A robust Bayesian model to quantify and adjust for study quality and conflict of interest in meta-analyses

Matthias C. M. Troffaes [1] **Lorenzo Casini** [2] **Jürgen Landes** [3] **Ullrika Sahlin** [4]

[1]Department of Mathematical Sciences, Durham University, United Kingdom
[2]Department of Philosophy, University of Bologna, Italy
[3]Munich Center for Mathematical Philosophy and Open Science Center, LMU Munich, Germany
[4]Centre for Environmental and Climate Science, Lund University, Sweden

Abstract

Meta-analyses are vital for synthesizing evidence in medical research, but conflicts of interest can introduce research bias, undermining the reliability of the synthesized findings. This paper proposes a new robust Bayesian meta-analysis model. The model inflates uncertainty of low-quality studies and incorporates a bias term for studies subject to conflicts of interest. Using a random-effects model and sensitivity analysis with bounded probabilities, the model enables robust adjustments for conflicts of interest in meta-analytic contexts. A case study on antidepressant trials illustrates the potential application of the model.

Keywords. conflict of interest, meta-analysis, sensitivity analysis

1. Introduction

Medical studies investigate the causal effects of medical interventions: how quickly does a drug cure a disease? And which adverse reactions does it cause? Randomised controlled trials are widely agreed to be the "gold standard" for quantifying a drug's intended effects [48]. Patients taking part in such trials are randomly assigned to one of two groups, namely the treatment group, where they are administered the drug under study, or the control group, where they are given a placebo or a comparator treatment. Crucially, while the trial is in progress, neither patients nor their physicians are aware of the outcome of the allocation procedure.

Unfortunately, we cannot simply rely on the reported effect size measured in a single trial. Indeed, different trials investigating the same drug can arrive at different conclusions [42]. There are a number of reasons for this. First, trials often only study few patients [7, 50, 55]. Second, trials may be subject to various risks of bias [46], thereby increasing variability and bias in reported effect sizes. In particular, a large majority of trials are subject to conflicts of interest [47]. Such studies may, although they need not, bias truth-directed inquiry toward outcomes that conform with incentive structures [29, 36, 40]. Randomized controlled trials are expensive [53, 54] and thus often paid for by pharmaceutical companies. Physicians and academics paid by such companies to conduct and analyse trials might, although unethical, have strong incentives to make drugs look beneficial.

Small sample size (few patients per trial) causing a potentially large random error is ideally dealt with by increasing the sample size. Unfortunately, we cannot always conduct trials with more patients, due to the high costs involved. The next best thing is thus to aggregate multiple trials and thereby increase the number of patients. This is called *meta-analysis* [4, 17, 22]. Typically one uses the standard error of each trial to weigh its effects size, so that large studies carry more weight than small studies when aggregating the results. Still, this does not by itself address risks of bias.

Meta-analysis is an active field of research. The latest edition of the *Cochrane Handbook for Systematic Reviews of Interventions* [28] recommends taking note of potential study-specific risks of bias, including conflicts of interest, before carrying out a meta-analysis [6]. Nevertheless, it makes no proposal on how to quantify and *adjust* for their influence. Indeed, if some trials have conflict of interest, a meta-analysis may incorrectly reinforce the opinion that the effect size is statistically significant or even large [16]. More specifically, the handbook recognizes the advantage of Bayesian methods as they are able to adjust a meta-analysis for external evidence [25, §8.8.4.2, p. 210, §16.8.1, p. 518] [14, Box 10.13.a]. Moreover, it recognizes sensitivity analysis as a valuable tool when entertaining Bayesian methods [6, §7.6.2(4)]. Nevertheless, the handbook also states that Bayesian approaches to adjust *for risks of bias* are currently not developed enough to be recommended [6, §7.6.2]. Consequently, their use for this purpose is not encouraged, due to the strong assumptions they require, limited expertise, and the diverse intended audiences [6, §7.6.2(4)].

We hold the view that all available evidence—quality

and conflict of interest included—*ought to* be incorporated into the meta-analysis model and adjusted for [8, §3] [21]. To this end, we develop a Bayesian model in combination with sensitivity analysis, aiming to alleviate Cochrane's worries on the viability of Bayesian adjustment for bias. In particular, our main research questions are as follows:

1. Can we improve meta-analysis through careful modelling and prior elicitation of potential sources of bias, including quality and conflict of interest?
2. Can bounded probability help ensure that findings are robust especially when prior judgement is difficult, alleviating concerns on prior assumptions biasing the meta-analysis?

Although Bayesian methods are increasingly used especially when the number of studies is small [14, §10.13], almost all meta-analyses in the literature use frequentist methods (i.e. confidence intervals). Indeed, many researchers "regard it as controversial to combine objective trial data with subjective opinion" [14, §10.13]. Nevertheless, there exist several Bayesian proposals on how to adjust for bias in meta-analysis [12, 13, 32, 44, 57, 62]. However, there is no attempt to date to use knowledge about conflict of interest to correct the estimates produced by meta-analysis. This paper is a first step in trying to fill this gap in the literature.

We build on recent work [10, 44]. Casini and Landes [10] propose a Bayesian model for representing conflict of interest as influencing the observed effect *of a single study* both directly (by inflating the effect size) and indirectly, by affecting (e.g. improving) study quality, which in turn affects observed effects. Raices Cruz, Troffaes, Lindström, and Sahlin [44] study a robust Bayesian meta-analysis model for a binomial outcome that adjusts per-study error using quality scores based on the risk of bias table. They then propose bounding these scores to obtain a robust estimate of the aggregated effect.

Here, we move beyond these results by developing a robust Bayesian meta-analysis model that quantifies the influence of each study's quality and conflict of interest on the aggregated effect size. Robust Bayesian methods [3] combine Bayesian methods [14, §10.13], which rely on prior information in addition to the data, and sensitivity analysis [14, §10.14]. Whilst our model informs its priors by meta-epidemiological data, we use bounded probability (via sensitivity analysis informed by bounds on hyperparameters) to handle weak prior information where no single distribution can be reasonably justified. In particular, our sensitivity analysis focuses on conflict of interest, as its impact is extremely difficult to quantify precisely. This can help the researcher decide whether conflict of interest demands adjustment.

To demonstrate our approach, we apply our model on an existing meta-analysis on antidepressants. This choice is guided by two main considerations. First, the pharmaceutical industry of antidepressants is one of the most flourishing in medical research because of societal and financial pressures towards over-medicalisation [24, 51], increasing potential conflicts of interest [16]. Second, we may draw on evidence specific to conflict of interest in clinical trials of antidepressants [61]. In particular, we shall use the example of fluoxetine versus tricyclics [38, Analysis 2.1, §1.2.1, pp. 263-264], aggregating 18 trials, to illustrate our proposal. Since fluoxetine is one of the oldest selective serotonin reuptake inhibitors on the market, fluoxetine research has been widely investigated and systematically reviewed [38]. In our specific example, we found that conflict of interest had only a minor impact on the results of the meta-analysis, under a wide range of possible prior beliefs about what that impact might have been. This is helpful, because it strengthens the conclusions of the analysis.

The rest of the paper is organised as follows: Section 2 discusses the main quantities of interest, Section 3 introduces our Bayesian meta-analysis model, and Section 4 discusses elicitation of the model's hyper-parameters. Section 5 applies it to an existing meta-analysis comparing fluoxetine and tricyclics. Section 6 discusses limitations of our proposal. Section 7 concludes.

2. Quantities of Interest

For ease of reference, Table 1 summarizes notation. Observed data are denoted by capital letters (e.g. N_{ij}), unknown parameters by lower case Greek letters (e.g. δ_i), and hyper-parameters (to be elicited) by lower case Greek letters with a prime (e.g. μ'_δ).

