heterogeneous (see Appendix, Section 1.2). In such situations the bivariate beta binomial model may be heavily biased towards zero, as shown in the simulations (Section 4, scenarios 17-20). Additionally, the bivariate beta-binomial model was shown to perform well when the event rate is small; here it is on average 70%.

6 Discussion

In this paper we have presented a Bayesian meta-analysis model for synthesizing binary data obtained from a collection studies of different designs: studies of the usual parallel design, studies of a bilateral (split-body) design, and studies including a mixture of unilateral and bilateral interventions. Our model uses a bivariate binomial distribution that accounts for the correlations induced due to bilateral operations, and has several distinct advantages. It uses the exact likelihood of the data; it does not employ the normal approximation; it respects the randomization of the studies; it includes in the analysis data coming from studies with zero events in one or even both interventions, without a need for imputation. Our model has similarities with the bivariate beta-binomial (Sarmanov) model for meta-analysis (e.g. as described by Chen et al.\cite{21}). The two models differ, however, in the model parametrisation and the
assumption of random effects. The bivariate beta-binomial model assumes random effects for the arm-specific event probabilities while the mean log-odds ratio is not a parameter of the model. In our approach, random effects are assumed for study-specific log-odds ratios, and the mean log-odds ratio is a parameter directly estimated in the model. This feature allows us to assume an informative prior distribution for the log-odds ratio, based on available empirical distributions.

The bivariate beta-binomial model has been previously suggested as the optimal approach for meta-analysing rare events\(^8\). Our simulations showed that our model may lead to significant increase in precision, coverage and power, and a decrease in bias, as compared to the bivariate beta-binomial model. Moreover, according to our simulations, the bivariate beta-binomial model was found to perform very badly when the baseline risk was very variable in the included studies, even when no correlation was assumed. This was also demonstrated in one of the two real data-sets we used to illustrate our methods. The increase of precision in our model is more pronounced in fixed-effects meta-analyses. In the random-effects regime, increasing the precision of the study estimates by accounting for correlations can be partly counterbalanced by an increase in the estimate for heterogeneity.

We set our model in a Bayesian framework. This allows the inclusion of external information for the model parameters. However, for the case of rare events inferences from Bayesian meta-analyses may heavily depend on the choice of prior distribution for the parameters – even when these are thought to be uninformative\(^36\). Due to this fact, the use of frequentist approaches instead of Bayesian approaches is sometimes advocated. We understand the reasoning behind this view; in this paper however, we argue in favour of a Bayesian approach. We think that the scarcity of information due to the rarity of events should not be seen as an argument against the use of Bayesian methods. On the contrary, we think that meta-analysts should opt for using Bayesian methods when they have the opportunity to include high quality, trustworthy external evidence in the analysis.

https://mc.manuscriptcentral.com/smmr
The major limitation of our model lies on its complexity. The software code we provide might be computationally heavy when studies have very large sample sizes. This may be especially true if there are large, mixed-design studies in the dataset. Moreover, embarking into this complicated modelling will only make a difference in the estimates if the corresponding correlation is large. If the correlation is expected to be small (e.g. $\phi < 0.2$), then researchers can safely treat observations from patients that received both interventions as if they corresponded to different patients.

In our model we assumed exchangeability on the average risk. This might be a very strong assumption to make, e.g. if there are large differences in the randomization ratio across studies. Alternatively, researchers can assume exchangeability in the probability of having an event in one of the interventions, $p_{iA}$ or $p_{iB}$ in Equation (2). Choosing between the two, however, might be important when the data are sparse. Thus, the choice should be ideally guided by the availability of trustworthy external information that can be used to formulate informative priors. Another limitation of our model is that it does not account for the ordering and timing of the treatments. E.g., there might be studies where the treatments were administered simultaneously and there might be studies where treatment A is given first and treatment B later, or vice versa, first B then A.

One additional limitation of our model is that it cannot correctly account for correlations induced by patients that received the same intervention in multiple sites of their bodies i.e. a classical cluster design (e.g. the same intervention in both hands, both eyes, multiple teeth, multiple people in a household, etc.). This situation requires an extension of the approach described in this paper. Other possible extensions of our bivariate binomial meta-analysis model relate to the case of two correlated outcomes (for studies of the usual, parallel design), and for the meta-analysis of twin studies. Also, our model can be extended for the meta-analysis of the accuracy of multiple diagnostic tests. Finally, it would be interesting to explore the use of the non-central hypergeometric distribution\textsuperscript{15,44} for bilateral interventions, and compare it with the bivariate binomial approach described in this paper.
Dimou et al. discuss an alternative method that can be used to fill the (unknown) cross-classified information in the contingency tables of studies that do not report it. Their method uses information from studies that report the full tables. When no studies report the full tables, the authors suggest the use of the iterative proportional fitting algorithm, which is however based on strong assumptions. Both these methods do not account for uncertainty in the (unobserved) missing values of the contingency tables. Moreover, the methods described in that paper are not suitable to use in the case of rare events.

A different approach to synthesizing data from studies with bilateral design is to only use information on the number of discordant pairs, i.e. the number of patients with an event in A but not in B, and the number of patients with an event in B but not in A. This approach, however, could not be used for the case where some studies in the meta-analysis are of a unilateral design, and/or some studies are of a mixed design. E.g. this approach could not have been used in the examples presented in this paper.

To summarize, we think that the model we presented constitutes the best available method for meta-analysing binary outcomes in the presence of bilateral (split-body) interventions, and that its implementation is in practice straightforward.

Acknowledgements

We would like to thank Nikolaos Pandis and Haris Vasilieiadis for their help in choosing and interpreting the clinical examples we used. OE was supported by the Swiss National Science Foundation (grant title: «Enhancing methods for evaluating the comparative safety of medical interventions»). GR was funded by the German Research Foundation (DFG) (grant RU 1747/1-2). GS is a Marie Skłodowska-Curie Fellow (MSCA-IF-703254).

References

https://mc.manuscriptcentral.com/ssmr


**Table 1:** Summary of the type of data available for the CTS and myopia examples. $r_A$ ($r_B$) denotes the number of events and $N_A$ ($N_B$) the number of patients that received intervention $A$ ($B$). With $N$ we denote the total number of patients included in each study.

<table>
<thead>
<tr>
<th>Design</th>
<th>Data provided in the studies</th>
<th># of studies available</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unilateral design:</strong> Only one body part per patient was treated</td>
<td>$r_A$, $N_A$; $r_B$, $N_B$; $N = N_A + N_B$</td>
<td>CTS: 12; Myopia: 2</td>
</tr>
<tr>
<td><strong>Bilateral design:</strong> Each patient received both interventions in different body parts</td>
<td>$r_A$, $r_B$; $N = N_A = N_B$</td>
<td>CTS: 3; Myopia: 5</td>
</tr>
<tr>
<td><strong>Mixed design:</strong> Mixture of unilateral and bilateral interventions; some patients received one intervention in one body part, some received both interventions in different body parts</td>
<td>$r_A$, $N_A$; $r_B$, $N_B$; $\max(N_A, N_B) \leq N$; $N &lt; N_A + N_B$</td>
<td>CTS: 7; Myopia: 0</td>
</tr>
<tr>
<td><strong>Unknown design:</strong> No information on whether any patients received both interventions</td>
<td>$r_A$, $N_A$; $r_B$, $N_B$</td>
<td>CTS: 2; Myopia: 3</td>
</tr>
</tbody>
</table>

**Table 2:** 'Contingency' table for a study of bilateral design with $N$ patients, providing the full cross-classified information. $A^+$ ($A^-$) denotes events (non-events) in intervention $A$; likewise for treatment $B$. $n_1$ denotes the number of patients with events for both interventions. $n_2$ is the number of patients with an event for $A$ and a non-event for $B$, the opposite is $n_3$. $n_4$ denotes the number of patients with no events in any intervention.

<table>
<thead>
<tr>
<th></th>
<th>$B^+$</th>
<th>$B^-$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A^+$</td>
<td>$n_1$</td>
<td>$n_2$</td>
<td>$r_A$</td>
</tr>
<tr>
<td>$A^-$</td>
<td>$n_3$</td>
<td>$n_4$</td>
<td>$N - r_A$</td>
</tr>
<tr>
<td>Total</td>
<td>$r_B$</td>
<td>$N - r_B$</td>
<td>$N$</td>
</tr>
</tbody>
</table>
Table 3: Results from 20 scenarios, comparing the proposed bivariate binomial model to the bivariate beta-binomial model. All scenarios are analysed using informative priors on $\varphi$ for the bivariate binomial model, except otherwise noted. SZ (DZ): single (double)-zero studies. ‘Some SZ/DZ’ denotes scenarios with 0-30% SZ/DZ studies. ‘Many SZ/DZ’ denotes scenarios with >30%. $\mu$ denotes the true log-odds ratio, $\varphi$ the true correlation in each scenario and $\sigma$ the standard deviation of the log-odds of the average risk. Large heterogeneity in average risk corresponds to $\sigma > 2$. Gain in precision corresponds to the percentage reduction in the width of the 95% CrI, of our model vs. the beta-binomial. MSE: mean squared error.

