UNIVERSITY OF GÖTTINGEN

MASTER THESIS

Between-Trial Heterogeneity in Meta-Regression

Author:
Malte KRAMER

Supervisors:
Prof. Dr. Tim FRIEDE
Dr. Christian RÖVER

A thesis submitted in fulfillment of the requirements for the degree of Master of Science in Applied Statistics

at the

University Medical Center Göttingen
Department of Medical Statistics

4th July 2022

Declaration of Authorship

I, Malte KRAMER, declare that this thesis titled, "Between-Trial Heterogeneity in Meta-Regression" and the work presented in it are my own. I confirm that:

- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

Abstract

Using random effects meta-analysis requires an estimate for the between-trial heterogeneity. Between-trial heterogeneity is difficult to estimate when the number of studies is small, which is very common in clinical research. We propose using meta-regression rather than individual meta-analyses when (similar) subgroups are analyzed in order to have more data to estimate the between-trial heterogeneity. When conducting a meta-analysis in the presence of subgroups, often individual meta-analyses per subgroup are carried out. We investigate whether it is beneficial to use a meta-regression instead of individual meta-analyses. Data from the "Cochrane Database of Systematic Reviews" showed that the interval lengths for the overall effect μ shrunk when using meta-regression, indicating potentially higher precision. We then conducted a simulation study that showed that the coverage probability of meta-regression is closer to the predefined level while again shrinking the interval lengths for μ. This effect is even apparent if the true τ values in each subgroup differ substantially (i.e., by factor 2). The benefits of meta-regression, however, disappear if the difference in τ becomes larger.

Keywords: Between-Trial Heterogeneity, Meta-Analysis, Meta-Regression, Simulation, Cochrane Database of Systematic Reviews

Acknowledgements

I would like to thank Prof. Tim Friede and Dr. Christian Röver for the collaboration and helpful advice. Secondly, I want to thank Dieter Kramer for his support regarding this thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

CDSR	Cochrane database of systematic reviews
CI	Confidence / credible interval
DL	DerSimonian and Laird
ECDF	Empirical cumulative distribution function
HNP	Half-normal prior
ML	Maximum likelihood
MSE	Mean squared error
NNHM	Normal-normal hierarchical model
PM	Paule and Mandel
REML	Restricted maximum likelihood
UP	Uniform prior

Introduction

Meta-analyses are used to combine results of multiple related studies with similar characteristics. As an example, a meta-analysis of randomized controlled trials comparing a treatment to placebo can use the study estimates to synthesize an overall estimate of the treatment effect (Riley et al., 2011). Meta-analyses assume either a fixed effect or random effects across studies. A fixed effect meta-analysis assumes that a true common (fixed) effect is estimated in the studies, whereas a random effects meta-analysis allows that the true effect could vary due to between-trial heterogeneity. In order to generalize the conclusions beyond the actual studies included in the meta-analysis, the random effects model is often more appropriate since it includes uncertainty resulting from heterogeneity among studies (Borenstein et al., 2010). Especially for small numbers of studies in a meta-analysis, the between-trial heterogeneity is hard to estimate (Veroniki et al., 2016). Frequentist methods tend to estimate it to be zero (Williams et al., 2018).

If in the studies, subgroups that potentially behave similar are identified, the question arises how the analysis should be conducted. In general, besides an overall meta-analysis, often subgroup meta-analyses are performed. The assumption made in the overall meta-analysis are an equal overall effect μ and between trial heterogeneity τ across subgroups, whereas the individual subgroup meta-analyses do not assume that. Individual meta-analyses assume an overall effect μ and a common between-trial heterogeneity τ per subgroup.

Meta-regression is used to incorporate predictors in the meta-analysis. Either categorical or continuous predictors can be used. Meta-regression can be seen as a compromise between an overall meta-analysis and subgroup specific meta-analyses. Meta-regression assumes a common between-trial heterogeneity while allowing for subgroup specific estimates. In the case

where an overall meta-analysis is considered, the assumption of a common between trial heterogeneity in meta-regression should usually be reasonable, since the assumption of a common between-trial heterogeneity was already made. Meta-regression promises better properties, since more data is available to estimate the between-trial heterogeneity, and therefore we expect better confidence/ credible intervals (Dias et al., 2013). We furhter expect that for the frequentist methods the number of zero estimates for τ is reduced when using meta-regression.

We used empirical data from the Cochrane Database of Systematic Reviews to compute meta-regressions on data that contained subgroups. This was done to check the properties of meta-regression on real data compared to individual meta-analyses. Moreover, we conducted a simulation study to further check the properties of the meta-regression in comparison to individual meta-analyses.

Methodology

2.1 General remarks

First, the used methods are introduced. This includes meta-analysis and meta-regression. For these methods, between-trial heterogeneity and confidence /credible intervals for the overall effect(s) μ are described in detail. Lastly, choices for the design matrix in meta-regression are considered.

2.2 Study-level estimates and standard errors

Meta-analyses are used to combine findings from multiple studies (Friede et al., 2017a). Is is often used in systematic reviews to combine evidence by pooling multiple estimates (Glass, 1976; Viechtbauer, 2010).

In this work, we consider the results of clinical trials. We consider the case of binary data in the form of a 2×2 contingency table for each study. In this table, control and treatment group are summarized in terms of patients experiencing a predefined outcome or not.

TABLE 2.1: Contingency table for a single study

	Event	No Event	Total
Treatment	n_{11}	n_{12}	n_{1+}
Control	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

From Table 2.1, we compute a coefficient for the degree of association between treatment group and event occurrence, e.g., the log odds ratio and its standard errors that will be used in the meta-analysis. The log-odds ratio Y_i

is the logarithm of the odds ratio, which is defined as follows:

$$e^{Y_i} = \frac{(n_{11}/n_{1+})/(n_{12}/n_{1+})}{(n_{21}/n_{2+})/(n_{22}/n_{2+})} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \tag{2.1}$$

Odds are the ratio of the frequency that an event will happen over the frequency that it does not. The odds ratio is the ratio of the odds for experiencing an event if treated (n_{11}/n_{12}) and the odds if not treated (n_{21}/n_{22}).

An odds ratio of 1 would imply that the treatment had no effect. Using the log transformation directly leads to a log odds ratio of 0, indicating no effect. Furthermore, log-odds have symmetric properties that are useful (Röver, 2020).

The approximate standard error for the log-odds ratio can be computed by (Hedges & Olkin, 1985; Röver, 2020):

$$\sigma_{Y_i} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \tag{2.2}$$

2.3 The normal-normal hierarchical model

Once we have calculated the effect measures Y_i for multiple studies, a meta-analysis can be conducted. Standard models for meta-analysis assume either a fixed effect or random effects across studies (Friede et al., 2017b). For random effects meta-analysis, commonly a normal-normal hierarchical model (NNHM) is used. We assume that the study estimates Y_i follow an approximate normal distribution with mean θ_i and variance s_i^2.

$$Y_i|\theta_i \sim \mathcal{N}(\theta_i, s_i^2), i \in 1, \ldots, k \tag{2.3}$$

For simplicity, the standard errors are assumed to be known. At the next level, we assume normal distributed study effects θ_i.

$$\theta_i|\mu, \tau \sim \mathcal{N}(\mu, \tau^2), i \in 1, \ldots, k \tag{2.4}$$

The between-trial heterogeneity, here denoted by τ, is the standard error of the θ_i and captures the amount of heterogeneity between studies. If the heterogeneity is zero, the random effects model simplifies to a common-effect (or fixed effect) model. In this thesis, the terms common-effect model and

fixed-effect model are used interchangeably. One can combine the two stages, which leads to the marginal model (Friede et al., 2017b):

$$Y_i|\mu,\tau \sim \mathcal{N}(\mu, s_i^2 + \tau^2), i \in 1,\ldots,k \tag{2.5}$$

Inference for μ and the nuisance parameter τ can be done in either the frequentist or Bayesian framework. For known between-trial heterogeneity τ and non-informative prior for μ, both frameworks lead to analogous results (Friede et al., 2017a).

2.4 Frequentist methods

2.4.1 Estimation of μ

In the frequentist case, μ is usually inferred as follows:

$$\hat{\mu} = \sum_{i=1}^{k} w_i Y_i / \sum_{j=1}^{k} w_j \tag{2.6}$$

w_i are the inverse variance weights given by (Friede et al., 2017a):

$$w_i = 1/(s_i^2 + \tau^2), j \in 1,\ldots,k. \tag{2.7}$$

In order to estimate μ, τ is estimated beforehand and then plugged in. Estimates for τ are discussed in Section 2.4.2. The sum over all inverse variance weights is called the total precision:

$$w_+ = \sum_{i=1}^{k} w_i. \tag{2.8}$$

It is further notable that $\hat{\mu}$ has a standard error of:

$$\hat{\sigma}_\mu = 1/w_+. \tag{2.9}$$

2.4.2 Between-trial heterogeneity τ

Inference for μ is commonly based on first deriving an estimate for τ, from which the inverse variance weights w_i can be calculated. Afterwards μ can be estimated using τ and w_i as shown in (2.6). There are various proposed

methods to estimate the between trial heterogeneity τ. The methods used here are summarized in the following section. Additionally, the maximum likelihood method is considered.

The between-trial heterogeneity estimates described here can be found in Veroniki et al. (2016). The following summaries are heavily based on this publication. Additional estimators can be found in Veroniki et al. (2016).

DerSimonian and Laird (DL)

The heterogeneity estimator by DerSimonian and Laird (1986) is based on Cochran's Q-Statistic. It is a measure of deviation of subject specific effects and the pooled estimate from a fixed effect model. The estimator is given by:

$$\tau^2_{DL} = \max\left\{0, \frac{Q-(k-1)}{\sum w_{i,FE} - \frac{\sum w^2_{i,FE}}{\sum w_{i,FE}}}\right\}, \quad Q = \sum_{i=1}^{k} \frac{(y_i - \hat{\mu}_{FE})^2}{s_i^2} \tag{2.10}$$

with $w_{i,FE} = 1/s_i$. $\hat{\mu}_{FE}$ is an estimate for μ from a fixed effect model. Variances are by definition larger than or equal to zero. Therefore, negative estimates for τ are usually truncated to zero, which introduces a positive bias (Viechtbauer, 2005). Before the truncation, the estimate is unbiased if the sampling variances are known (DerSimonian & Laird, 1986). Potential reasons for bias across all methods other than the estimation of s_i^2 are a bias in the treatment effect or a correlation between treatment effects and their variances s_i^2. For large number of studies and small τ, the bias of the DL is found to be acceptable (Veroniki et al., 2016).

However, large τ in case of binary data introduce a large negative bias to the DL estimate (Bowden et al., 2011; Novianti et al., 2014; Sidik & Jonkman, 2007). The negative bias is probably related to effect measures that are based on 2×2 contingency tables. This is also a problem for all other methods. Large τ can lead to including 0 cells in the tables, such that the method of inverse variance weights becomes questionable. This results in what is called "single" and-or "double zero" studies. Further information on this topic can be found in Günhan et al. (2020).

The DL estimator is shown to asymptotically approach the Cramer-Rao lower bound (Jackson et al., 2010).

Jackson et al. (2010) also showed that the DL estimate is inefficient when the trial sizes differ significantly. However, including large number of studies leads to efficient inference on μ.