2.1. Effect sizes. In this paper, we consider a random effects model for meta-analysis [5]. The standard random effects model has the following form:

$$\delta_i \sim \mathrm{N}(\mu_\delta, \sigma_\delta) \tag{1}$$

where μ_δ is the true overall mean effect size (denoted by μ in [5, Eq. (3)]), σ_δ is the standard deviation of the effect size between studies (denoted by τ in [5, Fig. 4]), and δ_i is the true study-specific mean effect size (denoted by θ_i in [5, p. 100]). The main purpose of meta-analysis is to quantify our uncertainty about μ_δ using the outcome data from individual studies. In this paper, we add two additional quantities to improve our uncertainty quantification of the overall effect size μ_δ, namely *study quality* and *conflict of interest*.

2.2. Study quality. Study quality is typically based on multiple considerations. The current standard for quality evaluation is the *Grading of Recommendations Assessment, Development and Evaluation* (GRADE) tool [49]. When used to evaluate individual studies, GRADE mainly uses considerations of indirectness (available evidence not directly answering the clinical question of interest), imprecision (small sample size), and risk of bias

K	number of studies
i	study/trial
j	trial group (1=control, 2=treatment)
β_{ij}	per-study per-group mean response
σ_i	per-study std. dev. of outcomes
δ_i	per-study effect size
μ_δ	overall effect size
σ_δ	between-study std. dev. of effect size
γ	bias parameter for effect size
N_{i+}	number of positive risks of bias
N_{i-}	number of negative risks of bias
Q_i	quality score
C_i	conflict of interest score
N_{ij}	sample size
X_{ijk}	individual response
$\bar{X}_{ij}$	sample mean
S_{ij}	sample standard deviation
μ'_δ	prior mean of μ_δ
σ'_δ	prior std. dev. of μ_δ
μ'_γ	prior mean of γ
σ'_γ	prior std. dev. of γ
α'_σ	prior median of σ_i
λ'_σ	prior shape of σ_i
α'_δ	prior median of σ_δ
λ'_δ	prior shape of σ_δ
N	normal distribution (mean, std. dev.)
Gam	gamma distribution (shape, rate)
χ^2	chi squared distribution (d.f.)

Table 1. *Summary of notation.*

(inadequate allocation concealment, inadequate blinding, incomplete outcome data, selective reporting). Such considerations help evaluators form a qualitative judgement, which is then reported in systematic reviews.

However, since per-study GRADE scores are not available for the data that we use [38], here we operationalise study quality by means of the risk of bias table [27] which is reported for each study [38, Fig. 4, p. 41]. Following [44, §2.2], we translate the risk of bias table into a per-study numerical value $Q_i \in (0, 1]$. This value represents the ratio between the standard deviation σ_δ of the effect size of a study with best possible quality, and the standard deviation of the effect size δ_i of the ith study. Equivalently, with regards to δ_i, we inflate the standard deviation σ_δ by a factor $1/Q_i$. In this way, we model the influence of the quality Q_i of a specific study i on its effect size.

For the inference, what matters is that the Q_i are ordered so that the relative quality of the studies is reflected. As exact Q_i values are difficult to justify, a sensitivity analysis is recommended [44]. However since we mainly want to focus on the impact of conflict of interest, for simplicity, here we will keep the Q_i fixed, and defer a full sensitivity analysis over the Q_i to future work.

To translate the risk of bias table into a quality number, we aim to give the best study a quality of 1, and the worst study a quality of 0.2 (thus cautiously inflating its standard deviation by a factor 5). One way to achieve this goes as follows:

$$D_i := N_{i+} - N_{i-} \tag{2}$$

$$D_* := \min_{i=1}^{K} D_i \tag{3}$$

$$D^* := \max_{i=1}^{K} D_i \tag{4}$$

$$Q_i := 0.2 + 0.8 \frac{D_i - D_*}{D^* - D_*} \tag{5}$$

where N_{i+} is the number of positive points (i.e. increasing quality) in the table, and N_{i-} is the number of negative points (i.e. decreasing quality). From Table 2, D_i is minimal for $i = 16$ (one positive and three negatives, so $D_* = 1 - 3 = -2$) and maximal for $i = 12$ (two positives and no negatives, so $D^* = 2 - 0 = 2$). Almost all studies we consider have a quality of 0.4 to 0.6, as can be seen from Table 2. Note that we excluded 'other bias' for our quality assessment since this corresponds to the presence of industry funding, which we will represent separately as conflict of interest.

2.3. Conflict of interest. In the broadest sense, conflict of interest refers to "anything that may influence professional judgment", be that a financial interest by a pharmaceutical company, a personal interest, an academic interest, a political interest, or otherwise [31]. It has been noted that variables such as affiliation, impact factor, country, etc. are not statistically significant once industry funding and study quality are controlled for [19]. Recent literature reduces conflict of interest to presence of situations where one or more authors are either company employees of the industry or received any support from the industry for any of their work [16]. Although conflicts of interest are now widely reported, not all are reported [56]. Most findings relating conflict of interest and reported outcomes concern *reported* conflict of interest. We follow this tradition.

There is strong evidence that industry funding promotes studies aligning with the funders' interests [1, 18, 20, 31, 37]. When it comes to explaining this difference, though, and more specifically whether it is due to better study quality or not, opinions are divided. Some suggest that industry-sponsored studies are of better quality than other trials [45] or that there is no significant difference between the two [35, 39]. Still, others observe that industry-sponsored studies tend to selectively choose outcome data [41] or comparators [33], and to promote "p-hacking", selective outcome reporting, non-publication of negative trials, ghostwriting, and other practices that

i	study	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Q_i	R_8	C_i
1	Altamura 1989	?	?	?	?	?	?	?	0.6	?	0
2	Chouinard 1985	?	?	?	?	?	–	?	0.4	?	0
3	Demyttenaere 1998	?	?	?	?	?	–	?	0.4	–	1
4	De Ronchi 1998	?	?	?	?	?	–	?	0.4	?	0
5	Fawcett 1989	?	?	?	?	?	?	?	0.6	–	1
6	Feighner 1985b	?	?	?	?	?	?	?	0.6	?	0
7	Judd 1993	?	?	?	?	?	?	?	0.6	–	1
8	Keegan 1991	?	?	?	?	?	?	?	0.6	–	1
9	Kerkhofs 1990	?	?	?	?	?	–	?	0.4	–	1
10	Laakman 1988	?	?	?	?	?	–	?	0.4	?	0
11	Marchesi 1998	?	?	?	?	?	?	?	0.6	–	1
12	OKeane 1992	?	?	?	?	?	+	+	1.0	?	0
13	OntiverosSanchez 1998	?	?	?	?	?	?	?	0.6	–	1
14	Peters 1990	?	?	?	?	?	?	?	0.6	?	0
15	Preskorn 1991	?	?	?	?	?	?	?	0.6	–	1
16	Suleman 1997	?	?	–	–	–	+	?	0.2	–	0*
17	Versiani 1999	?	?	?	?	?	?	?	0.6	–	1
18	Yu 1997	?	?	?	?	?	?	?	0.6	?	0

Table 2. *Risk of bias assessments from [38, Fig. 4, p. 41], along with our quality and conflict of interest scores. Here, '+' means low risk, '–' means high risk, and '?' means unclear risk. R_1 to R_7 correspond to distinct types of bias (see [38] for details), and R_8 corresponds to 'other bias' which indicated industry funding. Note that C_8 for Suleman 1997 (starred in the table) was set to 0, as the study was sponsored by producers of both fluoxetine and tricyclics, whilst all other studies with industry funding were sponsored by producers of fluoxetine only.*

yield more positive outcomes [34]. Such practices are not tracked by available quality indicators, such as the risk-of-bias tool [46]. As a result, the correlation between conflict of interest and a difference in estimates is robust, even after controlling for a number of potential explanatory factors [20, 31]. We follow existing suggestions [37] to the point that, while industry sponsorship in itself is not a bias, it should be treated as a proxy for an *independent* bias-inducing mechanism. Technically, this is achieved by adding a bias term of the form γC_i onto μ_δ, where γ is a new parameter that measures any resulting bias on the effect size. This parameter needs to be given a prior distribution, but can also be learned from the data.