<table>
<thead>
<tr>
<th>#</th>
<th>Scenario description</th>
<th>Bivariate binomial</th>
<th>bivariate beta-binomial</th>
<th>Precision gained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean bias</td>
<td>MSE</td>
<td>Coverage</td>
</tr>
<tr>
<td>1</td>
<td>$\mu = 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No SZ/DZ</td>
<td>0.012</td>
<td>0.024</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Some SZ/DZ</td>
<td>0.006</td>
<td>0.044</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Many SZ/DZ</td>
<td>0.031</td>
<td>0.059</td>
</tr>
<tr>
<td>4</td>
<td>$\mu = 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No SZ/DZ</td>
<td>0.000</td>
<td>0.022</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Some SZ/DZ</td>
<td>0.006</td>
<td>0.032</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Many SZ/DZ</td>
<td>0.002</td>
<td>0.031</td>
</tr>
<tr>
<td>7</td>
<td>Some SZ/DZ and:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Informative priors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Vague priors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Misspecified priors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4: Results from analysing scenarios 21 and 22, for different choices of priors for the model parameters.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Informative priors used for</th>
<th>Mean bias</th>
<th>MSE</th>
<th>Coverage</th>
<th>Power</th>
<th>Increase in precision compared to non-informative priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 μ=0.5, φ=0.3</td>
<td>Many SZ/DZ</td>
<td>0.000</td>
<td>0.063</td>
<td>97%</td>
<td>44%</td>
<td>0.258</td>
</tr>
<tr>
<td>18 φ=0.3</td>
<td>Some SZ/DZ</td>
<td>0.024</td>
<td>0.028</td>
<td>98%</td>
<td>77%</td>
<td>0.221</td>
</tr>
<tr>
<td>19 φ=0.6</td>
<td>Many SZ/DZ</td>
<td>0.040</td>
<td>0.071</td>
<td>96%</td>
<td>60%</td>
<td>0.242</td>
</tr>
<tr>
<td>20 φ=0.0</td>
<td>Many SZ/DZ</td>
<td>0.011</td>
<td>0.046</td>
<td>100%</td>
<td>45%</td>
<td>0.260</td>
</tr>
</tbody>
</table>

**μ=0.5. Assumed large σ_u (variability in the average risk)**

Table 4: Results from analysing scenarios 21 and 22, for different choices of priors for the model parameters. μ denotes the log-odds ratio and φ the correlation. μ_u and σ_u denote the mean and standard deviation of the average risk.
Three possible configurations of the 2x2 table of a study of 100 bilateral interventions reporting only the margins of the contingency table. The first configuration corresponds to a large positive correlation ($\varphi=0.80$), the second to no correlation ($\varphi=0$), the third to a large negative correlation ($\varphi=-0.53$).

$127x71mm (300 x 300 DPI)$
Graphical analysis of the different ways in which the observed events can be distributed among different categories of patients in a mixed-design study. x, y and n1 are not reported.

239x165mm (150 x 150 DPI)
Meta-analysis for minor events, endoscopic vs. open surgical operation for CTS, using a range of alternative models: UNI-FE (B/F): univariate fixed-effect (Bayesian/Frequentist), UNI-RE (B/F): univariate random-effects (Bayesian/Frequentist), BB-FE: bivariate binomial fixed-effect (Bayesian), BB-RE: bivariate binomial random-effects (Bayesian). For all Bayesian models we present the median of the posterior distribution.

<table>
<thead>
<tr>
<th>Primary Model</th>
<th>OR (95% CrI)</th>
<th>τ (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB-RE, high φ</td>
<td>0.50 (0.30, 0.91)</td>
<td>0.77 (0.35, 1.41)</td>
</tr>
</tbody>
</table>

### Sensitivity analyses

#### Bayesian

<table>
<thead>
<tr>
<th>Model</th>
<th>OR (95% CrI)</th>
<th>τ (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNI-FE (B)</td>
<td>0.41 (0.28, 0.58)</td>
<td></td>
</tr>
<tr>
<td>UNI-RE (B)</td>
<td>0.50 (0.29, 0.95)</td>
<td>0.75 (0.26, 1.43)</td>
</tr>
<tr>
<td>BB-FE, low φ</td>
<td>0.40 (0.28, 0.57)</td>
<td></td>
</tr>
<tr>
<td>BB-RE, low φ</td>
<td>0.50 (0.30, 0.91)</td>
<td>0.74 (0.28, 1.41)</td>
</tr>
<tr>
<td>BB-FE, high φ</td>
<td>0.39 (0.28, 0.54)</td>
<td></td>
</tr>
<tr>
<td>BB-RE, low φ,</td>
<td>0.53 (0.29, 1.03)</td>
<td>1.00 (0.50, 1.38)</td>
</tr>
<tr>
<td>vague priors on τ</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Frequentist

<table>
<thead>
<tr>
<th>Model</th>
<th>OR (95% CrI)</th>
<th>τ (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNI FE (F)</td>
<td>0.41 (0.28, 0.58)</td>
<td></td>
</tr>
<tr>
<td>UNI-RE (F)</td>
<td>0.52 (0.29, 0.92)</td>
<td>0.83</td>
</tr>
<tr>
<td>Bivariate beta binomial</td>
<td>0.55 (0.24, 1.27)</td>
<td></td>
</tr>
</tbody>
</table>

Endoscopic better ← → open better
Surgical interventions for myopia correction

<table>
<thead>
<tr>
<th>Primary Model</th>
<th>OR (95% Cr.I.)</th>
<th>τ (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB-RE, high φ</td>
<td>1.24 (0.73, 1.90)</td>
<td>0.36 (0.06, 0.95)</td>
</tr>
</tbody>
</table>

Sensitivity analyses

**Bayesian**

- UNI-FE (B) 1.40 (1.00, 1.97)
- UNI-RE (B) 1.31 (0.77, 2.03) 0.29 (0.06, 0.88)
- BB-FE, low φ 1.39 (1.00, 1.94)
- BB-RE, low φ 1.29 (0.77, 2.01) 0.33 (0.09, 0.95)
- BB-FE, high φ 1.37 (1.00, 1.86)
- BB-RE, low φ 1.22 (0.61, 2.16) 0.60 (0.14, 1.31)

**Frequentist**

- UNI-FE (F) 1.40 (1.00, 1.95)
- UNI-RE (F) 1.40 (1.00, 1.95) 0.00
- Bivariate beta
  - Binomial 1.00 (0.94, 1.06)

PRK better ← ↔ LASIK better

Meta-analysis for uncorrected visual acuity (UCVA) of 20/20 or better, at six months after treatment, LASIK vs. PRK for myopia. Model abbreviations as per Figure 3.

159x129mm (150 x 150 DPI)
Appendix for “A model for meta-analysis of correlated binary outcomes: the case of split-body interventions”

1 Description of the data

1.1 CTS example

The available data is summarized in Table 1, and is also presented in the following forestplot. Note that for illustration purposes DZ studies are not shown in the graph and that for SZ studies we used a continuity correction.

Table 1: Available data for the CTS example. \( r_e (r_o) \) denotes events in endoscopic (open) surgical operation for CTS. \( N_e (N_o) \) denotes total number of patients in each arm.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study design</th>
<th>( r_e )</th>
<th>( N_e )</th>
<th>( r_o )</th>
<th>( N_o )</th>
<th>( N ) (patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unilateral</td>
<td>5</td>
<td>63</td>
<td>8</td>
<td>63</td>
<td>126</td>
</tr>
<tr>
<td>2</td>
<td>Unilateral</td>
<td>5</td>
<td>23</td>
<td>8</td>
<td>22</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>Unilateral</td>
<td>6</td>
<td>44</td>
<td>2</td>
<td>35</td>
<td>79</td>
</tr>
<tr>
<td>4</td>
<td>Unilateral</td>
<td>0</td>
<td>128</td>
<td>2</td>
<td>60</td>
<td>188</td>
</tr>
<tr>
<td>5</td>
<td>Unilateral</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>77</td>
<td>176</td>
</tr>
<tr>
<td>6</td>
<td>Unilateral</td>
<td>0</td>
<td>85</td>
<td>0</td>
<td>91</td>
<td>176</td>
</tr>
<tr>
<td>7</td>
<td>Unilateral</td>
<td>2</td>
<td>30</td>
<td>16</td>
<td>31</td>
<td>61</td>
</tr>
<tr>
<td>8</td>
<td>Unilateral</td>
<td>2</td>
<td>74</td>
<td>3</td>
<td>76</td>
<td>150</td>
</tr>
<tr>
<td>9</td>
<td>Unilateral</td>
<td>0</td>
<td>47</td>
<td>0</td>
<td>54</td>
<td>101</td>
</tr>
<tr>
<td>10</td>
<td>Unilateral</td>
<td>1</td>
<td>24</td>
<td>0</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>11</td>
<td>Unilateral</td>
<td>2</td>
<td>46</td>
<td>0</td>
<td>44</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>Unilateral</td>
<td>3</td>
<td>45</td>
<td>3</td>
<td>35</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>Bilateral</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>14</td>
<td>Bilateral</td>
<td>3</td>
<td>60</td>
<td>3</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>Bilateral</td>
<td>2</td>
<td>20</td>
<td>6</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>Mixed</td>
<td>2</td>
<td>82</td>
<td>3</td>
<td>65</td>
<td>122</td>
</tr>
<tr>
<td>17</td>
<td>Mixed</td>
<td>4</td>
<td>84</td>
<td>12</td>
<td>85</td>
<td>126</td>
</tr>
<tr>
<td>18</td>
<td>Mixed</td>
<td>3</td>
<td>53</td>
<td>21</td>
<td>52</td>
<td>71</td>
</tr>
<tr>
<td>19</td>
<td>Mixed</td>
<td>3</td>
<td>16</td>
<td>1</td>
<td>16</td>
<td>29</td>
</tr>
<tr>
<td>20</td>
<td>Mixed</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>21</td>
<td>Mixed</td>
<td>12</td>
<td>34</td>
<td>23</td>
<td>36</td>
<td>62</td>
</tr>
<tr>
<td>22</td>
<td>Mixed</td>
<td>0</td>
<td>97</td>
<td>0</td>
<td>95</td>
<td>147</td>
</tr>
<tr>
<td>23</td>
<td>Unknown</td>
<td>1</td>
<td>32</td>
<td>0</td>
<td>36</td>
<td>.</td>
</tr>
<tr>
<td>24</td>
<td>Unknown</td>
<td>1</td>
<td>50</td>
<td>0</td>
<td>50</td>
<td>.</td>
</tr>
</tbody>
</table>
1.2 Myopia example