Paule Mandel (PM)

Following Dersimonian and Kacker (2007), Cochrane's Q statistic can be generalized to:

$$Q_a = \sum a_i (y_i - \hat{\mu}_a)^2 \tag{2.11}$$

$a_i \in \mathbb{R}_+$ being the weight given to each study. And $\mu_a = \sum \frac{a_i y_i}{\sum a_i}$. Analogous to the DL, the expectation of Q_a is given by:

$$E(Q_a) = \tau^2 \left(\sum a_i - \frac{\sum a_i^2}{\sum a_i} \right) + \sum a_i s_i^2 - \frac{a_i s_i^2}{\sum a_i} \tag{2.12}$$

Solving for τ^2 gives the generalized method of moments (GMM) estimator.

$$\tau^2_{GMM} = \max \left\{ 0, \frac{Q_a - \left(\sum a_i s_i^2 - \frac{\sum a_i^2 s_i^2}{\sum a_i} \right)}{\sum a_i - \frac{\sum a_i^2}{\sum a_i}} \right\} \tag{2.13}$$

Other estimates (as well as the DL estimate) can be transformed to this general form.

Paule and Mandel (1982) used a form of Q_a with $a_i = w_{i,RE} = 1/(s_i^2 + \tau^2)$ and the fact that it follows a χ^2 distribution under the assumptions made in the random effects model.

$$Q = \sum w_{i,RE} (y_i - \hat{\mu}_{RE}(\tau^2))^2 \sim \chi^2_{k-1}. \tag{2.14}$$

$\hat{\tau}^2_{PM}$ can be found by iterating through values of τ^2 in $Q(\tau^2)$, until it is equal to the expected value: $k-1$. If $Q(0) \geq k-1$, this procedure leads to a unique estimate of $\hat{\tau}^2_{PM}$. If $Q(0) < k-1$, no positive solution can be found and $\hat{\tau}^2_{PM}$ is set to 0 (van Aert & Jackson, 2018).

When underlying assumption do not hold, the PM estimator is more robust compared to the DL estimator (Rukhin et al., 2000). For small sample sizes k

and small τ, the PM estimator is shown to be positively biased, whereas for large k and large τ, the PM estimator shows a negative bias (Sidik & Jonkman, 2007). Panityakul et al. (2013) moreover, suggests using the PM estimator instead of DL.

Maximum likelihood (ML)

The ML between-trial heterogeneity estimator is an iterative method that is asymptotically efficient. ML is asymptotically unbiased, and the variance is approaching the Cramer-Rao lower bound (Veroniki et al., 2016). It is based on the marginal distribution, $y_i \sim N(\mu, s_i^2 + \tau^2)$ and the estimate τ^2_{ML} can be found maximizing:

$$ll(\mu, \tau^2) = -\frac{k}{2}(2\pi) - \frac{1}{2}\sum ln(s_i^2 + \tau^2) - \frac{1}{2}\sum \frac{(y_i - \mu)^2}{(s_i^2 + \tau^2)} \tag{2.15}$$

$$\hat{\mu}_{RE}(\hat{\tau}^2_{ML}) = \frac{\sum w_{i,RE} y_i}{w_{i,RE}} \tag{2.16}$$

$$\hat{\tau}^2_{ML} = \max\left\{0, \frac{\sum w_{i,RE}((y_i - \hat{\mu}_{RE}(\hat{\tau}^2_{ML}))^2 - s_i^2)}{w^2_{i,RE}}\right\} \tag{2.17}$$

The estimates are obtained, iterating over $\hat{\tau}^2_{ML}$, $\hat{\mu}_{RE}(\hat{\tau}^2_{ML})$ until convergence. In each iteration, negative estimates for the between-study heterogeneity $\hat{\tau}^2_{ML}$ are set to 0. Methods to maximize the likelihood are expectation-maximization-algorithm, Newton-Raphson and Fisher-Scoring. Iterative estimators have the disadvantage that they can fail to converge. For small numbers of studies included in the meta-analysis, the likelihood can be flat and thus hard to maximize. In these cases, it can be useful to either apply closed form estimators or incorporate informative priors if Bayesian models are used (Pullenayegum, 2011; Rhodes et al., 2015; Turner et al., 2012).

For large τ^2 and small numbers of studies per analysis, the between study estimate is found to be negatively biased (Chung et al., 2013; Sidik & Jonkman, 2007; Veroniki et al., 2016). Overall, it is not suggested to use the ML estimate (Panityakul et al., 2013; Viechtbauer, 2005).

Restricted maximum likelihood (REML)

The REML estimator is analogous to the ML estimate, an iterative procedure. It has the same asymptotic behavior. The REML estimator is used to correct the large negative bias that is induced by the ML estimate, up to a degree. $\hat{\tau}^2_{REML}$ is obtained by equating the derivative of the restricted log likelihood with respect to τ^2 equal to zero.

$$\begin{aligned} ll(\tau^2) = &-\frac{k}{2}ln(2\pi) - \frac{1}{2}\sum ln(s_j^2 + \tau^2) \\ &- \frac{1}{2}\sum \frac{(y_i - \hat{\mu}_{RE}(\hat{\tau}^2_{ML}))^2}{s_i^2 + \tau^2} - \frac{1}{2}ln\left(\sum \frac{1}{s_i^2 + \tau^2}\right) \end{aligned} \tag{2.18}$$

Afterwards, solving for $\hat{\tau}^2$ results in:

$$\tau^2_{REML} = \max\left\{0, \frac{\sum w^2_{i,RE} + (\theta_i - \hat{\mu}_{RE}(\hat{\tau}^2_{ML})) - s_i^2}{w^2_{i,RE}} + \frac{1}{w_{i,RE}}\right\} \tag{2.19}$$

with $w_{i,RE} = 1/(s_i^2 + \hat{\tau}^2_{REML})$. Analogous to ML, $\hat{\tau}^2_{REML}$ is calculated by iterating until convergence with an initial $\hat{\tau}^2_{REML} \geq 0$. Especially for small numbers of studies included in the analysis, the REML estimator underestimates τ^2 (Veroniki et al., 2016).

2.4.3 Confidence Intervals for the overall effect μ

There are multiple ways to compute the confidence intervals (CI(s)) for the overall effect μ.

Z-CI (normal approximation)

The first CI for μ is based on the normal approximation (Röver et al., 2015; Sánchez-Meca & Marín-Martínez, 2008).

$$CI_z : \hat{\mu} \pm z_{(1-\alpha/2)}\, \hat{\sigma}_\mu \tag{2.20}$$

The standard error for $\hat{\mu}$ given an estimate for τ is defined in (2.9). $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution, where $1-\alpha$ is the targeted (or nominal) coverage probability. The goodness of the normal approximation is dependent on a large number of studies k (Hartung & Knapp, 2001a, 2001b; Hartung & Makambi, 2003; Röver et al., 2015). This is due to

the fact that both between-trial heterogeneity and study specific standard errors (s_i) are estimated and not known. As shown by Viechtbauer (2005), the standard error of $\hat{\mu}$ is underestimated on average, resulting in CIs that are too narrow.

t-CI

In order to address the coverage problems from the Z-CI, CIs based on t-distribution have been proposed (Follmann & Proschan, 1999; Hartung & Makambi, 2002; Sánchez-Meca & Marín-Martínez, 2008). Here a t-distribution with $k-1$ degrees of freedom instead of the standard normal distribution is used

$$CI_t : \hat{\mu} \pm t_{(k-1),(1-\alpha/2)}\ \hat{\sigma}_\mu \tag{2.21}$$

$t_{(k-1),(1-\alpha/2)}$ is the $(1-\alpha/2)$-quantile of a t-distribution with k-1 degrees of freedom. CIs based on a t-distribution are wider compared to those of a standard normal, therefore this should lead to higher coverage (Follmann & Proschan, 1999).

KnHa-CI

Hartung and Knapp (2001b) and Sidik and Jonkman (2002) have proposed another adjusted CI based on a t-distribution. In contrast to the t-CI, the following quadratic form is computed:

$$q = \frac{1}{k-1}\sum_i w_i\ (y_i - \hat{\mu})^2 \tag{2.22}$$

Hartung (1999) has shown that $q\hat{\sigma}_\mu^2$ is an unbiased non-negative estimator of σ_μ^2. This estimator of σ_μ^2 is used in the CI (Röver et al., 2015). The adjusted CI is then given by:

$$CI_{KnHa} : \hat{\mu} \pm \sqrt{q}\ \hat{\sigma}_\mu\ t_{(k-1),(1-\alpha/2)} \tag{2.23}$$

Adhoc-CI

However, the KnHa-CI can be shorter than the normal one if $\sqrt{q} < \frac{z_{(1-\alpha/2)}}{t_{(k-1),(1-\alpha/2)}}$.

To ensure that the adjusted KnHa-CI (Adhoc-CI) is at least the length of the Z-CI, we substitute $q* = max\{1, q\}$ for q. The CI is then given by:

$$CI_{adhoc} : \hat{\mu} \pm \sqrt{q*}\ \hat{\sigma}_\mu\ t_{(k-1),(1-\alpha/2)} \tag{2.24}$$

As mentioned in Röver et al. (2015), the PM estimator is defined such that $q* = 1$ or less if there is no solution. This leads to $q*$ always being equal to 1.

2.5 Bayesian methods

2.5.1 General remarks

Bayesian models allow for the inclusion of prior knowledge into the analysis. This can be seen as an advantage or a disadvantage. Bayesian inference readily allows accounting for the estimation uncertainty of τ in the inference stage.

Bayesian models with carefully chosen priors usually give a uni-modal posterior distribution for τ, with a point estimate > 0, e.g., the posterior median.

Further, Bayesian models allow calculating the probability that μ or τ is larger (or smaller) than a given value from the posterior distribution. The posterior distribution is a combination of prior and likelihood, and thus summarizes what we know after the data have been observed.

2.5.2 Priors for μ and τ

The NNHM in the Bayesian Framework involves the selection of a prior for μ and τ. For convenience, we assume that we can factor out the prior density: $p(\mu, \tau) = p(\mu) \times p(\tau)$ (Röver, 2020). When choosing priors in Bayesian models, we consider proper or improper, and informative or uninformative priors. A prior is proper when the prior integrates to one: $\int_{-\infty}^{\infty} p(.) = 1$. It is considered improper if it does not: $\int_{-\infty}^{\infty} p(.) \neq 1$. Still, both may lead to proper posterior distributions.

Informative priors are not well-defined. They can be seen as priors that are not dominated by the likelihood (for small amounts of data). They have direct impact on the posterior distribution. With informative priors, we can include prior information to a model. An uninformative prior on the other hand is dominated by the likelihood. Therefore, it has minimal impact on the posterior distribution. A typical uninformative prior for, e.g., a location parameter is the uniform distribution over the real line.

As a prior distribution for μ, we consider either a non-informative prior in the form of a uniform distribution over the real line or a normal distribution with parameters μ_p, σ_p^2 as an informative prior. Both uniform or normal prior

for μ lead to a normally distributed conditional posterior for a given τ with mean $\mu(\tau)$ and variance $\sigma(\tau)$ (Röver, 2020).

For $\sigma_p \to \infty$ the normal prior tends to the uniform prior. In the case of log odds as an endpoint, one can consider using a normal distribution centered around zero as a "natural" choice, since log odds of zero would indicate an odds ratio of one and thus indicate no effect (Röver, 2020).

When choosing a prior distribution for τ, one has to keep in mind that the variance by definition is larger or equal to zero. To account for that, the prior distributions need to be defined on $\mathbb{R}_+$. For example, one can choose half-normal, half-Cauchy or a half-Student-t distribution as a prior (Röver, 2020).