Our example compares fluoxetine against tricyclics. All studies in our data report industry funding from producers of fluoxetine, have unclear funding source, or (in just one case) have funding from industries of both fluoxetine and tricyclics. Accordingly, we model conflict of interest by means of a binary variable $C_i \in \{0, 1\}$, where 1 denotes 'funding from fluoxetine producer' (so clear reason for potential bias towards fluoxetine) and 0 denotes 'funding unclear or from both producers' (so no clear reason for potential bias in either direction); see Table 2. In general, one could have $C_i \in \{-1, 0, 1\}$ or even $C_i \in \mathbb{R}$ if needed.

3. Likelihood and Priors

For each study i and group j, let N_{ij} denote the sample size. Assume that each individual response X_{ijk} is independently sampled from a normal distribution with mean β_{ij} and standard deviation σ_i (so we have the same standard deviation across groups but allow different standard deviations across studies):

$$X_{ijk} \sim \mathrm{N}(\beta_{ij}, \sigma_i) \tag{6}$$

for $k \in \{1, \ldots, N_{ij}\}$. Recall $j \in \{1, 2\}$, and let

$$\delta_i := \frac{\beta_{i2} - \beta_{i1}}{\sigma_i} \tag{7}$$

denote the *standardized mean difference*. This is a commonly used measure for the effect size when comparing the responses of two groups [26], and normalizes the results across studies when not all studies use the same response scale, so effect sizes can be meaningfully compared across studies [11, §2.2]. For the remainder of this paper, unless stated otherwise, the term *effect size* will refer to the standardized mean difference δ_i.

Let $\bar{X}_{ij}$ and S_{ij} denote the sample mean and sample standard deviation of the responses of study i and group $j \in \{1, 2\}$. We use the following model (see Figure 1):

$$\bar{X}_{i2} - \bar{X}_{i1} \sim \mathrm{N}\left(\delta_i \sigma_i, \sigma_i \sqrt{\frac{1}{N_{i1}} + \frac{1}{N_{i2}}}\right) \tag{8}$$

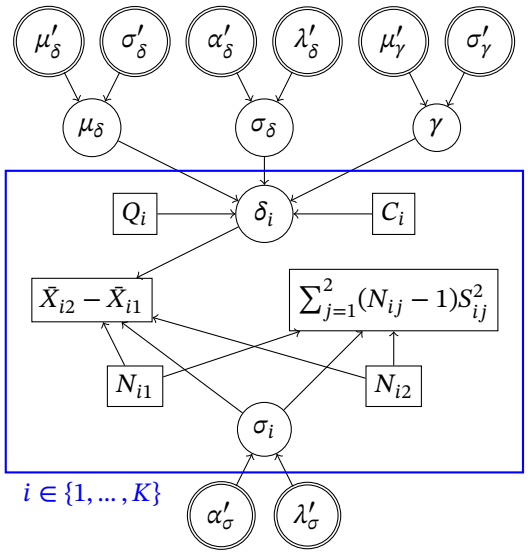

Figure 1. *Graphical meta-analysis model for quality and conflict of interest adjustment. The rectangle contains the submodel for each study in the meta-analysis. Rectangular nodes represent observed data, circular nodes represent parameters, and double cirular nodes represent hyperparameters.*

$$\sum_{j=1}^{2}(N_{ij}-1)S_{ij}^2 \sim \text{Gam}\left(\frac{\sum_{j=1}^{2}(N_{ij}-1)}{2}, \frac{1}{2\sigma_i^2}\right) \tag{9}$$

$$\delta_i \sim \text{N}(\mu_\delta + \gamma C_i, \sigma_\delta/Q_i) \tag{10}$$

with these prior distributions:

$$\log(\sigma_i) \sim \text{N}(\log(\alpha'_\sigma), \lambda'_\sigma) \tag{11}$$

$$\mu_\delta \sim \text{N}(\mu'_\delta, \sigma'_\delta) \tag{12}$$

$$\log(\sigma_\delta) \sim \text{N}(\log(\alpha'_\delta), \lambda'_\delta) \tag{13}$$

$$\gamma \sim \text{N}(\mu'_\gamma, \sigma'_\gamma) \tag{14}$$

Equations (8) and (9) follow from the normal sampling assumption. Indeed, note that

$$\bar{X}_{ij} \sim \text{N}\left(\beta_{ij}, \frac{\sigma_i}{\sqrt{N_{ij}}}\right) \tag{15}$$

$$(N_{ij}-1)S_{ij}^2/\sigma_i^2 \sim \chi^2(N_{ij}-1) \tag{16}$$

Equation (8) then follows from the well-known properties of the normal distribution, along with the definition of δ_i, and Equation (9) follows from the fact that $\chi^2(n) \sim \text{Gam}(n/2, n/2)$ and well-known properties of the gamma distribution (for instance, see [15, Ex. 8.3.1]). Note that, by the central limit theorem, Equations (8) and (9) will still hold approximately even if the responses are not precisely normally distributed, provided the N_{ij} are not too small.

Because we are not primarily interested in β_{ij}, we have set up the model not to depend on β_{ij}. This removes the need to specify priors for β_{ij}, resulting in fewer hyperparameters, a smaller model, and faster simulations.

Equation (10) enhances the standard random effects model (see Equation (1)) to handle study quality and conflict of interest. As discussed in Section 2, we add a bias term to μ_δ for studies that have conflict of interest (i.e. where $C_i \neq 0$), and we inflate the standard deviation σ_δ for studies that have lower quality (i.e. where $Q_i < 1$).

The above model is based on [44, §2], with some differences:

- Whereas [44, §2] handles binomial data using the log odds ratio as effect size, our model handles continuous data using the standardized mean difference as effect size.
- The quality Q_i is used to scale the standard deviation rather than the variance. This is just a matter of scale, but using the standard deviation is perhaps more natural for elicitation.
- An extra bias term γC_i is included to model the influence of conflict of interest.
- We use a log-normal prior to model uncertainty in the standard deviation σ_δ of the effect size across studies, instead of a non-informative inverse gamma prior on σ_δ^2. This allows us to easily elicit a weakly informative prior on σ_δ. Besides improving posterior inferences, weakly informative priors on variance parameters are known to improve stability in Monte Carlo simulations [52].

4. Hyper-parameter elicitation

To use the model, we need to elicit (or at least bound) various hyperparameters. The case study in Section 5 uses data from [38], and we are fortunate that [38] contains a very large number of other meta-analyses comprising similar drug trials, all concerning fluoxetine. Therefore, where possible, we have based our elicitation by inspecting these other meta-analyses to ensure priors cover reasonable ranges.

4.1. Prior median and shape for σ_i: α'_σ and λ'_σ. Recall σ_i is the standard deviation of the study-specific responses. Learning of σ_i is mostly driven by Equation (9). If $N_{i1} + N_{i2}$ is reasonably large (say at least 20), which it typically will be, the choice of the prior will only have a limited effect on the posterior, and thus a weakly informative prior should suffice. This is also why we describe all σ_i by the same prior regardless of the study i, thereby simplifying model elicitation.