The available data for this example is summarized in Table 2, also shown in the following forest plot.

*Table 2: Available data for the myopia example. $r_L(r_P)$ denotes events in LASIK (PRK) operation for myopia. $N_L(N_P)$ denotes total number of patients in each arm.*

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Study type</th>
<th>$r_L$</th>
<th>$N_L$</th>
<th>$r_P$</th>
<th>$N_P$</th>
<th>$N$ (patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Unilateral</td>
<td>1</td>
<td>25</td>
<td>3</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>Unilateral</td>
<td>16</td>
<td>61</td>
<td>13</td>
<td>68</td>
<td>129</td>
</tr>
<tr>
<td>3</td>
<td>Bilateral</td>
<td>46</td>
<td>50</td>
<td>47</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>Bilateral</td>
<td>19</td>
<td>28</td>
<td>12</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>Bilateral</td>
<td>23</td>
<td>25</td>
<td>24</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>Bilateral</td>
<td>30</td>
<td>33</td>
<td>32</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>Bilateral</td>
<td>43</td>
<td>44</td>
<td>42</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>Unknown</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Unknown</td>
<td>43</td>
<td>57</td>
<td>46</td>
<td>61</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Unknown</td>
<td>89</td>
<td>109</td>
<td>240</td>
<td>335</td>
<td>-</td>
</tr>
</tbody>
</table>

**Study**

<table>
<thead>
<tr>
<th>Study ID</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.24 (0.02, 2.47)</td>
</tr>
<tr>
<td>2</td>
<td>1.50 (0.66, 3.45)</td>
</tr>
<tr>
<td>3</td>
<td>0.73 (0.16, 3.46)</td>
</tr>
<tr>
<td>4</td>
<td>2.81 (0.95, 8.38)</td>
</tr>
<tr>
<td>5</td>
<td>0.48 (0.04, 5.65)</td>
</tr>
<tr>
<td>6</td>
<td>0.31 (0.03, 3.17)</td>
</tr>
<tr>
<td>7</td>
<td>2.05 (0.18, 23.44)</td>
</tr>
<tr>
<td>8</td>
<td>3.00 (0.11, 84.56)</td>
</tr>
<tr>
<td>9</td>
<td>1.00 (0.43, 2.32)</td>
</tr>
<tr>
<td>10</td>
<td>1.76 (1.03, 3.02)</td>
</tr>
</tbody>
</table>

PRK better ← → LASIK better
2 Details for the model used for analysing bilateral studies that do not provide cross-classifications

Assume a bilateral study \(i\), providing data for the number of events in each treatment \((r_{A,i} \text{ and } r_{B,i})\) and the number of patients, \(N_i\). Let us denote by \(p_{A,i}\) the probability of having an event in treatment A; likewise for B. The corresponding likelihood is as follows:

\[
L_i = \frac{N_i!}{\min(r_{A,i}, r_{B,i})} \sum_{n_{1,i}=\max(0, r_{A,i}+r_{B,i}-N_i)}^{\min(r_{A,i}, r_{B,i})} \frac{n_{1,i}!}{(r_{A,i} - n_{1,i})! (r_{B,i} - n_{1,i})! (N_i - r_{A,i} - r_{B,i} + n_{1,i})!} \times p_{1,i}^{n_{1,i}} p_{2,i}^{r_{A,i} - n_{1,i}} p_{3,i}^{r_{B,i} - n_{1,i}} p_{4,i}^{N_i - r_{A,i} - r_{B,i} + n_{1,i}}
\]

\[
p_{1,i} = p_{A,i} p_{B,i} + \varphi \sqrt{p_{A,i}(1 - p_{A,i})p_{B,i}(1 - p_{B,i})}
\]

\[
p_{2,i} = p_{A,i} - p_{1,i}
\]

\[
p_{3,i} = p_{B,i} - p_{1,i}
\]

\[
p_{4,i} = 1 - p_{1,i} - p_{2,i} - p_{3,i}
\]

\[
\text{logit}(p_{A,i}) = u_i + \frac{\theta_i}{2}
\]

\[
\text{logit}(p_{B,i}) = u_i - \frac{\theta_i}{2}
\]

\[
u_i \sim N(\mu_u, \sigma_u^2)
\]

\[
\theta_i \sim N(\mu, \tau^2)
\]

\[\mu_u, \mu, \sigma_u, \varphi \sim \ldots \text{ (prior distributions)}\]

The likelihood includes 3 independent parameters per study \(p_{1,i}, p_{2,i}, p_{3,i}\) (\(p_{4,i}\) can be calculated given these 3), but only 2 observed proportions are in the data (i.e. \(\frac{r_{A,i}}{N_i}, \frac{r_{B,i}}{N_i}\)). Including prior distributions for the 3 parameters would lead to identifiability issues. Instead, we add to the model external information regarding \(\varphi\), the 3 model parameters are mapped into \(p_{A,i}, p_{B,i}\) and \(\varphi\), rendering the model identifiable.

In this expression:

- \(n_{1,i}\) denotes the (unknown) number of patients having an event in both treatments.
- \(p_{1,i}\) denotes the probability of having an event in both treatments.
- \(p_{2,i}\) denotes the probability of having an event in treatment A but not treatment B.
• $p_{3,i}$ denotes the probability of having an event in treatment B but not treatment A.
• $p_{4,i}$ denotes the probability of having no events in either treatment.
• $\mu_{\alpha}$ denotes the mean log-odds of the average risk, and $\sigma_{\alpha}^2$ the corresponding variance.
• $\tau^2$ denotes the variance of random effects.
• $\varphi$ denotes the correlation.
• $\theta_i$ denotes the study-specific true treatment effects
• $\mu$ denotes the true underlying log-odds ratio
3 Alternative method for reconstructing the 2x2 table

In Section 3.3 of the main paper we discuss how researchers can stochastically reconstruct the 2x2 table (Table 2 in main paper) using the margins of the table and a prior distribution for the \( \varphi \) coefficient. Here we describe an alternative parameterization, which is based on the odds ratio of the cross-classified information.

We start by defining \( OR_{cc} = \frac{p_1 p_4}{p_2 p_3} \). By virtue of Equation 4 in the main paper, we can see that \( OR_{cc} > 1 \) corresponds to a positive correlation; \( OR_{cc} < 1 \) corresponds to negative correlation; \( OR_{cc} = 1 \) corresponds to zero correlation. Thus \( OR_{cc} \) can serve as an alternative measure of correlation.