Using a uniform distribution as a prior for τ needs more than 3 individual study estimates to produce an overall estimate. The uniform prior is scale invariant and not dependent on σ_i. The other mentioned priors do not come with a restriction on how many study estimates are needed.

2.5.3 Credible intervals

Bayesian credible intervals (CI(s)) are analogous to frequentist CIs. However, their interpretation is different. We consider either the so-called shortest interval and the central interval. The central interval is an equal-tailed interval using the respective posterior distributions $\alpha/2$ and $1-\alpha/2$ quantiles. Another choice is the highest posterior density interval (HPDI) which covers $1-\alpha$ posterior probability, where the posterior density is the largest (Gelman et al., 2015). Finding the shortest interval from the respective posterior distribution is closely related to the HPDI, but way easier to determine (Röver, 2020). For uni-modal distributions, they are even equivalent. In the following, the shortest interval will be used.

2.6 Meta-regression

Analogous to meta-analysis, meta-regression is a generalization of the NNHM. For this section, we use matrix notation. The first stage in meta-regression is the same as in the NNHM for meta-analysis:

$$y|\theta, s \sim \mathcal{N}(\theta, \Sigma) \quad \text{where} \quad \Sigma = \operatorname{diag}(s_1^2, \ldots, s_k^2),\ \theta = (\theta_1, \ldots, \theta_k) \tag{2.25}$$

The difference lies in the second stage. Here μ from before is replaced by the product of design matrix X and regression coefficient vector β. This is done so that predictors can be incorporated into the analysis.

$$\theta|X, \beta, \tau \sim \mathcal{N}(X\beta, \tau^2 I) \tag{2.26}$$

As before, we can express this as the marginal model:

$$y|X, \beta, \tau, s \sim \mathcal{N}(X\beta, \Sigma_\tau) \quad \text{where} \quad \Sigma_\tau = \Sigma + \tau^2 I \tag{2.27}$$

The choices for the matrix X are further explained in Section 2.7. We only consider predictors that contain indicator variables to represent subgroups. The choice of an appropriate design matrix X leads to individual estimates for the treatment effect per subgroup. Compared to individual meta-analyses where we estimate one individual τ for each subgroup, in meta-regression we estimate only one common τ for all. This directly implies a key assumption of this approach, namely, equal τ across subgroups. If we expect that the heterogeneity across subgroups differs substantially, meta-regression should be treated with caution unless the meta-regression model is extended to allow for different heterogeneity parameters in the subgroups.

When a particularly small number of studies per meta-analysis is used, Bayesian approaches using informative priors help to estimate the between-trial heterogeneity. However, the inclusion of prior beliefs about the distribution of τ causes the prior to influence the estimation of τ, especially when the data set is small. Adding more data to update the prior results in less impact from the prior. Therefore, meta-regression for subgroups should yield a more robust result in the case of informative priors.

The methods introduced in the meta-analysis context, namely CIs and between trial heterogeneity estimates, can be generalized to meta-regression and are thus not shown again (Jackson et al., 2014; Jackson & Riley, 2014; van Aert & Jackson, 2018).

Further, meta-regression with only one intercept is one large meta-analysis. Thus, the previously described simulation results for the simple meta-analysis should translate to meta-regression.

2.7 The design matrix in meta-regression

We want to investigate the impact of analyzing multiple subgroups simultaneously. For each subgroup, we would estimate individual overall effects, but at the same time a common between trial heterogeneity.

In order to get individual estimates for each subgroup, the design matrix X has to be defined appropriately: X has the dimensions $k \times p$ where k is the number of studies included in the meta-regression and p is the number of subgroups. For example, if the third study is part of a subgroup q, the third row is filled as follows: $X_{3,q} = 1$ and $X_{3,1} = \cdots = X_{3,q-1} = X_{3,q+1} = \cdots = X_{3,p} = 0$. A general design matrix is given by:

$$X_{k\times p} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{2.28}$$

This leads to p intercepts $\beta_{p\times 1} = (\beta_1, \ldots, \beta_p)$, one for each subgroup.

One could also include continuous predictors in meta-regression. For a given set of r continuous covariates combined with an intercept, the design matrix usually looks as follows:

$$X_{k\times r+1} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{r,1} \\ 1 & x_{1,2} & \cdots & x_{r,2} \\ \vdots & \vdots & \cdots & \vdots \end{pmatrix} \tag{2.29}$$

Here, one intercept β_1 and r additional parameters are estimated:

$$(\beta_1, \beta_2, \ldots, \beta_{r+1}). \tag{2.30}$$

Meta-regression results in a meta-analysis if the design matrix is a vector of ones: $\mathbf{1}_n$.

Empirical Study

3.1 Aim

To assess how meta-regression for subgroups performs in practice, we searched for meta-analyses based on studies that contains binary data. We found such data in the "Cochrane Database of Systematic Reviews".

3.2 Data

The "Cochrane Database of Systematic Reviews" (CDSR) is a journal and database for systematic reviews in health care. Each systematic review has been reviewed by the editorial team of CDSR. The reviews aim to synthesize empirical evidence. Studies that are considered for each meta-analysis have to meet certain pre-specified inclusion criteria to reduce potential biases. All data, estimates and more are available as standardized files on their website to download. The existence of standardized files makes the further analysis of the systematic reviews more efficient and less vulnerable to errors. The files are in "XML" format. They can be either used with the Review Manager (RevMan) or can be loaded into R via the `meta` package.

For the analysis of empirical data, all available CDSR files have been downloaded. In case of new evidence, the reviews are updated. Thus, the most recent version is used. The code for the web-scraper used to obtain the standardized files is based on python. It is available on request.

The simulation and analysis was performed in R. The methods of interest are implemented in `bayesmeta` (UP, HNP) and `metafor` (REML, DL, PM) (Röver, 2020; Viechtbauer, 2010).

3.3 Example meta-analyses and meta-regression

Usually in a Cochrane review, all subgroups are analyzed in separate meta-analyses. Common practice is to also include one estimate of a meta-analysis, where data from all subgroups is analyzed together. However, we are interested in the effect of using one meta-regression instead of multiple meta-analyses. To illustrate this, we show the results of one meta-regression and individual meta-analyses in the form of a forest plot in Figure 3.1. The estimated values of τ are shown in Table 3.1. We have used the Knapp-Hartung method to calculate the μ confidence intervals in the forest plot (Figure 3.1). For the Bayesian credible interval, we used the "shortest" intervals.

The example stems from the Cochrane review of Goodwin et al. (2013). In Goodwin et al. (2013), the addition of radiotherapy to breast conserving surgery for treatment of ductal carcinoma in situ was investigated. Ductal carcinoma in situ of the breast is a pre-malignant condition. Cochrane reviews usually contain one or more comparisons of interest. In the comparisons, usually two treatments are compared. Accordingly, certain controlled studies that have examined this pairwise comparison are used. In these comparisons, one or more outcomes are used to assess the treatment effect. Meta-analyses are then conducted for each of these outcomes. We used the first comparison and the fourth outcome. Namely, post-operative radiotherapy versus surgery alone was compared using the incidence of ipsilateral recurrence by surgical excision as the outcome. The data contain two subgroups. We selected this example at random from all CDSR data containing two subgroups and at least two but no more than six studies per subgroup. This subset was used for illustration purposes.

We found that τ in the second subgroup is estimated to be zero when using the frequentist methods (Table 3.1). For both first subgroup and meta-regression τ is estimated to be larger than zero. Among the frequentist τ point estimates larger than zero, the REML point estimate is the largest, followed by the DL point estimate, and the smallest is PM. Estimating a common τ in the meta-regression leads to larger τ estimates relative to the second subgroup and smaller τ point estimates relative to the first subgroup. A comparison of the individual meta-analyses with meta-regression shows that the between-trial heterogeneity CI is shorter for the latter in all methods. Further, the estimated between-trial heterogeneity in the Bayesian case with a UP is associated with sizeable uncertainty in the individual meta-analyses.

TABLE 3.1: τ point estimates with 95% CIs for the example Goodwin et al. (2013) (1.4)

Type	Method	Subgroup	τ	95% CI
Individual	UP	1	0.53	[0; 6.24]
Individual	HNP	1	0.27	[0; 0.75]
Individual	REML	1	0.22	[0; 1.85]
Individual	PM	1	0.19	[0; 1.85]
Individual	DL	1	0.21	[0; 1.85]
Individual	UP	2	0.61	[0; 7.74]
Individual	HNP	2	0.24	[0; 0.77]
Individual	REML	2	0.00	[0; 2.36]
Individual	PM	2	0.00	[0; 2.36]
Individual	DL	2	0.00	[0; 2.36]
Regression	UP	-	0.28	[0; 1.03]
Regression	HNP	-	0.22	[0; 0.61]
Regression	REML	-	0.18	[0; 0.89]
Regression	PM	-	0.09	[0; 0.89]
Regression	DL	-	0.12	[0; 0.89]

This is suggested by the 95% τ CIs in Table 3.1. The meta-regression 95% τ CI for the UP is much shorter in comparison. The UP across methods has the widest CIs across all methods. In the case of a HNP, the 95% τ CI for the meta-regression is shorter as well. The μ 95% CIs when using the KnHa-CI are found to be shorter when using meta-regression across all methods and subgroups, as it can be seen in Figure 3.1. Using the Z-CI in Figure A.1 (Appendix) suggests that the μ 95% CIs in meta-regression is shorter for the first subgroup and wider for the second subgroup when compared to individual meta-analyses.