The α'_σ hyper-parameter is equal to the prior median of σ_i. The λ'_σ hyper-parameter can be used to control the quantiles of the prior; see Table 3. For the data [38] that we will use in our case study, typical trials have standard deviations of responses between 5 and 10, so $\alpha'_\sigma = 7.5$

λ'	$P(\alpha' q_* \leq \sigma \leq \alpha' q^*)$	q_*	q^*
1	0.50	0.509	1.963
	0.95	0.141	7.099
2	0.50	0.260	3.853
	0.95	0.020	50.397
3	0.50	0.132	7.565
	0.95	0.003	357.771

Table 3. *25% and 2.5% quantiles q_* along with 75% and 97.5% quantiles q^* of the log-normal distribution, parametrized as $\log(\sigma) \sim \mathrm{N}(\log(\alpha'), \lambda')$ for $\lambda' \in \{1, 2, 3\}$ where α' represents the median.*

with $\lambda'_\sigma = 1$ seems reasonable. For a really weak prior, one could settle for $\lambda'_\sigma = 3$, at the expense of a slightly less informative posterior.

4.2. Prior mean and standard deviation of μ_δ: μ'_δ and σ'_δ. Recall μ_δ is the overall effect size. This is different from the study specific effect size δ_i which has additional variance and bias. If data is available from similar representative studies or meta-analyses, we can pool together the effect sizes of these studies to make an informed choice about μ'_δ and σ'_δ directly. On the cautionary side, we may take $\mu'_\delta = 0$ if we assume the effect can be positive or negative with equal probability. The standard normal quantiles can be used to elicit σ'_δ. For instance, if we believe that $P(-\mu^* \leq \mu_\delta \leq \mu^*) = 0.95$ for some given μ^*, and assuming $\mu'_\delta = 0$, we may set $\sigma'_\delta = \mu^*/1.96$. Other meta-analyses in [38] have overall effect sizes well within -0.5 and 0.5, so $\mu'_\delta = 0$ with $\sigma'_\delta = 0.25$ seems reasonable. For a really weak prior, one could settle for $\sigma'_\delta = 1$, at the expense of a slightly less informative posterior. Indeed, for effect sizes are measured in standardized mean differences, an effect size of more than 1 is generally considered extremely strong, and is extremely rare for the domain that we are studying.

However, if the meta-analysis concerns few studies, μ'_δ and σ'_δ will have a major impact on the inference. Therefore, in this case, instead of fixing their values, we propose bounding μ'_δ across a range of reasonable values.

4.3. Prior median and shape for σ_δ: α'_δ and λ'_δ. Recall σ_δ is the standard deviation of the study-specific effect size across studies. In the literature, the square of this parameter is sometimes referred to as the 'between-study variance' [58] and is denoted as τ^2. Estimates for it are commonly reported in meta-analyses [38]. The α'_δ hyper-parameter is equal to the prior median of σ_δ. The estimated between-study variances $\hat{\tau}^2$ across similar meta-analyses can be used to inform α'_δ. The λ'_δ hyper-parameter can be used to control the quantiles of the prior; again see Table 3. Other meta-analyses in [38], have a typical $\hat{\tau}^2$ between 0.1^2 and 0.7^2, so $\alpha'_\delta = 0.3$ with $\lambda'_\delta = 1$ seems reasonable.

Here too, if the meta-analysis concerns few studies, α'_δ and λ'_δ will have a major impact on the inference. Bounding α'_δ across a range of reasonable values can address this concern.

4.4. Prior mean and standard deviation of γ: μ'_γ and σ'_γ. Recall γ is the bias on the effect size for a study with one unit of conflict of interest. Little information may be available to inform these hyper-parameters. Large scale meta-analyses might give an informed view about potential effects of conflict of interest on effect sizes in specific research domains.

Two scenarios are of interest. First, we may wish to learn about whether conflict of interest has an impact. In that case, it makes sense to set $\mu'_\gamma = 0$ and to set σ'_γ reflecting a wider range of potential biases on the effect size. Second, we may wish to adjust for conflict of interest. In that case, we set μ'_γ equal to the expected bias adjustment for a study with one unit of conflict of interest, and σ'_γ then reflects our uncertainty around that expectation. Various sources of meta-epidemiological evidence may be used to inform μ'_γ and σ'_γ [1, 16, 18, 20, 30, 31, 35, 37, 43, 61] depending on the situation.

To illustrate our continuous outcome model, we use a study by Kjaergard and Als-Nielsen [31]. They estimated differences induced by conflict of interest in trials investigating (continuous) effects of interventions for a variety of conditions in the reference class of studies published in BMJ from January 1997 to June 2001. After controlling for potential confounders, they found that reported conflict of interest resulted in a bias of 0.48 (with standard error 0.13) towards positive outcomes on a 6 point scale. To translate their findings into our framework, note that an effect size of -0.8 is typically considered as a large effect and would be scored 1, whilst $+0.8$ would be scored 6. Going from 1 to 6 on the scale thus roughly corresponds to moving from -0.8 to 0.8 on the effect size. We may thus set

$$\mu'_\gamma = 0.48 \times 1.6/5 \simeq 0.15. \tag{17}$$

Once again, if the meta-analysis contains only few studies to discriminate between situations with and without conflict of interest, μ'_γ and σ'_γ will have a major impact on the inference. Therefore, if we are not clearly informed about γ a priori, it is critical bound μ'_γ across a range of values, rather than settling on a specific value. As we have considerable uncertainty around μ'_γ, in our meta-analysis, we will opt for a weakly informative set of priors. Specifically, we will set $\sigma'_\gamma = 0.3$, and bound $\mu'_\gamma \in [0, 0.3]$, to cover the full range of reasonably possible scenarios.

5. Case Study

To demonstrate our model on a real-world example, we performed a meta-analysis using our model, using

data from a meta-analysis by Magni et al. [38, Analysis 2.1, §1.2.1, pp. 263-264], which compares fluoxetine against tricyclics. Table 4 lists the data used. Note that we omitted one trial that reported an S_{ij} equal to zero, following [38] who also excluded this trial from their meta-analysis. For Bayesian computations, we implemented the model in Stan [9]. The results of our analysis is given in Table 5.

The first column, labelled 'frequentist', replicates the analysis from [38]. Unfortunately, Magni et al. [38] do not report the exact method used. To estimate δ_i, we calculated Hedges' g [23] (which is Cohen's d [11] with a bias correction) under the equal variance assumption, with error based on [60, Eq. (26)] assuming normal quantiles. We estimated σ_δ (often called τ in the literature) using the DerSimonian and Laird method [58, §3.1]. To estimate μ_δ, we used the Wald-type normal distribution method [59, §3.1.1(i)]. This approach gave us values nearly identical to the ones in [38].

The second column, labelled 'Bayes (no m.a.)' gives our Bayesian model a sanity check. We ran our model against the data but without the meta-analysis part of the model (whence, 'no m.a.'). Specifically, the model included only the likelihood given by Equations (8) and (9), along with an extremely vague priors for δ_i and σ_i:

$$\delta_i \sim \mathrm{N}(0, 100) \qquad \log \sigma_i \sim \mathrm{N}(\log(1), 3) \tag{18}$$

With a flat prior, one expects the credible intervals to be extremely close to the confidence intervals. And indeed, the first two columns are extremely close.

The third column, labelled 'Bayes (no adj.)', gives the outcomes of the Bayesian model without adjusting for quality and conflict of interest, i.e. simply with

$$\delta_i \sim \mathrm{N}(\mu_\delta, \sigma_\delta) \tag{19}$$

and prior distributions for σ_i, μ_δ and σ_δ as given in Section 4. It shows what a standard Bayesian meta-analysis might result in, and gives us a baseline to assess the impact of modelling quality and conflict of interest later. Interestingly, the Bayesian estimates for μ_δ and σ_δ are nearly identical to the frequentist estimates. Besides providing a baseline, the similarity to the frequentist estimates confirms that our Bayesian meta-analysis model is running well. The δ_i estimates here are very different from the ones in the previous columns: this is because each δ_i is estimated from all data jointly. To see how this makes sense, for example, note that the estimated values of μ_δ and σ_δ imply that the observed value for δ_1 (see first row) is nearly impossible. The Bayesian model has therefore adjusted its estimate for δ_1 towards more reasonable values. One is normally not interested in estimating the δ_i in a Bayesian meta-analysis, but it is nevertheless interesting and insightful to compare and understand the difference, so we point it out here.