We have \( p_A = p_1 + p_2, \ p_B = p_1 + p_3 \) and \( p_1 + p_2 + p_3 + p_4 = 1 \). Thus, given \( OR_{cc} \) and the margins (\( p_A \) and \( p_B \)) we can solve the system of equations to obtain \( p_1, p_2, p_3 \) and \( p_4 \). The solution for \( p_1 \) is the following:

\[
p_1 = \frac{C}{2(1 - OR_{cc})} \pm \frac{\sqrt{C^2 + 4 \ OR_{cc}(1 - OR_{cc})p_A p_B}}{2(1 - OR_{cc})}
\]

where

\[
C = 1 - (1 - OR)(p_A + p_B)
\]

The advantage of this method, as compared with the method based on \( \varphi \), is that \( OR_{cc} \) ranges from 0 to \( +\infty \), i.e. it is not bounded like \( \varphi \). The disadvantage is that it is much more difficult to decide or elicit clinically meaningful values for \( OR_{cc} \). Thus, we do not consider this approach any more in this paper.
4 Details for the model used for analysing mixed design studies
that do not provide cross-classifications

Assume a study $i$, which included a mixture of unilateral and bilateral interventions, and which provided information on the events in each treatment group ($r_{A,i}, r_{B,i}$), the total number of patients receiving each treatment ($N_{A,i}, N_{B,i}$), and the total number of patients in the study ($N_i$). The corresponding likelihood is as follows:

$$
L_i = \min(n_{A,i}, r_{A,i}) \min(n_{B,i}, r_{B,i}) \sum_{x_i=\max(0, r_{A,i} - N_{A,i})}^{N_{A,i}} \sum_{y_i=\max(0, r_{B,i} - N_{B,i})}^{N_{B,i}} \left( \frac{N_{A,i}^u}{x_i} \right) p_{A,i}^{x_i} (1 - p_{A,i})^{N_{A,i}^u - x_i} \times \left( \frac{N_{B,i}^u}{y_i} \right) p_{B,i}^{y_i} (1 - p_{B,i})^{N_{B,i}^u - y_i}
\times \left( \frac{N_{B,i}^b}{y_i} \right) p_{B,i}^{y_i} (1 - p_{B,i})^{N_{B,i}^b - y_i} \times \min(r_{A,i} - x_i, r_{B,i} - y_i) \sum_{n_{1,i}=\max(0, r_{A,i} - x_i, r_{B,i} - y_i - N_{B,i})}^{N_{B,i}^b - y_i} \frac{N_{B,i}^b}{n_{1,i}!(r_{A,i} - x_i - n_{1,i})!}
\times \frac{1}{(r_{B,i} - y_i - n_{1,i})!(N_{B,i}^b - r_{A,i} - r_{B,i} + x_i + y_i + n_{1,i})!}
\times p_{1,i}^{n_{1,i}} p_{2,i}^{r_{A,i} - x_i - n_{1,i}} p_{3,i}^{r_{B,i} - y_i - n_{1,i}} p_{4,i}^{r_{A,i} - x_i + y_i + n_{1,i}}
$$

$p_{1,i} = p_{A,i} p_{B,i} + \varphi \sqrt{p_{A,i}(1 - p_{A,i})p_{B,i}(1 - p_{B,i})}$

$p_{2,i} = p_{A,i} - p_{1,i}$

$p_{3,i} = p_{B,i} - p_{1,i}$

$p_{4,i} = 1 - p_{A,i} - p_{2,i} - p_{3,i}$

$logit(p_{A,i}) = u_i + \frac{\theta_i}{2}$

$logit(p_{B,i}) = u_i - \frac{\theta_i}{2}$

$u_i \sim N(\mu_u, \sigma_u^2)$

$\theta_i \sim N(\mu, \tau^2)$

$\mu_u, \mu, \tau, \sigma_u, \varphi \sim \ldots$ (prior distributions)

In this expression:

- $N_{A,i}^u$ denotes the number of unilateral patients receiving A. It is equal to $N_{A,i} = N_i - N_{B,i}$.
- $N_{B,i}^u$ denotes the number of unilateral patients receiving B. It is equal to $N_{B,i} = N_i - N_{A,i}$.
\( N^b_i \) denotes the number of bilateral patients. It is equal to \( N^b_i = N_{A,i} + N_{B,i} - N_i \).

\( x_i \) denotes the (unknown) number of events in treatment A that were observed in the unilateral patients. It holds that \( \max(0, r_{A,i} - N^b_i) < x_i < \min(N^u_{A,i}, r_{A,i}) \).

\( y_i \) denotes the (unknown) number of events in treatment B that were observed in the unilateral patients. It holds that \( \max(0, r_{B,i} - N^b_i) < y_i < \min(N^u_{B,i}, r_{B,i}) \).

The rest of the parameters are as described in the previous Section of this Appendix.
## 5 Summary table on how to choose the correct likelihood

*Table 3: Choosing the correct likelihood in each study depending on its design and the availability of cross-classified information.*

<table>
<thead>
<tr>
<th>Study design</th>
<th>All cross-classified information is available</th>
<th>cross classified information is not available</th>
</tr>
</thead>
<tbody>
<tr>
<td>unilateral</td>
<td>Use two univariate binomial distributions</td>
<td>(not applicable)</td>
</tr>
<tr>
<td>bilateral</td>
<td>Use a multinomial likelihood (equation 5 in paper)</td>
<td>Use a bivariate binomial likelihood (equation 6 in paper) with external information for φ</td>
</tr>
</tbody>
</table>
| mixed        | • Use two independent univariate binomial distributions for patients receiving unilateral interventions  
               • Use a multinomial likelihood for patients receiving bilateral interventions (equation 5 in paper) | Extended bivariate binomial distribution (equation 7 in paper) with external information for φ |
| unknown      | (not applicable)                             | If correlation is expected to be positive and we are interested in the most conservative estimate, treat as unilateral |
6 Eliciting prior distributions for the average risk

Here we discuss a method for eliciting expert opinion regarding the average risk. We also discuss how to combine opinions from different experts to formulate a prior distribution, possibly after weighting each expert’s opinion. The methods has been described in [1, 2].

We ask $N_E$ experts the following question:

“A total of $N$ patients were randomized to receive treatment $A$ or $B$ for the target condition (<specify treatments and condition>). What percentage of them do you expect to have a positive outcome (<specify outcome>)? Please provide a 95% confidence interval for this proportion.”

The $k$-th expert ($k = 1, \ldots, N_E$) provides an estimate for the 95% confidence interval for the baseline risk. Let us denote the average risk $p_b$. Assuming the expert’s opinion to come from a beta distribution we can construct a prior $\pi_k(p_b) = Beta(p_b; a_k, b_k)$ from the provided confidence interval. In [2] we had provided an R code for calculating the parameters $a_k$ and $b_k$ (freely available online at http://www.mtm.uoi.gr/images/BetaPriors%20for%20MONMA.txt.txt).

After calculating the individual $a_k$ and $b_k$ for each expert $k$, an overall prior can be obtained as a combination of the individual expert opinions:

$$f(p_b) \propto (\pi_1(p_b))^{w_1}(\pi_2(p_b))^{w_2} \ldots (\pi_{N_E}(p_b))^{w_{N_E}} = \prod_k (\pi_k(p_b))^{w_k}$$

(1)

The $w_k$ parameters are weights ($\sum_{k=1}^{N_E} w_k = 1$) that are assigned to experts. These can be chosen to reflect the credibility attached to each expert opinion. For example, we can define the weights based on the years of relevant clinical experience of each expert. Equation (1) suggests that the prior distribution for $p_b$ is a beta distribution $p_b \sim Beta(\sum w_k a_k, \sum w_k b_k)$. This can be used as a prior distribution for average risk.

In order to use this formulation, the model described in the paper (and also in the previous paragraph of this appendix) needs to be slightly changed:

$$\logit(p_{A,i}) = \logit(p_{b,i}) + \frac{\theta_i}{2}$$

$$\logit(p_{B,i}) = \logit(p_{b,i}) - \frac{\theta_i}{2}$$

$$p_{b,i} \sim Beta\left(\sum w_k a_k, \sum w_k b_k\right)$$
7 Eliciting prior distributions for the $\varphi$ coefficient (correlations)

We ask $N_E$ medical experts the following question:

“A number of patients received both A and B treatments (<specify treatments and condition>). We measure the number of events for a specific outcome (<specify outcome>). Some of the patients had an event only in A, some only in B, some in both treatments, and some in neither treatments.

We want to explore the relationship between having an event in one of the treatments with having an event in the other. For example, patients that experience an event in A may be more prone to have an event in B as well. Or, conversely, it might be the case that patients who had an event in A are less likely to have an event in B.

We want to quantify this relationship (correlation) using a measure ranging from -1 to +1. A correlation equal to -1 indicates a perfectly negative relationship between the two events, i.e. if a patient had an event in A he/she will not have the event in B.

A correlation equal to +1 indicates a perfect positive correlation, i.e. an event A guarantees an event in B.

A correlation of 0 means that the two events are completely independent, i.e. the probability of a patient having an event in A does not depend on whether or not he/she had an event in B.

Can you provide a range of values in which you are 95% certain that the true value of this correlation lies? Please provide values for $\varphi_{\text{lower}}$ and $\varphi_{\text{upper}}$, as described in the picture below.”

![Diagram showing the range of values for $\varphi$ with 95% confidence intervals.

https://mc.manuscriptcentral.com/smmr
We use the Fisher transformation, \( \zeta = \frac{1}{2} \log \left( \frac{\varphi + 1}{1 - \varphi} \right) \) to transform the elicited values for \( \varphi \) into the \((-\infty, +\infty)\), in order to construct a normal prior. Using the Fisher-transformed lower and upper values of the correlation (\( \zeta_{\text{lower}} \) and \( \zeta_{\text{upper}} \)) we can calculate \( \zeta_{\text{mean}} = (\zeta_{\text{upper}} + \zeta_{\text{lower}})/2 \) and \( \sigma_\zeta = (\zeta_{\text{upper}} - \zeta_{\text{lower}})/3.92 \).

Having calculated a \( \zeta_{\text{mean},k} \) and a \( \sigma^2_{\zeta,k} \) from each expert \( k \), we can combine all information by performing an inverse variance meta-analysis, where the input of each expert \( (\zeta_{\text{mean},k}) \) is weighted according to the certainty this expert places on the corresponding estimate (as quantified by \( \sigma^2_{\zeta,k} \)). In order to further weight expert opinions e.g. according to their experience, we can substitute the variance by \( \frac{\sigma^2_{\zeta,k}}{w_k} \) [3]. Here \( w_k \) measures the experience of expert \( k \), e.g. it can be taken to be equal to the years of experience in the field, or the number of RCTs in which this expert has been involved in.