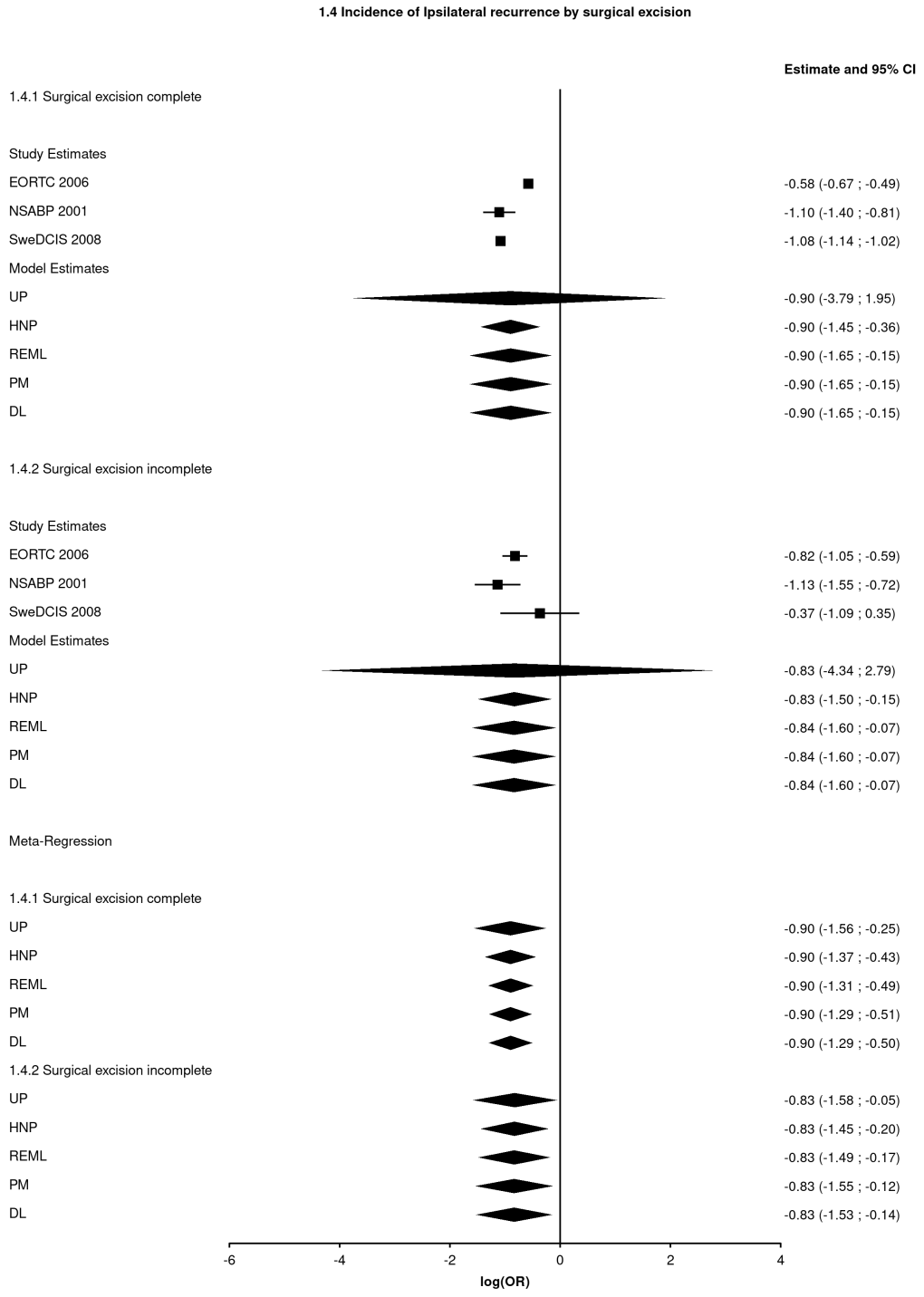

FIGURE 3.1: Example forest plot for first comparison and fourth outcome in Goodwin et al. (2013). Shown are the μ point estimates with 95% CIs for subgroup specific meta-analyses and meta-regression. This was done for the methods UP, HNP, REML, DL and PM. In the frequentist case, the KnHa CI was used.

3.4 Description of the empirical data

The data downloaded from the CDSR contained 6755 reviews. We only consider binary data that contain subgroups. This shrinks our number of reviews down to 3476. The downloaded data contains various variables. From the dataset, we used variables that define to which review, comparison, and subgroup each study belongs to. The number of reviews, comparisons, outcomes, subgroups and studies can be found in Table 3.2. Further, we used the binary data to compute the log-odds for each study. We only performed a descriptive analysis of the subgroups and did not further investigate the definition of the subgroups. Therefore, we are dependent on an accurate and consistent definition of the subgroups in the dataset.

TABLE 3.2: Summary of Cochrane database for subgroups subset with binary data

Type	Count
Reviews	3476
Comparisons	9614
Outcomes	32122
Subgroups	103528
Studies	306381

Figure 3.2 shows the relative share of number of studies per subgroups and number of studies per outcome. Based on our subset, the number of studies per outcome is at least two. The relative share decreases as the number of studies per outcome increases. The same is visible with the number of studies per subgroup. The most frequent number of studies per subgroup is one. Using the framework of meta-analysis should usually include at least two studies (Valentine et al., 2010). We keep the subgroups that only contain one study, since meta-regressions still can be used for a joint analysis. In the frequentist case, using only one study leads to a τ point estimate of zero. The respective CI for τ is then not computed. Only the use of an HNP in the Bayesian case leads to a CI for τ, which is based only on the prior.

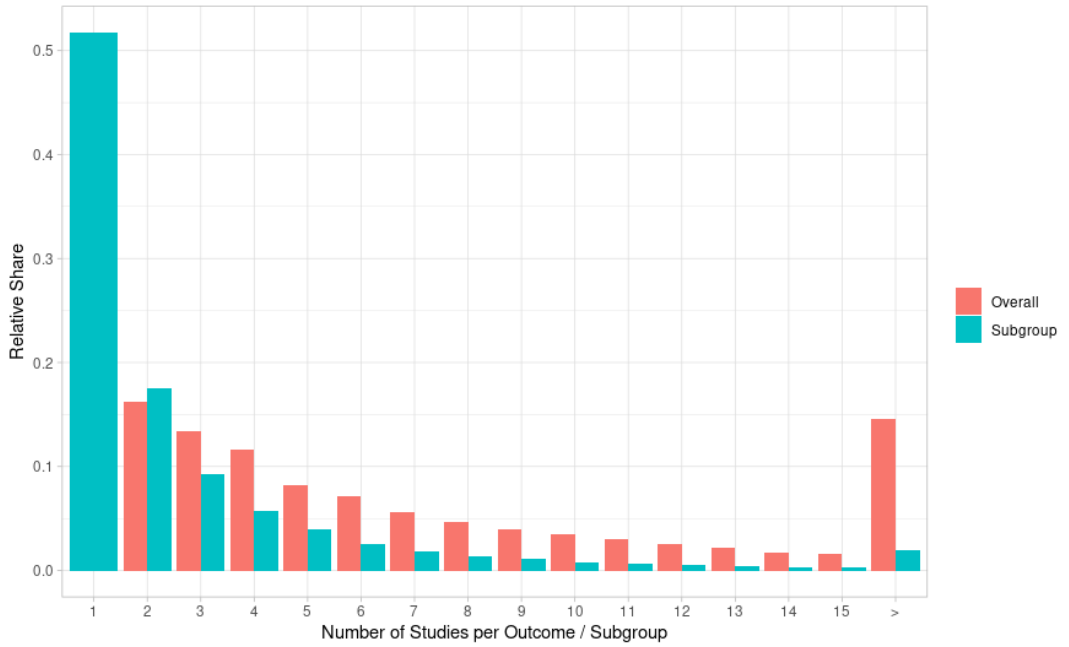

FIGURE 3.2: Relative share of study counts and subgroups per analysis. The y-axis values multiplied by 100 represent percentages.

The log-odds ratio estimates based on the binary data from the individual studies are shown in Figure 3.3 and approximately centered around zero. The distribution also appears to be symmetric. The right panel also shows the approximate variance (squared standard errors) of the log-odds ratio. If in either control or treatment group, all or no subjects experienced the predefined event, the standard errors are undefined. None, or all events in treatment and-or control group are commonly referred to as "single zero" or "double zero" studies in case of binary outcomes. In the case of "single" or "double zero" studies, so-called continuity correction is used to be able to calculate the approximate variances (Sweeting et al., 2004). This is done because Equation 2.2 cannot be calculated. To solve this problem, a continuity correction factor is added to each cell of the 2×2 contingency table. The continuity correction leads to large standard errors that are used in the meta-analysis/ -regression.

Removing the mentioned studies leads to a smoother representation in Figure 3.4. The distribution of the approximate variances then roughly follows a log-normal shaped distribution. The distribution of the log-odds ratios is smoother if the "single" and "double zero" studies are removed. We however do not investigate the influence of "single" and "double zero" studies

(Günhan et al., 2020). In the analysis below, we used all available data, including "single" and "double zero" studies.

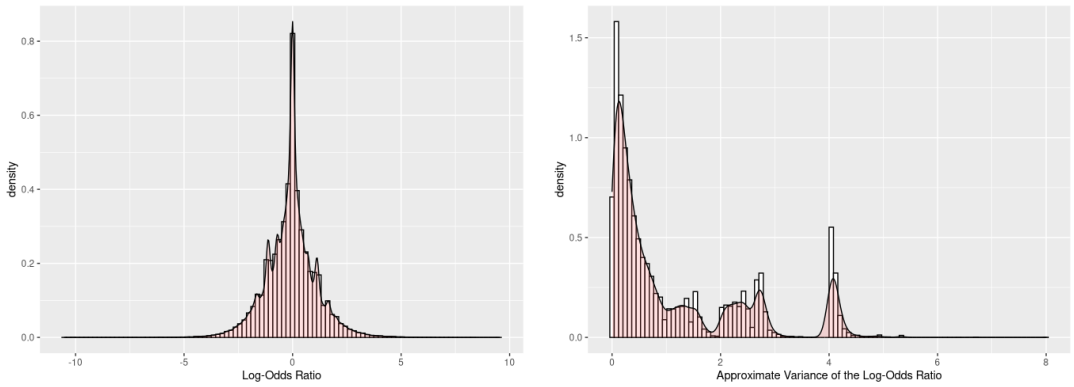

FIGURE 3.3: Log-Odds ratios and their approximate variances

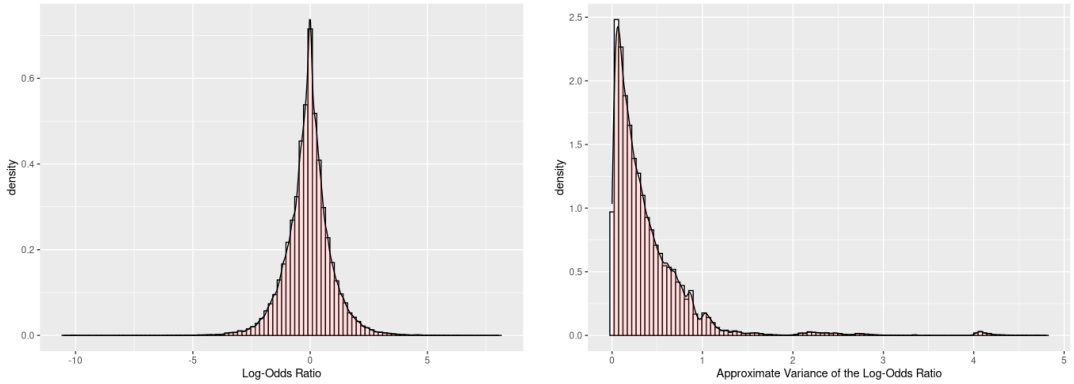

FIGURE 3.4: Log-Odds ratios and their approximate variances if "single" and "double zero" studies are removed

3.5 Estimation of meta-regressions and individual meta-analyses using the empirical data

Individual meta-analyses for the freqentist methods lead to a high number of τ point estimates equal to zero. When using meta-regression instead, we found that the share of zero point estimates for τ is reduced (Table 3.3). One has to keep in mind that the number of zero point estimates is also based on subgroups only containing one single study. This could be one explanation for the reduced number of zero point estimates. It is further noticeable that the number of zero point estimates in meta-analysis is larger than the relative share of subgroups containing only one study (Figure 3.2). In Table A.1 (Appendix) we only used subgroups that contained at least two studies. We found that the proportion of τ point estimates equal to zero is reduced when using meta-regression as well.

TABLE 3.3: Proportion of zero point estimates for τ. The values multiplied by 100 represent percentages.