The final three columns present a sensitivity analysis of our full model, for $\mu'_\gamma \in \{0, 0.3, -0.3\}$. Compared to 'Bayes (no adj.)', besides a small adjustment in the estimate for μ_δ (due to the bias adjustment term γC_i in the model), we also note that the credible intervals have widened by roughly 30%. This is expected: (i) we have one more parameter in the model (namely, γ) which results in an increased uncertainty in the estimate, and (ii) we have incorporated the quality of each trial, which has increased the variability between the δ_i estimates, indirectly also increasing the uncertainty in μ_δ.

Remember, in Section 4, we found that $\mu'_\gamma \in [0, 0.3]$ seemed like a reasonable bound. Since the estimate for μ_γ is monotone in μ'_γ, it suffices to look at $\mu'_\gamma \in \{0, 0.3\}$, i.e. the third last and second last columns. This leads us to the following posterior lower and upper expectations:

$$\underline{E}(\mu_\delta) = 0.05 \qquad \overline{E}(\mu_\delta) = 0.09 \tag{20}$$

$$\underline{E}(\gamma) = -0.03 \qquad \overline{E}(\gamma) = 0.05 \tag{21}$$

We also get these bounded credible intervals:

$$\underline{P}(\mu_\delta \in [-0.20, 0.32]) \geq 0.95 \tag{22}$$

$$\underline{P}(\gamma \in [-0.34, 0.37]) \geq 0.95 \tag{23}$$

For this specific meta-analysis, we conclude that the overall effect size is small, and any potential conflict of interest has only limited impact. The main value of this robust Bayesian analysis is that we know this conclusion holds, even against rather weak prior information about the potential impact of conflict of interest.

Out of interest, the final column shows what happens for $\mu'_\gamma = -0.3$. As expected, it has roughly the same effect as $\mu'_\gamma = 0.3$ but in the opposite direction.

The fact that, in this case study, our Bayesian meta-analysis aligns with the existing frequentist meta-analysis is purely coincidental. Had individual studies been different, our model could have led to a revision, rather than a corroboration, of the existing results. Due to space constraints, unfortunately, we did not provide an example to this effect.

6. Limitations

Whilst our analysis improves meta-analysis in light of study quality and conflict of interest, it does have a number of shortcomings.

To begin with, the manner in which different risks of bias affect between-study variability is hard to quantify. In our example, we simply related quality (which scales between-study variability) to the number of negative and positive entries in the risk of bias table, using a cautious (the worst quality study inflated the standard deviation by a factor 5 compared to the best quality study) but otherwise rather arbitrary mapping (see Equation (5)). This is a particularly weak point of our analysis, although

i	study	N_{i1}	M_{i1}	S_{i1}	N_{i2}	M_{i2}	S_{i2}	Q_i	C_i
1	Altamura 1989	11	13.0	2.0	11	10.0	1.0	0.6	0
2	Chouinard 1985	20	15.5	6.2	24	10.6	5.5	0.4	0
3	Demyttenaere 1998	35	9.9	6.3	31	7.2	4.5	0.4	1
4	De Ronchi 1998	32	14.2	8.3	33	13.9	9.4	0.4	0
5	Fawcett 1989	19	12.8	6.5	19	14.6	7.9	0.6	1
6	Feighner 1985b	22	15.0	6.2	22	19.0	5.5	0.6	0
7	Judd 1993	23	9.6	6.2	23	11.6	6.0	0.6	1
8	Keegan 1991	19	7.5	2.5	18	7.0	3.0	0.6	1
9	Kerkhofs 1990	9	8.4	6.2	10	9.8	4.6	0.4	1
10	Laakman 1988	22	8.5	8.0	28	7.0	5.2	0.4	0
11	Marchesi 1998	67	9.0	6.6	75	8.2	6.9	0.6	1
12	OKeane 1992	7	7.3	5.6	9	10.8	9.3	1.0	0
13	OntiverosSanchez 1998	21	7.8	6.2	21	5.8	5.5	0.6	1
14	Peters 1990	40	11.0	9.0	41	10.0	6.0	0.6	0
15	Preskorn 1991	29	13.7	7.8	31	15.6	6.1	0.6	1
16	Suleman 1997	15	7.2	2.5	15	7.0	2.6	0.2	0
17	Versiani 1999	77	10.5	8.9	79	8.7	7.7	0.6	1
18	Yu 1997	8	9.0	10.0	8	11.0	10.0	0.6	0

Table 4. *Trials used from Magni et al. [38, Analysis 2.1, §1.2.1, pp. 263-264]), along with our quality and conflict of interest scores. Trials with conflict of interest are highlighted.*