Using the outcome of this meta-analysis (pooled estimate and standard error) we can formulate a distribution for \( \zeta \), to use in a Bayesian setting as a prior.
8 R code for generating the data used in the simulations

For each scenario we generate 100 meta-analyses, for each meta-analysis we generate 20 studies. For the $k$-th meta-analysis we start by defining $\mu_k, \mu_{u,k}, \sigma_{u,k}, \varphi_k$. We draw $\tau_k^2 \sim \text{LN}(-2, 1.5)$. The parameters of this distribution were chosen to approximate empirical distributions described in [4]. For study $i$ in meta-analysis $k$ we generate $\theta_{i,k} \sim N(\mu_k, \tau_k^2), u_{i,k} \sim N(\mu_{u,k}, \sigma_{u,k}^2)$. We generate the number of patients in each study, $N_{i,k}$ by drawing from a uniform distribution, $U(50, 100)$.

```r
library(MASS)
Nstudies<- 20
Nmeta=100
mu=0.5
phi=0.6
tau.s=rlnorm(Nmeta, meanlog = -2, sdlog = 1.5)
Np=array(rep(0,Nmeta*Nstudies), dim=c(Nmeta,Nstudies))
for (k in 1:Nmeta){
  for (j in 1:Nstudies){
    Np[k,j]=round(runif(1,50,100))} #simulate number of patients
}
data4=array(rep(0,Nstudies*Nmeta*8), dim=c(Nmeta,Nstudies,20))
for (k in 1:Nmeta){
  Sigma <- matrix(c(tau.s[k],0,0,1),2,2)
  aaa=mvrnorm(n =Nstudies, c(mu,-3), Sigma)
  for (j in 1:Nstudies){
    data4[k,j,1]=aaa[j,1] #logor: logit(pA)-logit(pB)
    data4[k,j,2]=aaa[j,2] ###u: logit(pA)+logit(pB)/2
    data4[k,j,3]=-data4[k,j,1]/2+data4[k,j,2] ###logit(pA)
    data4[k,j,4]=+data4[k,j,1]/2+data4[k,j,2] ###logit(pB)
  }
}
for (k in 1:Nmeta){
  for (j in 1:Nstudies){
    data4[k,j,3]=exp(data4[k,j,3])/(1+exp(data4[k,j,3]))###pA
  }
}
for (k in 1:Nmeta){
  for (j in 1:Nstudies){
    data4[k,j,4]=exp(data4[k,j,4])/(1+exp(data4[k,j,4])) ###pB
  }
}
for (k in 1:Nmeta){
  for (j in 1:Nstudies){
    data4[k,j,20]=max(data4[k,j,3],data4[k,j,4])###pimin
    data4[k,j,19]=min(data4[k,j,3],data4[k,j,4])###pimax
    data4[k,j,18]=sqrt(data4[k,j,19]*(1-data4[k,j,20])/(data4[k,j,20]*(1-data4[k,j,19])))###pimax
    data4[k,j,16]=max(0,data4[k,j,3]+data4[k,j,4]-1)###test
  }
}
for (k in 1:Nmeta){
  for (j in 1:Nstudies){
    data4[k,j,17]=
    (data4[k,j,16]*(1-data4[k,j,3]-data4[k,j,4]+data4[k,j,16])-
    (data4[k,j,3]-data4[k,j,16])*(data4[k,j,4]-data4[k,j,16]))
  }
}```
\[
\sqrt{\text{data4}[k,j,3]*\text{data4}[k,j,4]*(1-\text{data4}[k,j,3])*(1-\text{data4}[k,j,4])}
\]

\[
\text{data4}[k,j,5]=\max(0.0000001,\text{data4}[k,j,3]*\text{data4}[k,j,4]+\min(\max(\phi,\text{data4}[k,j,17]),\text{data4}[k,j,18]))*\sqrt{\text{data4}[k,j,3]*\text{data4}[k,j,4]*(1-\text{data4}[k,j,3])*(1-\text{data4}[k,j,4])})
\]

\[
\text{data4}[k,j,6]=\max(0.0000001,\text{data4}[k,j,3]-\text{data4}[k,j,5])
\]

\[
\text{data4}[k,j,7]=\max(0.0000001,\text{data4}[k,j,4]-\text{data4}[k,j,5])
\]

\[
\text{data4}[k,j,8]=1-\text{data4}[k,j,5]-\text{data4}[k,j,6]-\text{data4}[k,j,7]
\]

\[
\text{aa2}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies},4))
\]

\[
\text{for (}k \text{ in } 1: \text{Nmeta}){
    \text{for (}j \text{ in } 1: \text{Nstudies}){
        \text{aa2}[k,j,1:4] = \text{rmultinom}(1, \text{Np}[k,j], \text{data4}[k,j,5:8])
    }
}\]

\[
\text{n11}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies}))
\]

\[
\text{n12}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies}))
\]

\[
\text{n13}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies}))
\]

\[
\text{n14}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies}))
\]

\[
\text{for (}k \text{ in } 1: \text{Nmeta}){
    \text{for (}j \text{ in } 1: \text{Nstudies}){
        \text{n11}[k,j] = \text{aa2}[k,j,1]
        \text{n12}[k,j] = \text{aa2}[k,j,2]
        \text{n13}[k,j] = \text{aa2}[k,j,3]
        \text{n14}[k,j] = \text{aa2}[k,j,4]
    }
}\]

\[
\text{nA}=\text{n11}+\text{n12}
\]

\[
\text{nB}=\text{n11}+\text{n13}
\]

\[
\text{Np}=\text{n11}+\text{n12}+\text{n13}+\text{n14}
\]

\[
\text{lower}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies}))
\]

\[
\text{range1}=\text{array}(\text{rep}(0,\text{Nmeta}*\text{Nstudies}), \text{dim}=\text{c}(\text{Nmeta},\text{Nstudies}))
\]

\[
\text{for (}k \text{ in } 1: \text{Nmeta}){
    \text{for (}j \text{ in } 1: \text{Nstudies}){
        \text{lower}[k,j] = \max(0, \text{nA}[k,j]+\text{nB}[k,j]-\text{Np}[k,j])
    }
}\]

\[
\text{for (}k \text{ in } 1: \text{Nmeta}){
    \text{for (}j \text{ in } 1: \text{Nstudies}){
        \text{range1}[k,j] = 1+\min(\text{nA}[k,j], \text{nB}[k,j])-\max(0, \text{nA}[k,j]+\text{nB}[k,j]-\text{Np}[k,j])
    }
}\]

\[
\text{sum(}\text{nA}[\cdot])
\]

\[
\text{sum(}\text{nB}[\cdot])
\]

\[
\text{dput(} \text{Nstudies})
\]

\[
\text{dput(} \text{nA})
\]

\[
\text{dput(} \text{nB})
\]

\[
\text{dput(} \text{Np})
\]

\[
\text{dput(} \text{lower})
\]

\[
\text{dput(} \text{range1})
\]
9 OpenBUGS code for analysing the simulated datasets

Data needed: Nmeta (number of meta-analyses), NS (number of studies), r.e.II (events in treatment A), r.o.II (events in treatment B), Npatients.II (#patients per study), lower.II (equal to max(0, r.e.II + r.o.II - Npatients.II)), range.II (equal to 1 + min(r.e.II, r.o.II) - max(0, r.e.II + r.o.II - Npatients)). All inputs except Nmeta and NS are data structures of dimensions NSxNmeta.

model{
  for (i in (1):(Nmeta)){
    for (k in (1):(NS)){
      pmax[k,i]<-max(p.e[k,i],p.o[k,i])
      pmin[k,i]<-min(p.e[k,i],p.o[k,i])
      phimax[k,i]<-sqrt(pmin[k,i]*(1-pmax[k,i])/(pmax[k,i]*(1-pmin[k,i])))
      # study-specific max value for phi
      fisher.phimax[k,i]<-1/2*log((1+phimax[k,i])/(1-phimax[k,i]))
      test[k,i]<-max(0,p.e[k,i]+p.o[k,i]-1)
      phimin[k,i]<-((test[k,i]*(1-p.e[k,i]-p.o[k,i]+test[k,i])-(p.e[k,i]-test[k,i])*(p.o[k,i]-test[k,i]))/
      sqrt(p.e[k,i]*p.o[k,i]*(1-p.e[k,i])*(1-p.o[k,i])))
      fisher.phimin[k,i]<-1/2*log((1+phimin[k,i])/(1-phimin[k,i]))
      # the elicited distribution for Fisher-transformed phi is truncated to a
      # study-specific range of allowed values
      for (i in (1):(Nmeta)){
        for (k in (1):(NS)){
          Fphi.study1[k,i]~ dunif(0,1)
          Fphi.study2[k,i] <-min(Fphi.study1[k,i],fisher.phimax[k,i])
          Fphi.study[k,i] <-max(Fphi.study2[k,i],phimin[k,i])
        }
      }
      # the 2x2 table is reconstructed
      for (i in (1):(Nmeta)){
        for (k in (1):(NS)){
          phi.study[k,i]<-tanh(Fphi.study[k,i])
          p[k,i,1]<-p.e[k,i]*p.o[k,i]+phi.study[k,i]*sqrt(p.e[k,i]*p.o[k,i]*(1-
          p.e[k,i])*(1-p.o[k,i]))
          p[k,i,2]<-p.e[k,i]-p[k,i,1]
          p[k,i,3]<-p.o[k,i]-p[k,i,1]
          p[k,i,4]<-1-p[k,i,1]-p[k,i,2]-p[k,i,3]
        }
      }
      # likelihood of the observations
      for (i in (1):(Nmeta)){
        for (m in (1):(NS)){
          dummy[m,i]<- 0  # dummy variable
          dummy[m,i] ~ dloglik(logLike.II[m,i])  # likelihood is exp(logLike.II[i])
        }
        for (k in 1:range.II[m,i]) {n11[m,i,k]<-lower.II[m,i]+k-1}
        for (k in 1:range.II[m,i]) {
          loglike.II[m,i,k]<- (logfact(Npatients.II[m,i,1]))
          +log(max(0.0000001,p[m,i,2]))*(r.e.II[m,i]-n11[m,i,k])
          +log(max(0.0000001,p[m,i,3]))*(r.o.II[m,i]-n11[m,i,k])
          +log(max(0.0000001,p[m,i,1]))*(n11[m,i,k])
          +log(max(0.0000001,p[m,i,4]))*(Npatients.II[m,i]-r.e.II[m,i]-
          r.o.II[m,i]+n11[m,i,k])
        }
      }
    }
  }
}

https://mc.manuscriptcentral.com/smmr
like[m,i,k]<exp(loglike.II[m,i,k])
A1[m,i]<max(0,sum(like[m,i,1:range.II[m,i]]))
logLike.II[m,i]<-log(A1[m,i])