Type	UP	HNP	REML	PM	DL
Meta-Analyses	0	0	0.76	0.78	0.78
Meta-Regression	0	0	0.41	0.46	0.46

As a further measure, we looked at the ratio of the mean of 95% CI of μ of individual meta-analysis and meta-regression. This is shown in Table 3.4. A ratio larger than one means that the meta-analysis CI is larger than the CI of the corresponding meta-regression. We found that the mean μ CI is wider in most cases. In the Bayesian case, we found that the effect of meta-regression is larger in case of a UP. A larger effect in this case means that the precision gain is greater for UP compared to HNP. This is as expected since the usage of an informative HNP leads to shorter CIs even for small number of studies per subgroup. In the frequenist methods, we can see that all estimations for τ lead to similar results. The comparison of the different CIs for μ shows that, only for the Z-CI, there is a slight advantage in the mean μ-CI lengths for the meta-analysis. All other CIs indicate that, on average, meta-regression shrinks the μ-CI lengths compared to meta-analysis despite fewer zero point estimates for τ. In Table A.2 (Appendix) we again used only subgroups that contained at least 2 studies. We found analogue results for the CI length ratios compared to the inclusion of all subgroups.

TABLE 3.4: Mean of μ 95% CI length ratio. The ratio is computed as the CI length of individual meta-analyses divided by the CI length of meta-regression. A ratio of < 1 indicates that the CI of individual meta-analyses are shorter and vice versa.

	UP	HNP	REML	PM	DL
shortest	2.22	1.08	.	.	.
Adhoc	.	.	1.75	1.76	1.76
KnHa	.	.	3.10	3.19	3.28
t	.	.	1.77	1.78	1.77
Z	.	.	0.99	0.99	0.99

Furthermore, we look at the mean difference of the τ point estimates resulting from meta-analysis and meta-regression (Table 3.5). A positive difference means that the point estimates from the meta-regression are smaller compared to the individual meta-analyses on average. The mean difference of the Bayesian point estimates tends to be larger than zero, whereas the frequentist mean difference is negative, possibly due to fewer zero point estimates

for τ. We would therefore conclude that the Bayesian meta-regression would result in on average smaller τ point estimates in the empirical data set. On the other hand, the frequentist τ is larger on average in the meta-regression.

TABLE 3.5: Mean difference of τ meta-analysis and τ meta-regression. A difference < 0 indicates that on average, the point estimate of τ is larger when using meta-regression

	UP	HNP	REML	PM	DL
$\tau_{Ind} - \tau_{Reg}$	0.21	0.02	-0.07	-0.07	-0.06

More than 95% of zeros in the difference of τ for the frequentist methods are based on the zero point estimates for the heterogeneity. The number of zero point estimates is lower in meta-regression by around 5 % if all data are considered. When the data are split by subgroup size (using the ICEMAN criterion based on subgroup size introduced in Chapter 3.6) the share of zero point estimates shrinks if larger "smallest subgroups" are considered. The on average smaller point estimate of τ in Table 3.5 can be possibly explained by the smaller amount of zero point estimates in the frequentist meta-regression.

In Figure 3.5 we show the distribution of the τ point estimates across the used methods in our analysis. As mentioned beforehand, the number of zero point estimates is large. The model using a UP is only computable if the number of studies is larger than three. Therefore, the distributions shown here are not directly comparable.

In contrast to the number of zero point estimates, the point estimates from using a HNP are highly influenced by the prior median at around 0.3. If there is only one study per subgroup, τ is close to the prior median.

Figure 3.6 shows the empirical cumulative distribution function (ECDF) of the τ point estimates. We show this figure to identify point masses, visible as jumps in the ECDF. Jumps in the ECDF show up in the case of a UP in individual meta-analyses. We did not find an explanation for this jump in the ECDF. Further jumps in the ECDF can be found when looking at the Frequentist methods. Here the inflated number of zero point estimates is clearly visible.

When looking at the log transformed τ point estimates, note that all zero point estimates are no longer defined and are therefore not shown. In Figure A.2 (Appendix) only the non-zero point estimates of τ are displayed. Using the UP on the empirical dataset lead to a peak in the log-transformed τ point

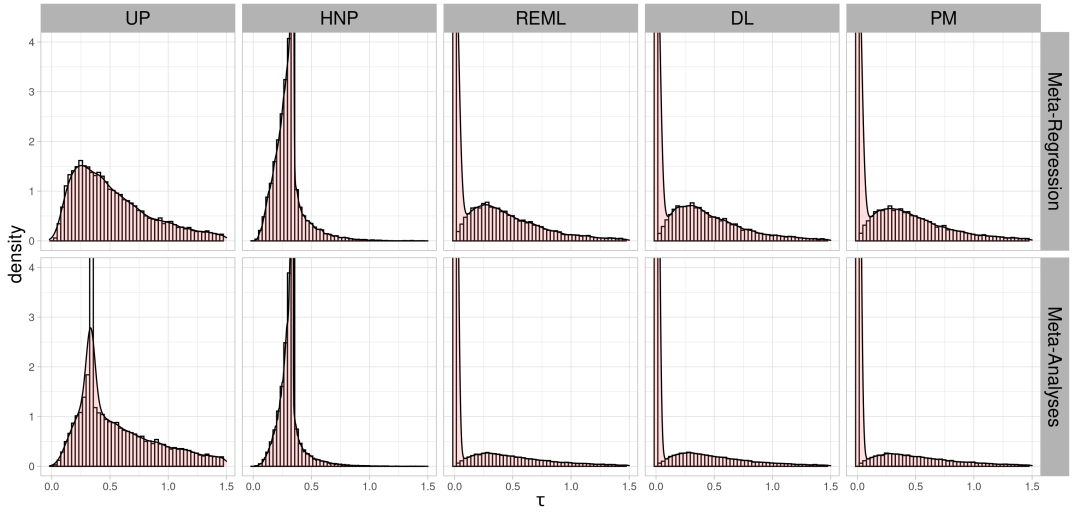

FIGURE 3.5: Histogram and kernel density of point estimates for τ in meta-regression and meta-analysis. The number of zero point estimates in this figure is not representative, since we restricted the y-axis.

estimates. This peak does not seem to be linked to "single" or "double zero" studies. If the data are split analogous to Table 3.6, we found that this effect is only visible if the smallest subgroups contained one or two studies. Using meta-regression does not leads to the peak shown in Figure A.2. In both meta-analysis and meta-regression, we observe a bimodal distribution for the REML point estimate. We suspect this to be a convergence issue.

Many subgroup analyses also include an overall analysis, but not all. Sometimes even subgroup analyses without a joint analysis are considered. The previous figures refer to all subgroups. Since it can be argued that in the cases with joint analysis the assumption of a common τ should be uncontroversial, these are again presented separately in the appendix. All analyses based on data where overall analyses were conducted are shown in Appendix A.4.3. Overall analysis means that additional to individual subgroup meta-analyses, also one large meta-analysis with pooled data from all subgroups is considered. Meta-analyses estimating a pooled point estimate with data of all subgroups consider the assumption of equal τ across subgroups as reasonable. The results are very similar. However, the ratio of the of μ CI lengths is smaller but still larger than 1 (Table: A.5). The mean differences of τ individual meta-analyses and meta-regression also show the same direction (Table: A.6). The overall share of zero point estimates is reduced when using meta-regression as well (Table: A.4).

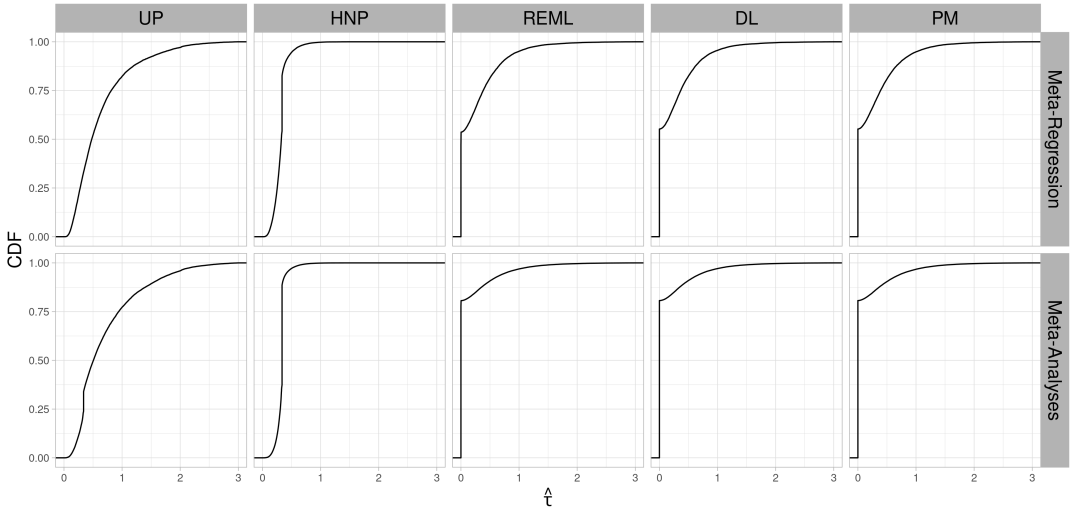

FIGURE 3.6: Empirical cumulative distribution function of τ point estimates in meta-regression and individual meta-analyses

3.6 Subgroup of the empirical dataset

The analysis of predefined subgroups is generally used to identify effect modifiers. To evaluate the credibility of including effect modifiers e.g., regarding a subgroup effect, or interaction with a continuous variable, Schandelmaier et al. (2020) published the Instrument for assessing the Credibility of Effect Modification Analyses (ICEMAN). It can be used as a guidance on whether to trust the findings of a certain modification or to treat it with caution. In their paper, they define different criteria to assess effect modifiers. One criterion is based on the size of the smallest subgroup in the meta-regression. This is shown below in Table 3.6. They define four groups, from very small to large. In case of subgroups, they define the credibility based on the number of trials in the smallest subgroup.

TABLE 3.6: ICEMAN Criterion: Is the number of trials large?

	Number of trials:	
Group	in the smallest subgroup	in continuous meta-regression
Very small	1–2	5 or less
Rather small	3-4	6-10
Rather large	5-9	11 to 15
Large	10 or more	more than 15

Simulation

4.1 Setup

In this chapter, we want to investigate individual meta-analyses compared to meta-regression in the presence of subgroups. In contrast to Chapter 3 in which real-life data were analyzed, this chapter investigates the behavior of individual meta-analyses and meta-regression in a structured way and under control of relevant parameters. We will consider the following:

- Comparison of individual meta-analyses and meta-regression as well as comparing different methods
- Effects of violation of common τ assumption in meta-regression
- In the Bayesian case: check whether a prior equal to the true distribution of τ leads to the nominal coverage probability
- Demonstrate that μ is unbiased in our simulations

To simulate the data, we used the following framework. For the simulation, we set the number of subgroups to 3 with the number of individual trials k per subgroup in $\{2, 5, 10\}$. The number of subgroups represents a small, rather large and large number of trials according to the ICEMAN criterion (Table 3.6).

We define true values for μ and τ and then generate y_i. s_i will be sampled beforehand from a uniform distribution. Sampling s_i from a uniform distribution with parameters 0.2 and 1 represents roughly sample sizes between

16 and 400 (Röver et al., 2019).

$$s_i \sim U(0.2, 1) \tag{4.1}$$
$$\theta_i \sim N(\mu, \tau^2) \tag{4.2}$$
$$y_i \sim N(\theta_i, s_i^2) \tag{4.3}$$

We will look at four simulation scenarios. In Scenario I, we used fixed values for τ:

$$\tau \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 1, 2\} \tag{4.4}$$

As demonstrated in simulation Scenario IV, the choice of μ does not influence the other scenarios aside from the location of μ, since the point estimate for μ is unbiased. This was shown analytically in, e.g., Jackson et al. (2014). Therefore, we set μ equal to zero in all scenarios except the fourth one.