par.	frequentist	Bayes (no m.a.)	Bayes (no adj.)	Bayes ($\mu'_\gamma = 0$)	Bayes ($\mu'_\gamma = 0.3$)	Bayes ($\mu'_\gamma = -0.3$)
δ_1	1.83 [0.80, 2.85]	1.92 [0.88, 2.96]	0.45 [-0.03, 1.24]	0.42 [-0.06, 1.15]	0.39 [-0.10, 1.14]	0.45 [-0.04, 1.21]
δ_2	0.83 [0.21, 1.45]	0.85 [0.23, 1.47]	0.37 [-0.03, 0.91]	0.48 [-0.00, 1.05]	0.46 [-0.03, 1.04]	0.50 [0.01, 1.08]
δ_3	0.48 [-0.01, 0.97]	0.49 [-0.01, 0.98]	0.27 [-0.06, 0.69]	0.32 [-0.07, 0.77]	0.33 [-0.05, 0.78]	0.31 [-0.10, 0.76]
δ_4	0.03 [-0.45, 0.52]	0.03 [-0.45, 0.52]	0.07 [-0.29, 0.40]	0.06 [-0.35, 0.45]	0.05 [-0.35, 0.44]	0.07 [-0.34, 0.47]
δ_5	-0.24 [-0.88, 0.39]	-0.25 [-0.90, 0.39]	-0.03 [-0.49, 0.35]	-0.04 [-0.50, 0.34]	-0.02 [-0.47, 0.37]	-0.07 [-0.52, 0.31]
δ_6	-0.67 [-1.28, -0.06]	-0.69 [-1.30, -0.07]	-0.20 [-0.75, 0.20]	-0.18 [-0.72, 0.23]	-0.21 [-0.75, 0.22]	-0.16 [-0.71, 0.27]
δ_7	-0.32 [-0.90, 0.26]	-0.33 [-0.89, 0.24]	-0.07 [-0.53, 0.29]	-0.08 [-0.55, 0.29]	-0.06 [-0.51, 0.31]	-0.11 [-0.55, 0.26]
δ_8	0.18 [-0.47, 0.82]	0.18 [-0.46, 0.84]	0.12 [-0.28, 0.54]	0.10 [-0.30, 0.53]	0.13 [-0.28, 0.55]	0.08 [-0.33, 0.49]
δ_9	-0.25 [-1.15, 0.66]	-0.26 [-1.17, 0.66]	0.01 [-0.52, 0.47]	-0.05 [-0.69, 0.50]	-0.03 [-0.66, 0.51]	-0.08 [-0.69, 0.46]
δ_{10}	0.22 [-0.34, 0.78]	0.23 [-0.34, 0.78]	0.15 [-0.22, 0.55]	0.17 [-0.26, 0.61]	0.15 [-0.27, 0.60]	0.19 [-0.23, 0.63]
δ_{11}	0.12 [-0.21, 0.45]	0.12 [-0.21, 0.45]	0.11 [-0.16, 0.38]	0.10 [-0.18, 0.38]	0.11 [-0.16, 0.39]	0.08 [-0.19, 0.36]
δ_{12}	-0.42 [-1.42, 0.58]	-0.46 [-1.44, 0.52]	-0.02 [-0.61, 0.44]	0.04 [-0.37, 0.39]	0.01 [-0.40, 0.36]	0.08 [-0.33, 0.44]
δ_{13}	0.33 [-0.27, 0.94]	0.34 [-0.27, 0.96]	0.19 [-0.18, 0.61]	0.16 [-0.22, 0.59]	0.18 [-0.19, 0.61]	0.14 [-0.24, 0.56]
δ_{14}	0.13 [-0.31, 0.57]	0.13 [-0.30, 0.57]	0.11 [-0.22, 0.44]	0.11 [-0.22, 0.45]	0.09 [-0.23, 0.42]	0.13 [-0.20, 0.46]
δ_{15}	-0.27 [-0.78, 0.24]	-0.27 [-0.79, 0.24]	-0.07 [-0.49, 0.27]	-0.08 [-0.50, 0.27]	-0.06 [-0.49, 0.29]	-0.10 [-0.51, 0.24]
δ_{16}	0.08 [-0.64, 0.79]	0.08 [-0.64, 0.80]	0.09 [-0.34, 0.52]	0.08 [-0.54, 0.70]	0.07 [-0.54, 0.70]	0.09 [-0.53, 0.70]
δ_{17}	0.22 [-0.10, 0.53]	0.22 [-0.10, 0.53]	0.17 [-0.08, 0.45]	0.16 [-0.11, 0.44]	0.17 [-0.09, 0.44]	0.15 [-0.12, 0.43]
δ_{18}	-0.19 [-1.17, 0.79]	-0.21 [-1.20, 0.80]	0.03 [-0.50, 0.50]	0.03 [-0.50, 0.50]	0.00 [-0.53, 0.46]	0.07 [-0.47, 0.53]
μ_δ	0.10 [-0.09, 0.29]		0.09 [-0.09, 0.27]	0.09 [-0.16, 0.32]	0.05 [-0.20, 0.28]	0.13 [-0.11, 0.37]
σ_δ	0.27		0.25 [0.06, 0.52]	0.14 [0.04, 0.29]	0.14 [0.04, 0.29]	0.14 [0.04, 0.29]
γ				-0.03 [-0.34, 0.28]	0.05 [-0.25, 0.37]	-0.11 [-0.43, 0.19]

Table 5. *Effect sizes δ_i from each trial $i \in \{1, \ldots, K\}$, and overall effect size μ_δ from standard frequentist meta-analysis and our Bayesian meta-analysis, based on data reported in Magni et al. [38, Analysis 2.1, §1.2.1, pp. 263-264]). Frequentist values denote maximum likelihood estimate and 95% confidence interval. Bayesian estimates denote posterior expectation and 95% (quantile based) credible interval. Trials with conflict of interest are highlighted.*

it is obviously still better than not adjusting at all for quality (as in standard frequentist meta-analysis). One could treat quality as an uncertain parameter, that can be learned from the risk of bias table. In this way, we could assess much better how much each risk contributes to quality. This would need a very large amount of data, which was simply not available. In this case, a sensitivity analysis (as in [44]), using a set of quality scores, provides a potential solution. Some assessment is still needed to provide sensible bounds for this set, and it is not entirely clear how to inform such bounds.

Secondly, in our example we interpreted conflict of interest as a binary variable, $C_i = \{0, 1\}$. This may be too limited, especially if there are many studies in the meta-analysis with varying levels of conflict of interest. To make C_i a more graded variable, one could have first authors count more than other authors, or employment count more than support/honoraria, or undeclared conflict of interest count more than declared conflict of interest. One should also allow C_i take negative values, to represent the direction of bias. A more refined operationalisation of conflict of interest is left to future work.

Thirdly, in our example, conflict of interest did not feed into the quality score (see Table 2). However, there may be a relationship between conflict of interest and quality [40]. On the one hand, sponsored studies have greater financial means to *improve* the quality of study (e.g. longer study duration, better data analytics tools). On the other hand, sponsored studies are subject to financial incentives exerting biases on reported results, which may not be easily tracked by current bias assessment tools [46]. Incorporating conflict of interest into the quality score could help disentangle the "beneficial" effects of conflict of interest from its biasing effects.

Unfortunately, there is not as much literature on the relation between conflict of interest and quality as there is on the relation between conflict of interest and reported outcomes. Bekelman, Li, and Gross [2] find that industry sponsorship is associated with restrictions on publication and data sharing, which may contribute to publication bias; however they do not verify, let alone quantify, this dependence. Als-Nielsen, Chen, Gluud, and Kjaergard [1] find that industry-funded trials are less likely to use adequate methods for allocation concealment, which affects risk of bias. Lundh, Lexchin, Mintzes, Schroll, and Bero [37] further focus on the relation between conflict of interest and risk of bias but report conflicting findings. All in all, at present it is very difficult to quantify the influence of conflict of interest on quality. We leave the development of modelling tools that incorporate an explicit dependence of quality on conflict of interest as an avenue for future work.

Our main goal here was to introduce a new model for bias adjustment. As regards the determination of the hyper-parameters, we took a pragmatic approach based on [31, 38]. Ideally, one would strive to judiciously use as much as possible of the available meta-epidemiological evidence and expert opinions to inform our judgments [1, 16, 18, 20, 30, 31, 35, 37, 43, 61].

Finally, we only presented one application (due to space constraints), involving a moderate amount of studies. Sensitivity analysis will have a much larger impact for meta-analyses with fewer studies, especially with additional sensitivity analysis over μ'_δ, α'_δ and Q_i.

7. Conclusion

In this paper, we build on previous work [10, 44] to create a robust Bayesian model that can adjust meta-analyses for risks of bias in individual trials, focusing on quality and conflict of interest. Since the presence of conflict of interest is well known to be correlated with reported outcomes, the absence or presence of conflict of interest provides relevant information for the assessment of data. The assessment of the presence and strength of conflict of interest is a delicate matter requiring human judgement, which can be aided by drawing on the meta-epidemiological literature. Our robust Bayesian model allows us to rigorously study and explore consequences of conflict of interest assessments alleviating concerns about the sensitivity of the end result. Robust Bayesian analysis appears to us as an appropriate tool to carry out this vital task.

In this regard, our proposal aims to show how, despite widespread intuitions that systematic reviews ought to be conducted in a frequentist paradigm for reasons of objectivity [14, §10.13], not only are Bayesian methods instrumental to quantitatively taking into account meta-epidemiological evidence or expert opinions, but—in so doing—they are actually *conducive to objectivity*, since only by taking all of the available evidence into account may one form a more strongly justified credence about true effect sizes. We conclude:

1. Current practice in meta-analysis can be improved through careful modelling and prior elicitation of potential sources of bias, including quality and conflict of interest.
2. The use of bounded probability can help ensure that findings are robust especially when prior judgement is difficult, alleviating concerns of bias due to prior assumptions in the meta-analysis, and thereby strengthening conclusions.

Additional Author Information

Acknowledgements. Jürgen Landes gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 528031869.

Competing interests. All authors declare that they have no conflict of interest.

References

[1] Bodil Als-Nielsen, Wendong Chen, Christian Gluud, and Lise L. Kjaergard. "Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events?" In: *JAMA* 290.7 (Aug. 2003), pp. 921–928. DOI: 10.1001/jama.290.7.921.