##prior distributions and parameterizing
for (i in (1):(Nmeta)){
  for (k in 1:(NS)){
    logit(p.e[k,i]) <- theta[k,i,1]-theta[k,i,2]/2
    logit(p.o[k,i]) <- theta[k,i,1]+theta[k,i,2]/2
    theta[k,i,1:2] ~dnorm(ml[i,1:2],prectau[i,1:2,1:2])
  }
  m1[i,1]<aa[i]
  m1[i,2]~unif(-2,4)
  ss[i,1,1]<-(bb[i]*bb[i])
  ss[i,1,2]<-0
  ss[i,2,1]<-0
  ss[i,2,2]<tau.s[i]
  prectau[i,1:2,1:2]<inverse(ss[i,1:2,1:2])
  tau.s[i]<unif(0,5)
  aa[i]<unif(-10,10)
  bb[i]<unif(0,5)
10 Simulations

10.1 Scenarios 1-20

Here we provide a description of scenarios 1-20. In the following table \( \varphi \) is the correlation coefficient assumed in each scenario, and \( \zeta \) is the Fisher transformation of \( \varphi \). \% SZ (\%DZ) shows the percent of single-zero (double-zero) studies in each scenario. \( \mu_\zeta \) and \( \sigma_\zeta \) denote the mean and the variance of the average log-odds of event in the studies. For each scenario we simulated 100 meta-analyses. Each meta-analysis included 20 studies. Each study had 50-100 simulated patients.

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>Scenario description</th>
<th>( \zeta )</th>
<th>( \mu_\zeta )</th>
<th>( \sigma_\zeta )</th>
<th>% SZ (% DZ)</th>
<th>( \mu )</th>
<th>Priors used in the analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Zero relative effects Low correlation No SZ and DZ studies</td>
<td>0.31</td>
<td>-1.5</td>
<td>0.5</td>
<td>0.2% (0.0%)</td>
<td>0.0</td>
<td>N(0.3,0.1(^2)) U(-4,0) U(0,3) LN(-2,1.5(^2)) U(-2,2)</td>
</tr>
<tr>
<td>2</td>
<td>Zero relative effects Low correlation, Some SZ and DZ studies</td>
<td>0.31</td>
<td>-3.5</td>
<td>1.0</td>
<td>20% (10%)</td>
<td>0.0</td>
<td>N(0.3,0.1(^2)) U(-5,-1) U(0,3) LN(-2,1.5(^2)) U(-2,2)</td>
</tr>
<tr>
<td>3</td>
<td>Zero relative effects Low correlation, Many SZ and DZ studies</td>
<td>0.31</td>
<td>-4.0</td>
<td>1.0</td>
<td>25% (19%)</td>
<td>0.0</td>
<td>N(0.3,0.1(^2)) U(-6,-1) U(0,3) LN(-2,1.5(^2)) U(-2,2)</td>
</tr>
<tr>
<td>4</td>
<td>Zero relative effects Large correlation No SZ and DZ studies</td>
<td>0.69</td>
<td>-1.5</td>
<td>0.5</td>
<td>0.9% (0.2%)</td>
<td>0.0</td>
<td>N(0.7,0.2(^2)) U(-4,0) U(0,3) LN(-2,1.5(^2)) U(-2,2)</td>
</tr>
<tr>
<td>5</td>
<td>Zero relative effects Large correlation Some SZ and DZ studies</td>
<td>0.69</td>
<td>-3.5</td>
<td>1.0</td>
<td>14% (14%)</td>
<td>0.0</td>
<td>N(0.7,0.2(^2)) U(-5,-1) U(0,3) LN(-2,1.5(^2)) U(-2,2)</td>
</tr>
<tr>
<td></td>
<td>Relative Effects</td>
<td>Correlation</td>
<td>Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>-------------</td>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>6</td>
<td>Zero relative effects</td>
<td>Large correlation</td>
<td>Many SZ and DZ studies</td>
<td>0.69</td>
<td>-4.0</td>
<td>1.0</td>
<td>15% (20%)</td>
</tr>
<tr>
<td>7</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>No SZ and DZ studies</td>
<td>0.31</td>
<td>-1.5</td>
<td>0.5</td>
<td>0.8% (0.1%)</td>
</tr>
<tr>
<td>8</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>Some SZ and DZ studies</td>
<td>0.31</td>
<td>-3.5</td>
<td>1.0</td>
<td>21% (9%)</td>
</tr>
<tr>
<td>9</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>Some SZ and DZ studies</td>
<td>0.31</td>
<td>-3.5</td>
<td>1.0</td>
<td>21% (9%)</td>
</tr>
<tr>
<td>10</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>Some SZ and DZ studies</td>
<td>0.31</td>
<td>-4.0</td>
<td>1.0</td>
<td>28% (18%)</td>
</tr>
<tr>
<td>11</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>Many SZ and DZ studies</td>
<td>0.31</td>
<td>-4.0</td>
<td>1.0</td>
<td>28% (18%)</td>
</tr>
<tr>
<td>12</td>
<td>Non-zero relative effects</td>
<td>Large correlation</td>
<td>No SZ and DZ studies</td>
<td>0.69</td>
<td>-1.5</td>
<td>0.5</td>
<td>0.8% (0.1%)</td>
</tr>
<tr>
<td>13</td>
<td>Non-zero relative effects</td>
<td>Large correlation</td>
<td>Some SZ and DZ studies</td>
<td>0.69</td>
<td>-3.5</td>
<td>1.0</td>
<td>18% (12%)</td>
</tr>
<tr>
<td>14</td>
<td>Non-zero relative effects</td>
<td>Large correlation</td>
<td>Some SZ and DZ studies</td>
<td>0.69</td>
<td>-3.5</td>
<td>1.0</td>
<td>18% (12%)</td>
</tr>
<tr>
<td>15</td>
<td>Non-zero relative effects</td>
<td>Large correlation</td>
<td>Vague priors</td>
<td>0.69</td>
<td>-3.5</td>
<td>1.0</td>
<td>18% (12%)</td>
</tr>
<tr>
<td>16</td>
<td>Non-zero relative effects</td>
<td>Large correlation</td>
<td>Many SZ and DZ studies</td>
<td>Misspecified priors</td>
<td>0.69</td>
<td>-4.0</td>
<td>1.0</td>
</tr>
<tr>
<td>17</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>Many SZ and DZ studies</td>
<td>Vague priors</td>
<td>0.31</td>
<td>-3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>18</td>
<td>Non-zero relative effects</td>
<td>Low correlation</td>
<td>Some SZ and DZ studies</td>
<td>Vague priors</td>
<td>0.31</td>
<td>-2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>19</td>
<td>Non-zero relative effects</td>
<td>Large correlation</td>
<td>Many SZ and DZ studies</td>
<td>Vague priors</td>
<td>0.69</td>
<td>-3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>20</td>
<td>Non-zero relative effects</td>
<td>Zero correlation</td>
<td>Some SZ and DZ studies</td>
<td>Vague priors</td>
<td>0.00</td>
<td>-3.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>
### 10.2 Scenarios 21 and 22

In the following table we present an overview of scenarios 21-22 and the priors used in their analyses.

**Table 5: Detailed description of the scenarios 21 and 22, and of the priors used in their analyses**

<table>
<thead>
<tr>
<th>Scenario #</th>
<th>Scenario description</th>
<th>Description of the simulated datasets</th>
<th>( \zeta )</th>
<th>( \mu_x )</th>
<th>( \sigma_x )</th>
<th>% SZ (% DZ)</th>
<th>( \mu )</th>
<th>( \zeta )</th>
<th>( \mu_x )</th>
<th>( \sigma_x )</th>
<th>( \tau^2 )</th>
</tr>
</thead>
</table>
| 21        | Non-zero relative effects  
Low correlation  
Many SZ and DZ studies  
10 studies per meta-analysis | \[ \begin{align*} 0.31 & \quad -3.0 & \quad 1.0 & \quad 25\% (14\%) \quad 0.5 \\ U(0,10) & U(-10,10) & U(0,5) & U(0,5) \\ U(0,10) & U(-10,10) & U(0,5) & LN(-2,1.5^2) \\ N(0.3,0.1^2) & U(-10,10) & U(0,5) & U(0,5) \\ U(0,10) & U(-4,2) & U(0,2) & U(0,5) \end{align*} \] | | | | | | | | | |
| 22        | Non-zero relative effects  
Large correlation  
Many SZ and DZ studies  
10 studies per meta-analysis | \[ \begin{align*} 0.69 & \quad -3.0 & \quad 1.0 & \quad 20\% (17.6\%) \quad 0.5 \\ U(0,10) & U(-10,10) & U(0,5) & U(0,5) \\ U(0,10) & U(-10,10) & U(0,5) & LN(-2,1.5^2) \\ N(0.7,0.1^2) & U(-10,10) & U(0,5) & U(0,5) \\ U(0,10) & U(-4,2) & U(0,2) & U(0,5) \end{align*} \] | | | | | | | | | |
11 OpenBUGS code used for the analysis of real data-sets
Here we provide the OpenBUGS code as well as the data used for the analysis of the CTS example.

model{
### two interventions, endoscopic (e) and open (o) for CTS

##### UNILATERAL STUDIES ######
#likelihood in each study i
for (i in 1:NS1){
r.e.I[i]~dbin(p.e[i],N.e.I[i])
r.o.I[i]~dbin(p.o[i],N.o.I[i])
}

# parameters of normal prior distribution for the Fisher-
#transformed correlation coefficient (\( \zeta \)), needed for bilateral and mixed studies
m.fisherphi<-0.2 # mean
sigma.fisherphi<-0.1 # standard deviation
prec.fisherphi<-1/(sigma.fisherphi*sigma.fisherphi)# precision

# study specific max-min values for phi
for (i in (NS1+1):(NS1+NS2+NS3)){
pmax[i]<-max(p.e[i],p.o[i])
pmin[i]<-min(p.e[i],p.o[i])