Scenario II aims to assess the robustness of the meta-regression. We simulate a scenario where the meta-regression assumption of equal τ across subgroups is violated. For that, we set τ for the subgroups to:

$$(\tau_1 / j, \tau_2, \tau_3 * j) \text{ with } j \in 1, 1.1, 1.2, \ldots, 2 \tag{4.5}$$

where j is the factor that shrinks τ_1 and inflates τ_3. We look at the situation where $\tau_1 = \tau_2 = \tau_3 = 0.3$. τ_2 is therefore always set to the value of 0.3. We decided to set the factor j to be equidistant in steps of 0.1 from 1 to 2.

In addition to fixed values for τ, we can also sample τ from a distribution. In Scenario III, all subgroups share a common τ. We would assume that if the prior for τ matches the sampling distribution, the quantile of the true values for τ (PIT values) should follow roughly a uniform distribution. If, in addition, μ were drawn from a distribution matching its prior, the resulting distribution of the quantile of the true value for τ would follow a uniform distribution. We set $\mu = 0$ and thus expect only roughly a uniform distribution. In Scenario III, we set the true distribution from τ to a half-normal distribution with scale parameter 0.5, which is the same as the distribution that is assumed for τ in the HNP. Both Scenario III and IV are used to show that the simulation works properly.

In simulation Scenario IV, we want to demonstrate that the choice of μ does

not influence our results. For that, we set the value of μ in the three subgroups equal to:

$$\mu_1, \mu_2, \mu_3 = 0, 0.2, 0.5 \tag{4.6}$$

Across all simulations, we use multiple tools to investigate the properties of the investigated methods. To assess whether the parameters tend to systematically vary from the true parameter values, we use the bias and the mean squared error (MSE). The bias of a variable is defined as follows:

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta \tag{4.7}$$

$\hat{\theta}$ represents the estimated values and θ the true value. The MSE on the other hand is calculated by the expectation of the squared difference of estimated and true values:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{4.8}$$

The coverage probability will also be used. The coverage probability of a method is defined as the proportion of times a CI covers the true parameter. Ideally, the coverage probability will attain the predefined level, e.g., 95%. If it covers a higher percentage than the predefined one, it is considered conservative. If it does not, it is considered liberal.

To compare the results from individual meta-analysis and meta-regression, we will use ratios of parameters, e.g., ratio of coverage probabilities or ratio of CI lengths. We will always divide the parameter point estimate from meta-analysis by the parameter from meta-regression. Therefore, a ratio larger than one indicates that the parameter of meta-analysis is larger in comparison to the meta-regression parameter and vice versa. A ratio of one would indicate that both parameters are equal.

The methods used in the simulations for estimation of τ in the frequentist case are REML, PM and DL. For the Bayesian methods, we use either a UP or a HNP for τ (combined with a UP for μ). The point estimate in the Bayesian case is the posterior median for both τ and μ. For the μ in the frequentist case, we use Z-, t-, KnHa- and the Adhoc-CI. In the Bayesian case, we use the shortest CI. The CIs were computed via implemented functions in the packages `bayesmeta, metafor` (Röver, 2020; Viechtbauer, 2010).

4.2 Scenario I: Effects of varying heterogeneity

Figure 4.1 shows an example density plot of the parameter point estimates for μ across all methods, where the true value is indicated by a dashed line. Jackson et al. (2014) derived that the frequentist point estimate for μ is unbiased. Our simulation shows the expected behavior with μ fluctuating around zero. A similar behavior can be found for the Bayesian point estimates. We found further that with increased τ, the distribution of μ shows a larger variance. This can be seen in Figures A.6 and A.7 (Appendix).

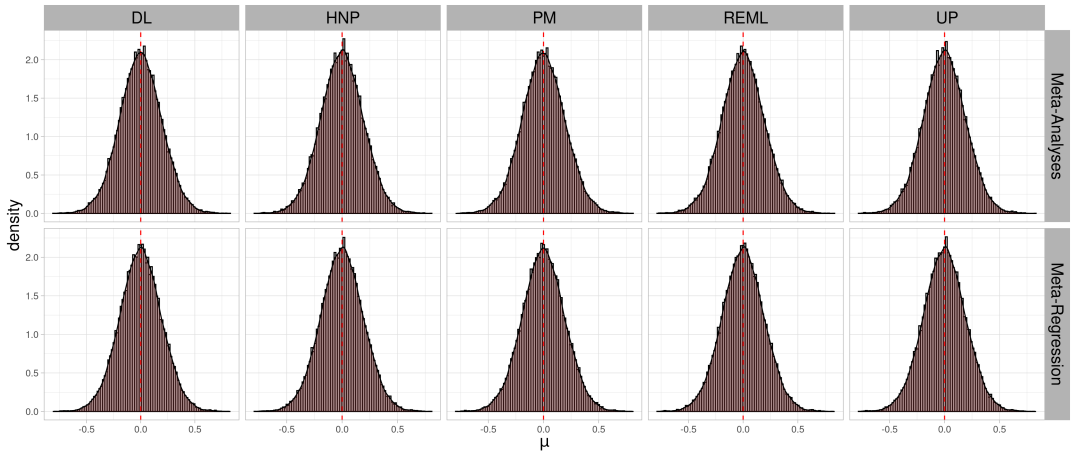

FIGURE 4.1: Distribution of point estimates of the overall effect μ for $\tau = 0.3$, a sample size per subgroup of k =10 and a true value of $\mu = 0$

In Figure 4.2 we plot the bias of τ against the amount of between-trial heterogeneity that we introduced. We found that for small amounts of heterogeneity the frequentist methods tend to be positively biased, whereas if we increase τ the bias becomes negative. A positive bias for very small values of τ can be expected, since we do not allow negative τ point estimates.

The same holds for the Bayesian case with a HNP. Here we can see how the choice of an informative prior influences the posterior. The bias is very low when it is near to the prior median (around 0.3). If the true τ is smaller than the prior median, a positive bias is introduced. A true τ that is larger than the prior median, however, results in a negative bias. This can be seen as an advantage since the point estimate is corrected towards the expectation being the prior median. Using meta-regression instead of individual meta-analyses reduces this effect. For the Bayesian case with a UP, we found that it is positively biased for all choices of τ. The results for individual meta-analyses are in general in agreement with the findings in Friede et al. (2017a).

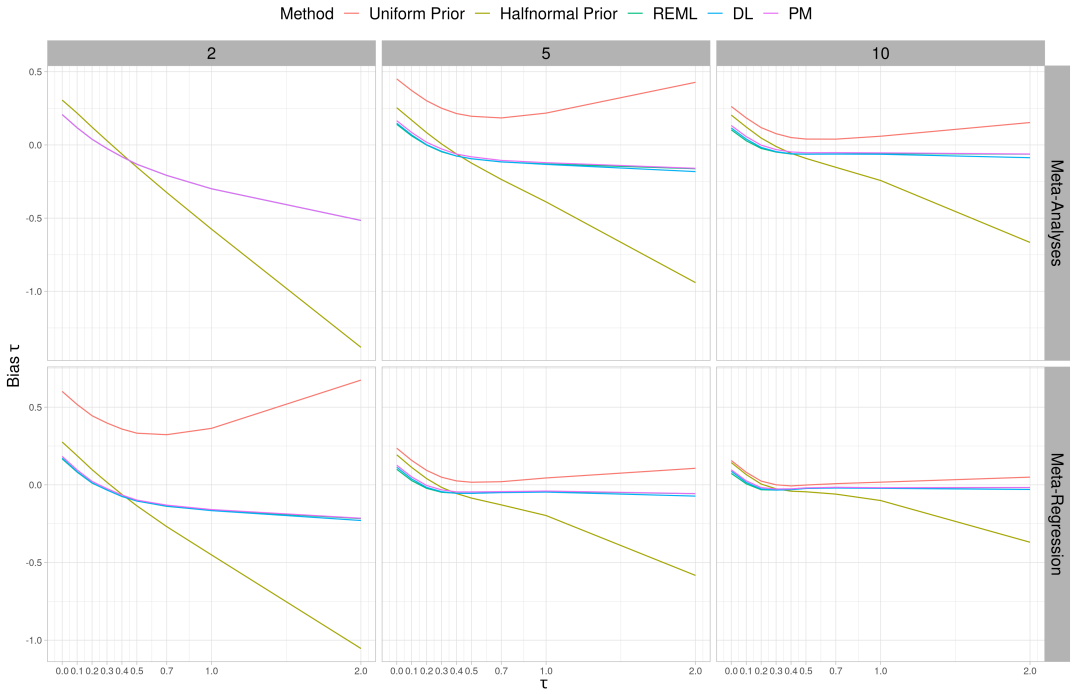

FIGURE 4.2: Bias in estimating τ for varying numbers of studies per subgroup and varying true values of τ for both individual meta-analyses and meta-regression

It is further evident that a larger sample size reduces the bias. The sample size per subgroup is indicated by the gray shaded bar. When comparing meta-analysis and meta-regression, we found that the larger sample size of the meta-regression reduces the bias as well. In this scenario, the usage of meta-regression effectively means a larger sample size for the estimation of τ, since we can use all subgroups at once.

In Figure 4.3, a box plot of the estimated value minus the true value for meta-regression and individual meta-analysis is shown. We found that the variance of this value increases as we increase τ. In the case of individual meta-analysis, the variability is over all choices of τ larger compared to meta-regression, but also increases for increased τ.

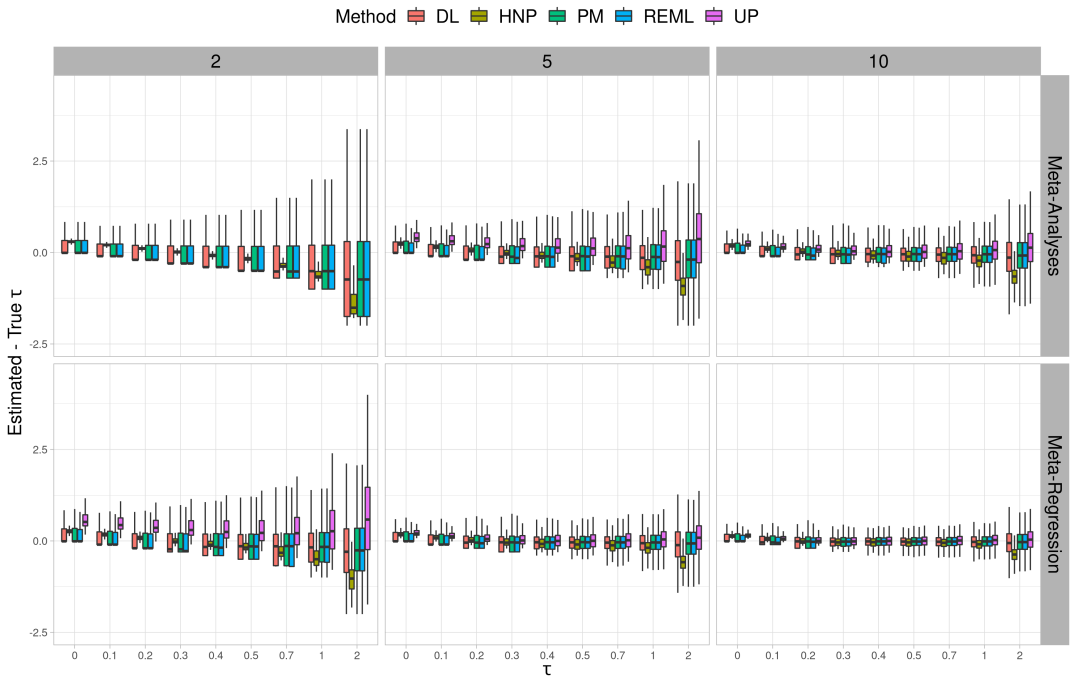

FIGURE 4.3: Box plot: Difference of τ point estimate and true τ for varying sample size per subgroup, and varying true τ in individual meta-analyses and meta-regression. Outliers were removed in this figure. Outliers are values that are out of the interquartile range (from the 25 or 75 percentile) times 1.5.