[2] Justin E Bekelman, Yan Li, and Cary P Gross. "Scope and impact of financial conflicts of interest in biomedical research: a systematic review." In: *JAMA* 289.4 (Jan. 2003), pp. 454–465. DOI: 10.1001/jama.289.4.454.

[3] James O. Berger. "Robust Bayesian analysis: sensitivity to the prior". In: *Journal of Statistical Planning and Inference* 25.3 (1990), pp. 303–328. DOI: 10.1016/0378-3758(90)90079-A.

[4] Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hanna R. Rothstein. *Introduction to Meta-Analysis*. Chichester: Wiley, 2009. DOI: 10.1002/9780470743386.

[5] Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. "A basic introduction to fixed-effect and random-effects models for meta-analysis". In: *Research synthesis methods* 1.2 (Apr. 2010), pp. 97–111. DOI: 10.1002/jrsm.12.

[6] Isabelle Boutron et al. "Chapter 7: Considering bias and conflicts of interest among the included studies [last updated August 2022]". In: *Cochrane Handbook for Systematic Reviews of Interventions*. Ed. by Julian P. T. Higgins et al. Version 6.5 (updated August 2024). Cochrane, 2024. URL: https://training.cochrane.org/handbook.

[7] Katherine S. Button et al. "Power failure: why small sample size undermines the reliability of neuroscience". In: *Nature Reviews Neuroscience* 14 (2013), pp. 365–376. DOI: 10.1038/nrn3475.

[8] Rudolf Carnap. "On the Application of Inductive Logic". In: *Philosophy and Phenomenological Research* 8.1 (1947), pp. 133–148. DOI: 10.2307/2102920.

[9] Bob Carpenter et al. "Stan: A probabilistic programming language". In: *Journal of statistical software* 76.1 (2017). DOI: 10.18637/jss.v076.i01.

[10] Lorenzo Casini and Jürgen Landes. "Bias, Conflict of Interest, and the Principle of Total Evidence". In: *Logical Reasoning with Imperfect Information in Social Settings*. Ed. by Mario Piazza, Matteo Tesi, and Pietro Vigani. Forthcoming. 2025.

[11] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates, 1988. DOI: 10.4324/9780203771587.

[12] Francesco De Pretis and Jürgen Landes. "A softmax algorithm for evidence appraisal aggregation". In: *PLoS ONE* 16.6 (2021), pp. 1–23. DOI: 10.1371/journal.pone.0253057.

[13] Francesco De Pretis, Jürgen Landes, and Barbara Osimani. "E-Synthesis: a Bayesian Framework for Causal Assessment in Pharmacosurveillance". In: *Frontiers in Pharmacology* 10 (2019). DOI: 10.3389/fphar.2019.01317.

[14] Jonathan J. Deeks, Julian P. T. Higgins, Douglas G. Altman, Joanne E. McKenzie, and Areti Angeliki Veroniki. "Chapter 10: Analysing data and undertaking meta-analyses [last updated November 2024]". In: *Cochrane Handbook for Systematic Reviews of Interventions*. Ed. by Julian P. T. Higgins et al. Version 6.5 (updated August 2024). Cochrane, 2024. URL: https://training.cochrane.org/handbook.

[15] Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. 4th ed. Pearson New International Edition. Pearson, 2014.

[16] Shanil Ebrahim, Sheena Bance, Abha Athale, Cindy Malachowski, and John P. A. Ioannidis. "Meta-analyses with industry involvement are massively published and report no caveats for antidepressants". In: *Journal of Clinical Epidemiology* 70 (Feb. 2016), pp. 155–163. DOI: 10.1016/j.jclinepi.2015.08.021.

[17] Evangelos Evangelou and Areti Angeliki Veroniki, eds. *Meta-Research: Methods and Protocols*. New York: Springer, 2022. DOI: 10.1007/978-1-0716-1566-9.

[18] Alberto Falk Delgado and Anna Falk Delgado. "The association of funding source on effect size in randomized controlled trials: 2013-2015 - a cross-sectional survey and meta-analysis." In: *Trials* 18.125 (Mar. 2017). DOI: 10.1186/s13063-017-1872-0.

[19] Maria Elena Flacco et al. "Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor". In: *Journal of Clinical Epidemiology* 68.7 (July 2015), pp. 811–820. DOI: 10.1016/j.jclinepi.2014.12.016.

[20] Lee S. Friedman and Elihu D. Richter. "Relationship between conflicts of interest and research results." In: *J Gen Intern Med* 19.1 (Jan. 2004), pp. 51–56. DOI: 10.1111/j.1525-1497.2004.30617.x.

[21] I. J. Good. "On the Principle of Total Evidence". In: *The British Journal for the Philosophy of Science* 17.4 (1967), pp. 319–321. DOI: 10.1093/bjps/17.4.319.

[22] Jessica Gurevitch, Julia Koricheva, Shinichi Nakagawa, and Gavin Stewart. “Meta-analysis and the science of research synthesis”. In: *Nature* 555.7695 (2018), pp. 175–182. DOI: 10.1038/nature25753.

[23] Larry V. Hedges and Ingram Olkin. *Statistical Methods for Meta-Analysis*. Academic Press, 1985. DOI: 10.1016/C2009-0-03396-0.

[24] Michael P. Hengartner. *Evidence-biased Antidepressant Prescription*. Chichester: Palgrave Macmillan Cham, 2022. DOI: 10.1007/978-3-030-82587-4.

[25] Julian P. T. Higgins and Sally Green, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley, 2008. DOI: 10.1002/9780470712184.

[26] Julian P. T. Higgins, Tianjing Li, and Jonathan J. Deeks. “Chapter 6: Choosing effect measures and computing estimates of effect [last updated August 2023]”. In: *Cochrane Handbook for Systematic Reviews of Interventions*. Ed. by Julian P. T. Higgins et al. Version 6.5 (updated August 2024). Cochrane, 2024. URL: https://training.cochrane.org/handbook.

[27] Julian P. T. Higgins, Jelena Savović, Matthew J. Page, Roy G. Elbers, and Jonathan A. C. Sterne. “Chapter 8: Assessing risk of bias in a randomized trial [last updated October 2019]”. In: *Cochrane Handbook for Systematic Reviews of Interventions*. Ed. by Julian P. T. Higgins et al. Version 6.5 (updated August 2024). Cochrane, 2024. URL: https://training.cochrane.org/handbook.

[28] Julian P. T. Higgins et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.5 (updated August 2024). Cochrane, 2024. URL: https://training.cochrane.org/handbook.

[29] John P. A. Ioannidis. “Hijacked evidence-based medicine: stay the course and throw the pirates overboard”. In: *Journal of Clinical Epidemiology* 84 (2017), pp. 11–13. DOI: 10.1016/j.jclinepi.2017.02.001.

[30] John P. A. Ioannidis. “What meta-research has taught us about research and changes to research practices”. In: *Journal of Economic Surveys* (2024). DOI: 10.1111/joes.12666.

[31] Lise L. Kjaergard and Bodil Als-Nielsen. “Association between competing interests and authors’ conclusions: epidemiological study of randomised clinical trials published in the BMJ”. In: *BMJ* 325.7358 (Aug. 2002), p. 249. DOI: 10.1136/bmj.325.7358.249.

[32] Jürgen Landes, Barbara Osimani, and Roland Poellinger. “Epistemology of Causal Inference in Pharmacology”. In: *European Journal for Philosophy of Science* 8 (1 2018), pp. 3–49. DOI: 10.1007/s13194-017-0169-1.