phimax[i]<-sqrt(pmin[i]*(1-pmax[i])/(pmax[i]*(1-pmin[i])))
fisher.phimax[i]<-1/2*log((1+phimax[i])/(1-phimax[i]))
test[i]<-max(0,p.e[i]+p.o[i]-1)
phimin[i]<-((test[i]*(1-p.e[i]-p.o[i]+test[i])-(p.e[i]-test[i])*(p.o[i]-test[i]))/
sqrt(p.e[i]*p.o[i]*(1-p.e[i])*(1-p.o[i])))
fisher.phimin[i]<-1/2*log((1+phimin[i])/(1-phimin[i]))
}

# the elicited distribution for Fisher-transformed phi is
# truncated to the study-specific range of allowed values
for (i in (NS1+1):(NS1+NS2+NS3)){
    Fphi.study1[i] ~ dnorm(m.fisherphi,prec.fisherphi)
    Fphi.study2[i] <-min(Fphi.study1[i],fisher.phimax[i])
    Fphi.study[i] <-max(Fphi.study2[i], fisher.phimin[i])
}

# probabilities of the 2x2 table are calculated
for (i in (NS1+1):(NS1+NS2+NS3)){
phi.study[i]<-tanh(Fphi.study[i])
p[i,1]<-p.e[i]*p.o[i]+phi.study[i]*sqrt(p.e[i]*p.o[i]*(1-
p.e[i])*(1-p.o[i]))
p[i,2]<-p.e[i]-p[i,1]
p[i,3] <- p.o[i] - p[i,1]
p[i,4] <- 1 - p[i,1] - p[i,2] - p[i,3]

###################################################################
#####  BILATERAL STUDIES  ######
###################################################################

# defining the likelihood of the observations
for (i in (NS1+1):(NS1+NS2)){
dummy[i] <- 0  # dummy variable
dummy[i] ~ dloglik(logLike.II[i])  # likelihood is exp(logLike.II[i])
for (k in 1:range.II[i-NS1]) {n11[i,k] <- lower.II[i-NS1]+k-1}
for (k in 1:range.II[i-NS1]) {
  loglike.II[i,k] <-
  (logfact(Npatients.II[i-NS1]))
  +log(max(0.000001,p[i,2]))*(r.e.II[i-NS1]-n11[i,k])
  +log(max(0.000001,p[i,3]))*(r.o.II[i-NS1]-n11[i,k])
  +log(max(0.000001,p[i,1]))*(n11[i,k])
  +log(max(0.000001,p[i,4]))*(Npatients.II[i-NS1]-r.e.II[i-NS1]-
                             r.o.II[i-NS1]+n11[i,k])
  -
  (logfact(n11[i,k]))
  +(logfact((r.e.II[i-NS1]-n11[i,k])))
  +(logfact((r.o.II[i-NS1]-n11[i,k])))
  +(logfact((Npatients.II[i-NS1]-r.e.II[i-NS1]-r.o.II[i-
                             NS1]+n11[i,k])))
}
like[i,k] <- exp(loglike.II[i,k])
}
A1[i] <- max(0, sum(like[i,1:range.II[i-NS1]]))
logLike.II[i] <- log(A1[i])

###################################################################
#####  MIXED STUDIES  ######
###################################################################

for (ns in (NS1+NS2+1):(NS1+NS2+NS3)){
dummy[ns] <- 0
dummy[ns] ~ dloglik(logLike.III[ns])

for (i in lower.e.III[ns-NS1-NS2]:upper.e.III[ns-NS1-NS2]){  # the likelihood of each possible splitting of the endoscopic
  like_e[ns, i+1] <-
pow(p.e[ns],i)*
pow(1-p.e[ns],N.e.u.III[ns-NS1-NS2]-i)
  *exp(logfact(N.e.u.III[ns-NS1-NS2]))/
  (exp(logfact(i)) * exp(logfact(N.e.u.III[ns-NS1-NS2]-i))))
for (i in lower.o.III[ns-NS1-NS2]:upper.o.III[ns-NS1-NS2]) {
    # the likelihood of each possible splitting of the open
    like_o[ns, i+1]<-pow(p.o[ns],i)*pow(1-p.o[ns],N.o.u.III[ns-NS1-NS2]-i)*exp(logfact(N.o.u.III[ns-NS1-NS2]))/(exp(logfact(i))*exp(logfact(N.o.u.III[ns-NS1-NS2]-i)))
}

for (i in (r.e.III[ns-NS1-NS2]-upper.e.III[ns-NS1-NS2]):(r.e.III[ns-NS1-NS2]-lower.e.III[ns-NS1-NS2])){
    ## defining the part of the likelihood due to bilaterals
    for (j in (r.o.III[ns-NS1-NS2]-upper.o.III[ns-NS1-NS2]):(r.o.III[ns-NS1-NS2]-lower.o.III[ns-NS1-NS2])){
        for (k in 0:Ntest) {
            Test[ns,i+1,j+1,k+1]<- (exp(logfact(N.b.III[ns-NS1-NS2]))
                *pow(p[ns,2],i-k)
                *pow(p[ns,3],j-k)
                *pow(p[ns,1],k)
                *pow(p[ns,4], N.b.III[ns-NS1-NS2]-i-j+k)
            /
            (exp(logfact(k))
            *exp(logfact(i-k*step(min(i,j)-k)))
            *exp(logfact(j-k*step(min(i,j)-k)))
            *exp(logfact((N.b.III[ns-NS1-NS2]-i-j+k)*step(k-max(0,i+j-N.b.III[ns-NS1-NS2])))
            ))
            *step(k-max(0,i+j-N.b.III[ns-NS1-NS2]))
            *step(min(i,j)-k)

            likeBilateralsSub[ns,i+1,j+1,k+1]<-
            Test[ns,i+1,j+1,k+1]*like_e[ns, r.e.III[ns-NS1-NS2]-i+1]*like_o[ns, r.o.III[ns-NS1-NS2]-j+1] }

    test1[ns,i+1,j+1]<- sum(likeBilateralsSub[ns,i+1,j+1,1:(Ntest+1)])
    test2[ns,i+1]<-sum(test1[ns,i+1,(r.o.III[ns-NS1-NS2]-upper.o.III[ns-NS1-NS2]+1):(r.o.III[ns-NS1-NS2]-lower.o.III[ns-NS1-NS2]+1)])
    test3[ns]<-max(0,sum(test2[ns, (r.e.III[ns-NS1-NS2]-upper.e.III[ns-NS1-NS2]+1):(r.e.III[ns-NS1-NS2]-lower.e.III[ns-NS1-NS2]+1)])

    for (ns in (NS1+NS2+1):(NS1+NS2+NS3)) {
        C1[ns]<- log(test3[ns])
        logLike.III[ns]<- C1[ns]
    }

    ## prior distributions, parameterizing
    for (i in 1:(NS1+NS2+NS3)) {
        logit(p.e[i]) <- theta[i,1]-theta[i,2]/2
        logit(p.o[i]) <- theta[i,1]+theta[i,2]/2
    }
}
\[ \theta[i,1:2] \sim \text{dmnorm}(m1[1:2], \text{prectau}[1:2,1:2]) \]

\}

\begin{verbatim}
theta[1]<-aa
theta[2]<-dunif(-2,2)

ss[1,1]<-(bb*bb)
ss[1,2]<-0#r[i]*bb[i]*sqrt(tau.s[i])
ss[2,1]<-0#r[i]*bb[i]*sqrt(tau.s[i])
ss[2,2]<-tau.s

prectau[1:2,1:2]<-inverse(ss[1:2,1:2])

tau.s~dlnorm(-1.43,0.48)
#r[i]~dunif(-1,1)

aa~dunif(-5,-1)
bb~dunif(0,3)
\end{verbatim}

#### DATA for CTS

list(
    NS1=14, NS2=3, NS3=7, # number of studies in each design (unilateral, bilateral, mixed).

    ###### data for unilateral studies
    r.e.I=c(5, 5, 6, 0, 0, 2, 0, 1, 3, 1, 1), # number of events in endoscopic, unilateral studies
    N.e.I=c(63, 23, 44, 128, 99, 85, 30, 74, 47, 24, 45, 32, 36, 50), # total number of patients in endoscopic, unilateral studies
    r.o.I=c(8, 8, 2, 2, 0, 0, 16, 3, 0, 0, 0, 3, 0, 0), # number of events in open, unilateral studies
    N.o.I=c(63, 22, 35, 60, 77, 91, 31, 76, 54, 24, 44, 35, 36, 50), # total number of patients in open, unilateral studies

    ###### data for bilateral studies
    r.e.II=c(1, 3, 2), # events endoscopic, bilateral studies
    r.o.II=c(2, 3, 6), # events open, bilateral studies
    Npatients.II=c(25, 60, 20), # number of patients, bilateral studies
    lower.II=c(0, 0, 0), # this gives the starting value of n1 in the sum of equation 3 in the main paper. It is a vector equal to max(0, r.e.II+r.o.II-Npatients.II) for each study
    range.II=c(2, 4, 3), # this gives the number of possible values for n1 in the sum of equation 5 in the main paper. It is a vector equal to 1+min(r.e.II, r.o.II)-max(0, r.e.II+r.o.II-Npatients.II) for each study

    ###### data for mixed studies

\end{verbatim}
r.e.III=c( 2, 4, 3, 3, 0,12, 0),## total events endoscopic, mixed studies
r.o.III=c( 3,12,21, 1, 0,23, 0), ## total events open, mixed studies
N.b.III=c(25,43,34, 3, 1, 8,45), ## patients receiving bilateral interventions, mixed studies
N.e.u.III=c(57,41,19,13, 7,26,52), ### patients unilateral, endoscopic, mixed studies
N.o.u.III=c(40,42,18,13, 8,28,50),### patients unilateral, open, mixed studies
Ntest=24, ## this is just an arbitrary large number. To facilitate convergence choose Ntest= max(r.e.III[i],r.o.III[i])+1,where i runs across all mixed studies.

## boundaries for the number of events in unilateral interventions, for mixed studies
## ENDOSCOPIC
lower.e.III=c(0,0,0,0,0,4,0), ## this vector is equal to max(0,r.e.III-N.b.III)
upper.e.III=c(2,4,3,3,0,12,0) ## this vector is equal to min(N.e.u.III,r.e.III)

## OPEN
lower.o.III=c(0,0,0,0,0,15,0), ## this vector is equal to max(0,r.o.III-N.b.III)
upper.o.III=c(3,12,18,1,0,23,0) ## this vector is equal to min(N.o.u.III,r.o.III)
12 R code used for the analysis of real data-sets

Here we provide the R2WinBUGS code for fitting the model presented in this paper. The code provided in the previous section of this appendix needs to be saved in a txt file. The data correspond to the CTS example.