In Figure 4.4 we show the distribution of τ point estimates for meta-analysis and sample size of 10 per subgroup. The method is shown by the gray shaded bar at the top. The sample size and the value of τ is shown by the gray shaded bars on the right-hand side. The true value of τ is indicated by the red dashed line. The frequentist methods, analogous to the findings in the Cochrane data, show an inflation of zero point estimates for τ. The inflation is larger for smaller sample sizes and for smaller true values of τ. We found that using meta-regression instead of individual meta-analyses leads to a smaller amount of zero point estimates for equal sample sizes and the same choice of τ. The point estimates using a UP tend to follow a skewed distribution for small values of the true τ parameter. The same holds for the HNP. In the case of the HNP, we can also see the influence of the prior on the estimation of τ. The mean of the distribution of estimated τ better represents the true value of τ, the closer we are to the prior median.

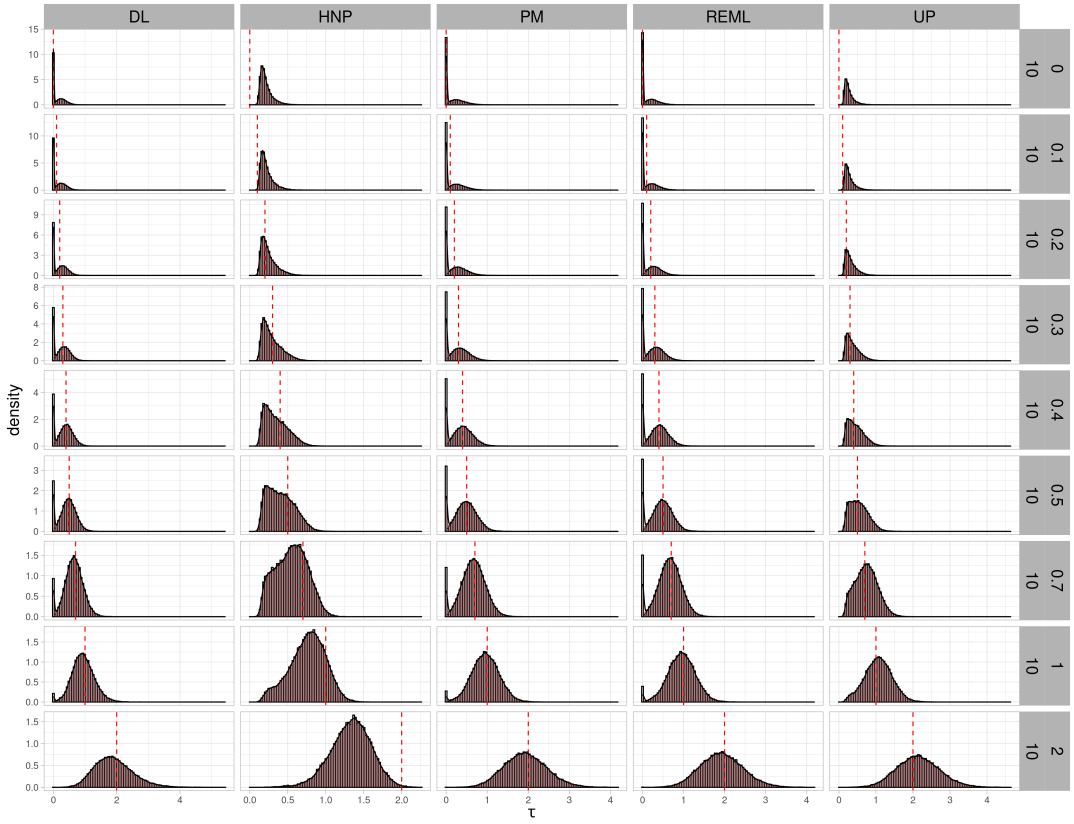

FIGURE 4.4: Distribution of τ point estimates for varying true values of τ in individual meta-analyses for 10 studies per subgroup. Note that the x-axis varies across methods. The gray bar on the right indicates the sample size per subgroup and the respective true τ

The mean difference of τ (meta-analysis and meta-regression) is shown in Figure 4.5. We can see that the mean difference for the frequentist methods is dependent on the amount of between-trial heterogeneity and on the number of studies included in the analysis. For small τ, the difference is positive, whereas for larger τ, the difference becomes negative. We can see the same behavior for the HNP. Only in the case of the UP, the difference is always positive, indicating that the meta-analysis point estimate is on average larger for all choices of τ. In case of the UP, meta-regression on average adjust τ downward. For all other methods, the adjustment is dependent on the true τ. When the true τ is small τ on average is adjusted downward, whereas for large τ it is adjusted upward. In the frequentist case, this is despite the fact that less τ are estimated to be zero in meta-regression.

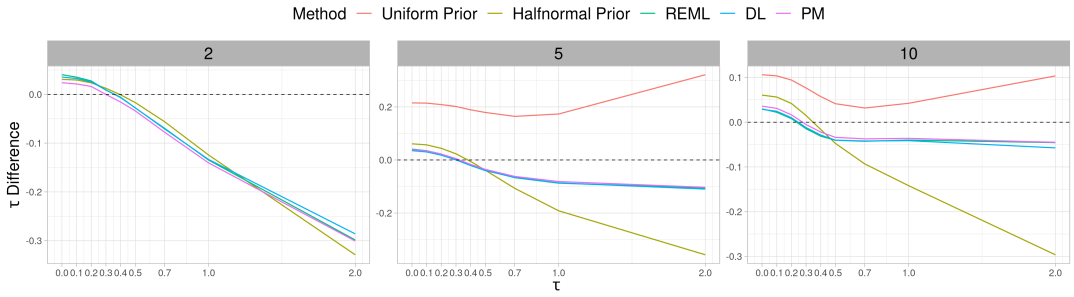

FIGURE 4.5: Mean difference of τ point estimates for varying sample size per subgroup. The difference is defined as individual meta-analysis point estimate minus meta-regression point estimate.

When using the `metafor` package, the `confint.rma.uni()` uses profile likelihood to compute the CI for τ. In case of numerical problems the bounds are arbitrarily set to [0,100] or, for the upper bound, ten times the point estimate of τ, whichever is greater. Therefore, the CI length for τ could be biased. We found in Figure 4.6 that the mean CI lengths for τ are reduced for meta-regression in comparison to separate analyses with an UP, REML, DL and PM. The change in CI lengths with the HNP are small compared to the other methods. For an increase in τ, the CIs are estimated to be wider. This is to be expected since the heterogeneity is increased.

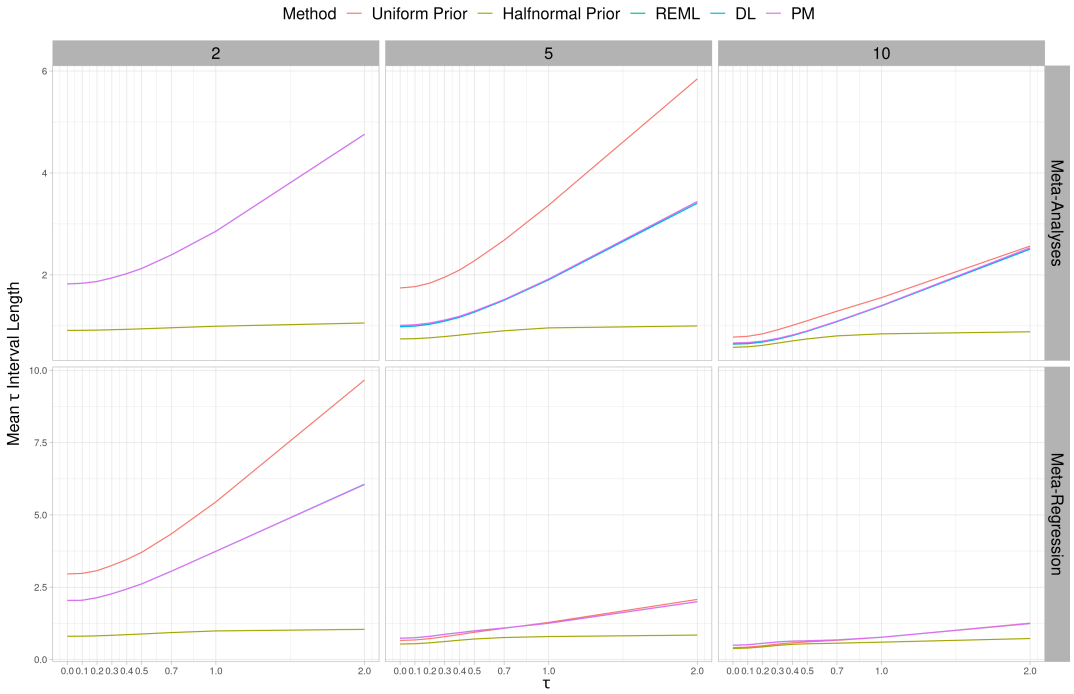

FIGURE 4.6: Mean τ 95% CI length for varying sample sizes per subgroup and varying τ in meta-regression and individual meta-analyses. The frequenist CI is based on profile likelihood and therefore does not differ across methods. In the frequentist case, the CI lengths are from the Z-CI.

In Figure 4.7 we display the coverage probability for individual meta-analyses. The gray shaded bars on the right show the CI that is used in each case. We found that the CI based on the normal approximation (Z) performs the worst. Here the aimed 95% are only covered for very small τ. The KnHa-CI also was not able to adhere to the 95%. However, we can observe that the coverage probability does not fall much below 93%. Both t- and Adhoc-CI tend to be conservative for small τ and for larger τ then cover roughly the 95%. In the case of the shortest CI, we found that the UP is more conservative with a lower sample size, but it overall covered the aimed 95% coverage probability. The effect of the HNP is clearly visible, as the CI is liberal for larger τ and conservative for smaller τ. One has to keep in mind that the difference between the frequentist τ point estimates are small compared to the differences of the CI computation (Z, t, KnHa, Adhoc).