[33] D. N. Lathyris, N. A. Patsopoulos, G. Salanti, and J. P. A. Ioannidis. “Industry sponsorship and selection of comparators in randomized clinical trials”. In: *European Journal of Clinical Investigation* 40.2 (2010), pp. 172–182. DOI: 10.1111/j.1365-2362.2009.02240.x.

[34] Joel Lexchin. “Those who have the gold make the evidence: how the pharmaceutical industry biases the outcomes of clinical trials of medications.” In: *Science and Engineering Ethics* 18.2 (June 2012), pp. 247–261. DOI: 10.1007/s11948-011-9265-3.

[35] Joel Lexchin, Lisa A. Bero, Benjamin Djulbegovic, and Otavio Clark. “Pharmaceutical industry sponsorship and research outcome and quality: systematic review”. In: *BMJ* 326.7400 (2003), pp. 1167–1170. DOI: 10.1136/bmj.326.7400.1167.

[36] Andreas Lundh and Lisa Bero. “The ties that bind”. In: *BMJ* 356 (2017). DOI: 10.1136/bmj.j176.

[37] Andreas Lundh, Joel Lexchin, Barbara Mintzes, Jeppe B Schroll, and Lisa Bero. “Industry sponsorship and research outcome”. In: *Cochrane Database of Systematic Reviews* 2.2 (Feb. 2017), MR000033. DOI: 10.1002/14651858.MR000033.pub3.

[38] L. R. Magni et al. “Fluoxetine versus other types of pharmacotherapy for depression”. In: *Cochrane Database of Systematic Reviews* 7 (2013). DOI: 10.1002/14651858.CD004185.pub3.

[39] John H. Montgomery et al. “An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia”. In: *Controlled Clinical Trials* 25.6 (2004), pp. 598–612. DOI: 10.1016/j.cct.2004.09.002.

[40] National Academies of Sciences, Engineering, and Medicine. *Sponsor Influences on the Quality and Independence of Health Research*. Ed. by Dara Rosenberg, Elizabeth Boyle, Alexandra McKay, and Joe Alper. Washington: National Academies Press, 2023. DOI: 10.17226/27056.

[41] Piia K. Peura, Janne A. Martikainen, Timo T. Purmonen, and Juha H. O. Turunen. “Sponsorship-related outcome selection bias in published economic studies of triptans: systematic review”. In: *Med Decis Making* 32.2 (Mar. 2012), pp. 237–245. DOI: 10.1177/0272989X11403834.

[42] Florian Prinz, Thomas Schlange, and Khusru Asadullah. “Believe it or not: how much can we rely on published data on potential drug targets?” In: *Nature Reviews Drug Discovery* 9.10 (712 2011), pp. 328–329. DOI: 10.1038/nrd3439-c1.

[43] Pascal Probst et al. “Industry Bias in Randomized Controlled Trials in General and Abdominal Surgery: An Empirical Study”. In: *Annals of Surgery* 264.1 (2016), pp. 87–92. DOI: 10.1097/SLA.0000000000001372.

[44] Ivette Raices Cruz, Matthias C. M. Troffaes, Johan Lindström, and Ullrika Sahlin. “A robust Bayesian bias-adjusted random effects model for consideration of uncertainty about bias terms in evidence synthesis”. In: *Statistics in Medicine* 41.17 (2022), pp. 3365–3379. DOI: 10.1002/sim.9422.

[45] T Reynolds. “Industry-funded versus publicly funded trials: are the standards the same?” In: *JNCI: Journal of the National Cancer Institute* 93.21 (Nov. 2001), pp. 1590–1592. DOI: 10.1093/jnci/93.21.1590.

[46] “RoB 2: a revised tool for assessing risk of bias in randomised trials”. In: *BMJ* (Aug. 2019). DOI: 10.1136/bmj.l4898.

[47] M. Roseman et al. “Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments”. In: *JAMA* 305.10 (2011), pp. 1008–17. DOI: 10.1001/jama.2011.257.

[48] D. L. Sackett, W. M. C. Rosenberg, J. A. M. Gray, R. B. Haynes, and W. S. Richardson. “Evidence based medicine: what it is and what it isn’t”. In: *BMJ* 312.7023 (1996), pp. 71–2. DOI: 10.1136/bmj.312.7023.71.

[49] Holger Schünemann, Jan Brożek, Gordon Guyatt, and Andrew Oxman. *GRADE handbook for grading quality of evidence and strength of recommendations*. Ed. by The GRADE Working Group. Updated October 2013. Oct. 2013. URL: https://guidelinedevelopment.org/handbook.

[50] Sonal Singh and Yoon K. Loke. “Drug safety assessment in clinical trials: methodological challenges and opportunities”. In: *Trials* 13.1 (2012). DOI: 10.1186/1745-6215-13-138.

[51] Jacob Stegenga. *Medical Nihilism*. Oxford: Oxford University Press, 2018.

[52] Xin Tong and Zijun Ke. “Assessing the Impact of Precision Parameter Prior in Bayesian Nonparametric Growth Curve Modeling”. In: *Frontiers in Psychology* 12 (Mar. 2021), p. 624588. DOI: 10.3389/fpsyg.2021.624588.

[53] Gail A. Van Norman. “Drugs and Devices”. In: *JACC: Basic to Translational Science* 1.5 (2016), pp. 399–412. DOI: 10.1016/j.jacbts.2016.06.003.

[54] Gail A. Van Norman. “Update to Drugs, Devices, and the FDA”. In: *JACC: Basic to Translational Science* 5.8 (2020), pp. 831–839. DOI: 10.1016/j.jacbts.2020.06.010.

[55] Jan P. Vandenbroucke and Bruce M. Psaty. “Benefits and risks of drug treatments: How to combine the best evidence on benefits with the best data about adverse effects”. In: *JAMA* 300.20 (2008), pp. 2417–2419. DOI: 10.1001/jama.2008.723.

[56] Matthew Vassar, Samuel Shepard, Simran Demla, and Daniel Tritz. “Correlation analysis of financial conflicts of interest and favourability of results or conclusions in addiction medicine systematic reviews and meta-analysis”. In: *BMJ Open* 12.8 (2022), e054325. DOI: 10.1136/bmjopen-2021-054325.

[57] Pablo Emilio Verde. “A bias-corrected meta-analysis model for combining, studies of different types and quality”. In: *Biometrical Journal* 63.2 (2020), pp. 406–422. DOI: 10.1002/bimj.201900376.

[58] Areti Angeliki Veroniki et al. “Methods to estimate the between-study variance and its uncertainty in meta-analysis”. In: *Research Synthesis Methods* 7.1 (2016), pp. 55–79. DOI: 10.1002/jrsm.1164.

[59] Areti Angeliki Veroniki et al. “Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis”. In: *Research Synthesis Methods* 10.1 (2019), pp. 23–43. DOI: 10.1002/jrsm.1319.

[60] Wolfgang Viechtbauer. “Approximate Confidence Intervals for Standardized Effect Sizes in the Two-Independent and Two-Dependent Samples Design”. In: *Journal of Educational and Behavioral Statistics* 32.1 (2007), pp. 39–60. DOI: 10.3102/1076998606298034.

[61] Ahmed Waqas, Ammad Anwar Baig, Muhammad Ali Khalid, Kapil Kiran Aedma, and Sadiq Naveed. “Conflicts of interest and outcomes of clinical trials of antidepressants: An 18-year retrospective study”. In: *Journal of Psychiatric Research* 116 (2019), pp. 83–87. DOI: 10.1016/j.jpsychires.2019.05.029.

[62] N. J. Welton, A. E. Ades, J. B. Carlin, D. G. Altman, and J. A. C. Sterne. “Models for potentially biased evidence in meta-analysis using empirically based priors”. In: *Journal of the Royal Statistical Society* 172 (2009), pp. 119–36. DOI: 10.1111/j.1467-985X.2008.00548.x.