```r
library(R2WinBUGS)
library(BRugs)

NS1=14
NS2=3
NS3=7

r.e.I=c( 5, 5, 6, 0, 0, 2, 2, 0, 1, 2, 3,1,1) # number of events in endoscopic, unilateral studies
N.e.I=c(63,23,44,128,99,85,30,47,24,46,45,32,50) ## total number of patients in endoscopic, unilateral studies
r.o.I=c( 8, 8, 2, 0, 0,16, 3, 0, 0, 3,0,0) ## number of events in open, unilateral studies
N.o.I=c(63,22,35,60,77,91,31,76,54,24,44,35,36,50)## total number of patients in open, unilateral studies
r.e.II=c( 1, 3, 2 ) ## events endoscopic, bilateral studies
r.o.II=c( 2, 3, 6) ## events open, bilateral studies
Npatients.II=c(25,60,20)

r.e.III=c( 2, 4, 3, 3, 0,12, 0) ## total events endoscopic, mixed studies
r.o.III=c( 3,12,21, 1, 0,23, 0) ## total events open, mixed studies
N.b.III=c(25,43,34, 3, 1, 8,45) ## patients with bilateral interventions, mixed studies. This vector is equal to: Npatients(in endoscopic)+Npatients(open)-Npatients(all), see section 3.4 of the paper
N.e.u.III=c(57,41,19,13, 7,26,52) ### patients unilateral endoscopic, mixed studies
N.o.u.III=c(40,42,18,13, 8,28,50)### patients unilateral open, mixed studies

Ntest=max(r.e.III[i],r.o.III[i])+1 ## this is just an arbitrary large number

for(i in 1:length(Npatients.II)) {lower.II[i]=max(0,r.e.II[i]+r.o.II[i]-Npatients.II[i])
for(i in 1:length(Npatients.II)) {range.II[i]=1+min(r.e.II[i],r.o.II[i])-max(0,r.e.II[i]+r.o.II[i]-Npatients.II[i])
for(i in 1:length(Npatients.II)) {lower.e.III[i]=max(0,r.e.III[i]-N.b.III[i])
for(i in 1:length(Npatients.II)) {upper.e.III[i]=min(N.e.u.III[i],r.e.III[i])
for(i in 1:length(Npatients.II)) {lower.o.III[i]=max(0,r.o.III[i]-N.b.III[i])
for(i in 1:length(Npatients.II)) {upper.o.III[i]=min(N.o.u.III[i],r.o.III[i])
```

https://mc.manuscriptcentral.com/smmr
```r
data1 <- list("NS1", "NS2", "NS3","r.e.I","N.e.I","r.o.I","N.o.I","r.e.II","r.o.II","Npatients.II","lower.II","range.II","r.e.III","r.o.III","N.b.III","N.e.u.III","N.o.u.III","Ntest","lower.e.III","upper.e.III","lower.o.III","upper.o.III")

setwd("C:/…")  # input working directory here
bugs.data(data1, dir = getwd(), digits = 5, data.file = "data1.txt")

parameters1 = c("or","tau.s")
sim=openbugs(data=data1,parameters=parameters1,
inits=NULL,model.file="model.bivRE.txt",
n.chains=1, n.iter=2000, n.burnin=1000,
n.thin=1,
n.sims = 2000,
bugs.directory="C:/Program Files (x86)/OpenBUGS/OpenBUGS323/",
working.directory = NULL, over.relax = FALSE, seed=NULL)
# a txt file with the model, named model.bivRE must be included in
the working directory
# make sure the path file for OpenBUGS is correctly specified

print(sim)
plot(sim)
```

The following code can be used to fit the model of Equation (2) in the main paper, in a frequentist setting:

```r
library(lme4)
y =c( 5, 5, 6, 0, 0, 2, 2, 0, 1, 2, 3,1, 1, 1, 3, 2, 2, 4, 3, 3, 0,12, 0, 8, 8, 2, 2, 0, 0,16, 3, 0, 0, 3,0,0, 2, 3, 6, 3,12,21, 1, 0,23, 0)
n =c(63,23,44,128,99,85,30,74,47,24,46,45,32,50, 25,60,20, 82,84,53,16, 8,34,97,63,22,35,60,77,91,31,76,54,24,44,35,36,50, 25,60,20,65,85,52,16, 9,36,95)
stud=c(1:24, 1:24)
treat=c(rep(0,24), rep(1,24))
treat.s=treat-0.5
gmFE <- glmer(cbind(y, n - y)~treat.s+(1|stud), family = binomial(link="logit"))  ### fixed effects
gmFE
exp(-confint(gmFE,method="Wald"))

gmRE <- glmer(cbind(y, n - y)~treat.s+(1|stud)+(1|stud:treat.s), family = binomial(link="logit"))  ### random effects
gmRE
exp(-confint(gmRE,method="Wald"))
```
References