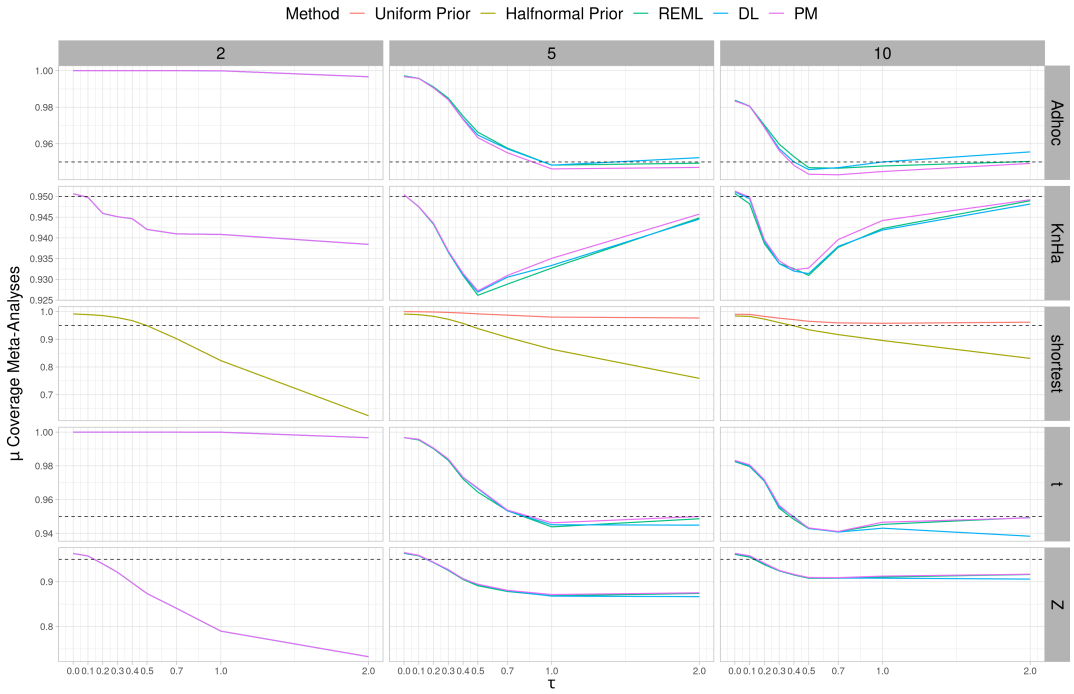

FIGURE 4.7: CI coverage probability of the true μ in individual meta-analyses for varying sample size per subgroup and true τ. The shown y-axis values multiplied by 100 represent percentages.

Figure 4.8 shows that the overall coverage probability is better across all methods when using meta-regression. In the cases where the 95% were not attained, the coverage probability is higher than before, e.g., Z, KnHa, HNP. For the CIs that tended to be conservative before, we found that the coverage probability is less conservative for the small sample sizes and small τ.

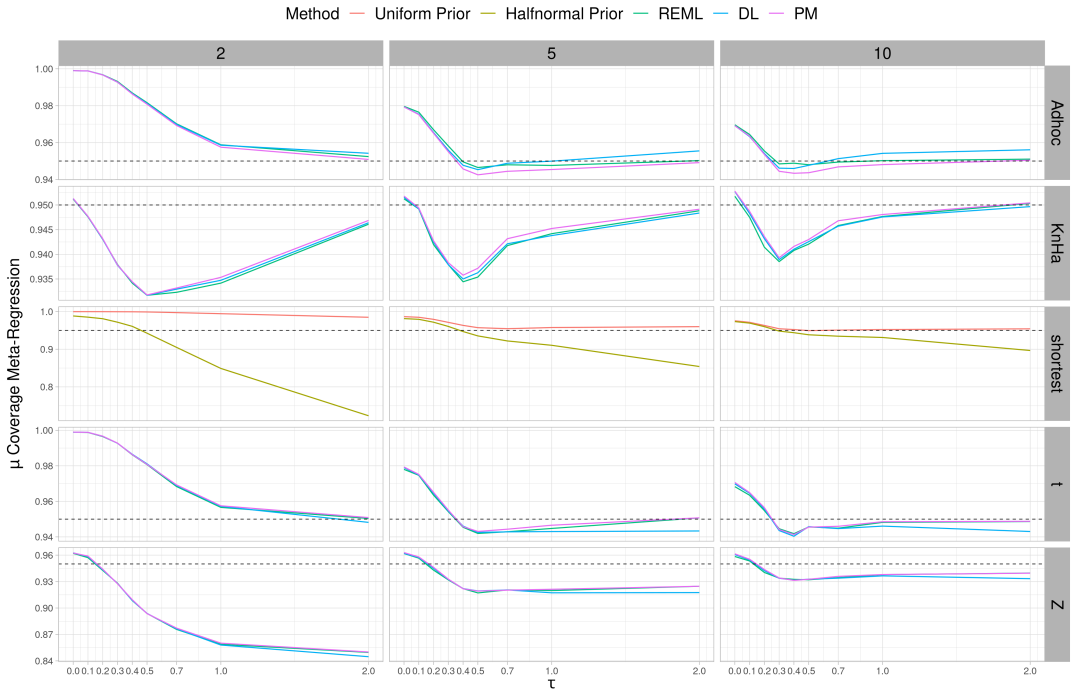

FIGURE 4.8: CI coverage probability of the true μ in meta-regression for varying sample size per subgroup and true τ. The shown y-axis values multiplied by 100 represent percentages.

We were also able to show that the μ CIs shrunk on average when we used meta-regression in comparison to individual meta-analyses. This is shown in Figure 4.9. If the ratio is larger than one (indicated by the dashed line) then the meta-analysis CI is wider. We found that for Adhoc-, KnHa-, t- and the shortest-CI with UP that the meta-regression shrinks the CIs for all sample sizes and choices of τ. However, we found that the shrinkage for small sample sizes was more evident. For the shortest-CI with a HNP and the Z-CI, we found that the CIS for large τ are actually wider in the case of meta-regression. This is probably due to the coverage of individual meta-analysis CIs for small τ being larger than the aimed level of 95% and for large τ being smaller. The coverage of meta-regression CIs then should be closer to the 95% level. We also found that the larger the sample size, the later the ratio drops below one.

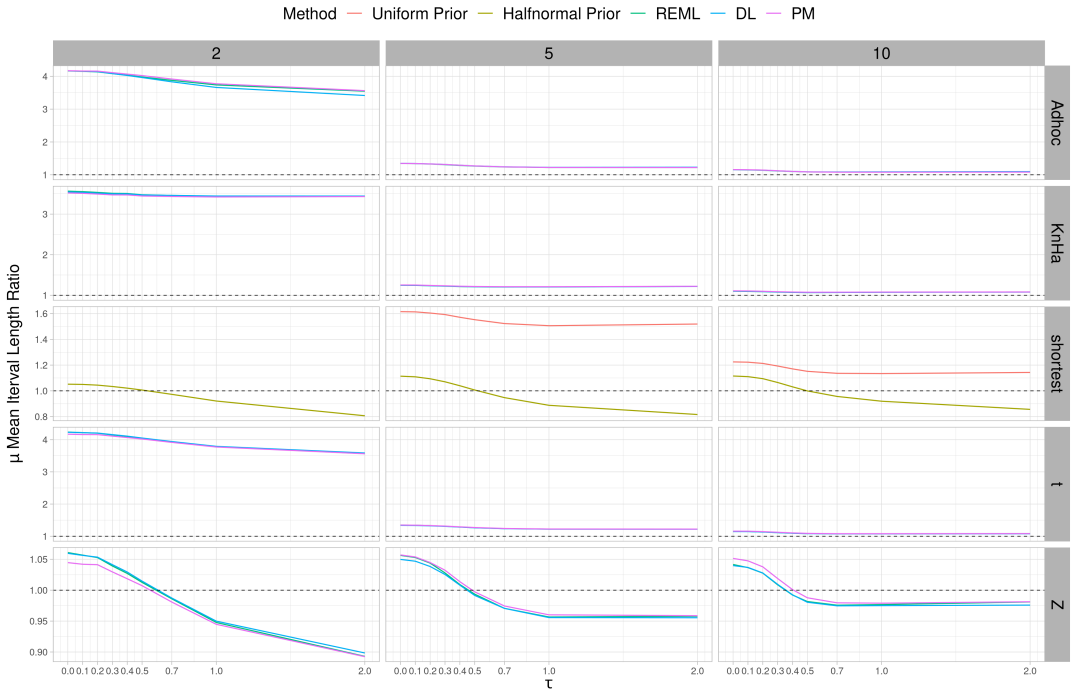

FIGURE 4.9: 95% μ CI length ratio meta-analysis and meta-regression for varying sample size per subgroup, true τ and CIs. A ratio larger than 1 indicates that the CIs of individual meta-analyses are wider on average compared to meta-regression.

In Figure 4.10 we show the ratio of the MSE of μ for meta-analysis and meta-regression. Except for the case with the HNP, we can show that the use of meta-regression lowers the MSE over all choices of τ. In the case of the HNP, we found that for τ close to the prior mean, the MSE appears to be lower for meta-analysis. For the frequentist methods, we can show that the MSE becomes smaller for larger τ. However, the differences in MSE are very small. This is to be expected since the point estimate of μ is only very indirectly influenced. Different estimation of τ only leads to slight changes in the study weighs. More relevant are the effect on the precision, e.g., CI length and coverage.

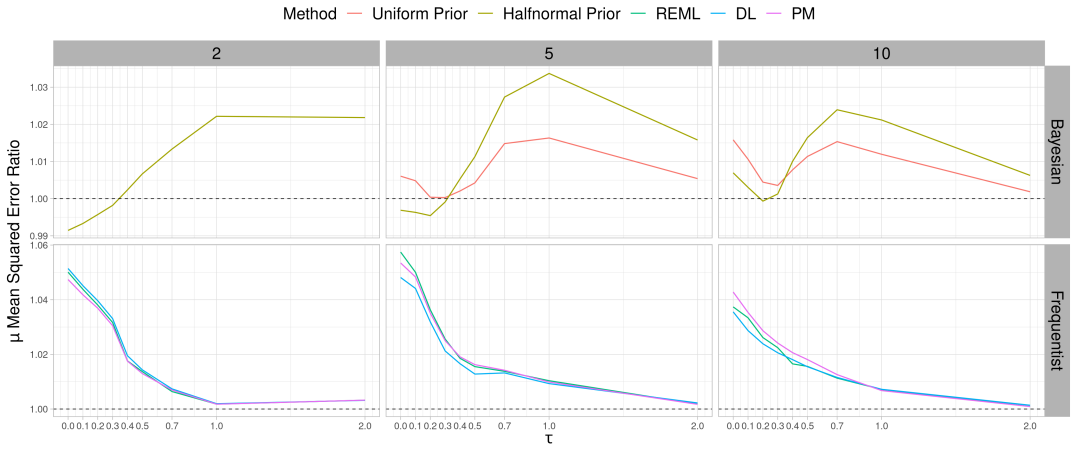

FIGURE 4.10: Ratio of μ MSE for varying sample size per subgroup and true τ in individual meta-analyses and meta-regression. A ratio of larger than 1 indicates that the MSE for individual meta-analyses is larger compared to meta-regression.

4.3 Scenario II: Effects of assumption violation

In Scenario II, we want to investigate the robustness of the meta-regression approach by generating data that does not meet the assumptions made in the meta-regression. In Figure 4.11 we plot the coverage probability of μ against the factor we used to vary τ across the three subgroups for individual meta-analyses. The assumptions made by the individual meta-analysis are not violated. One τ is always 0.3 in the simulations, where the other two either shrunk or are increased by the factor on the x-axis. For small factors j , we would expect the simulation to produce results similar to those in Scenario I. We observe a fairly constant coverage probability across most methods. Taking a look at the shortest-CI in the case of a HNP, we can see the effect of the prior. The coverage probability for μ tends to shrink for an increase in the factor of τ. Since we plot the coverage of all τ in one plot, we have to recall what we found in Scenario I. We found that a τ smaller than the posterior median leads to a conservative CI, whereas a τ larger than the posterior mean leads to a liberal CI. We found that the coverage deteriorates for a larger difference in τ across subgroups. For the other methods we found similar coverage probabilities as in Scenario I with a τ of 0.3.

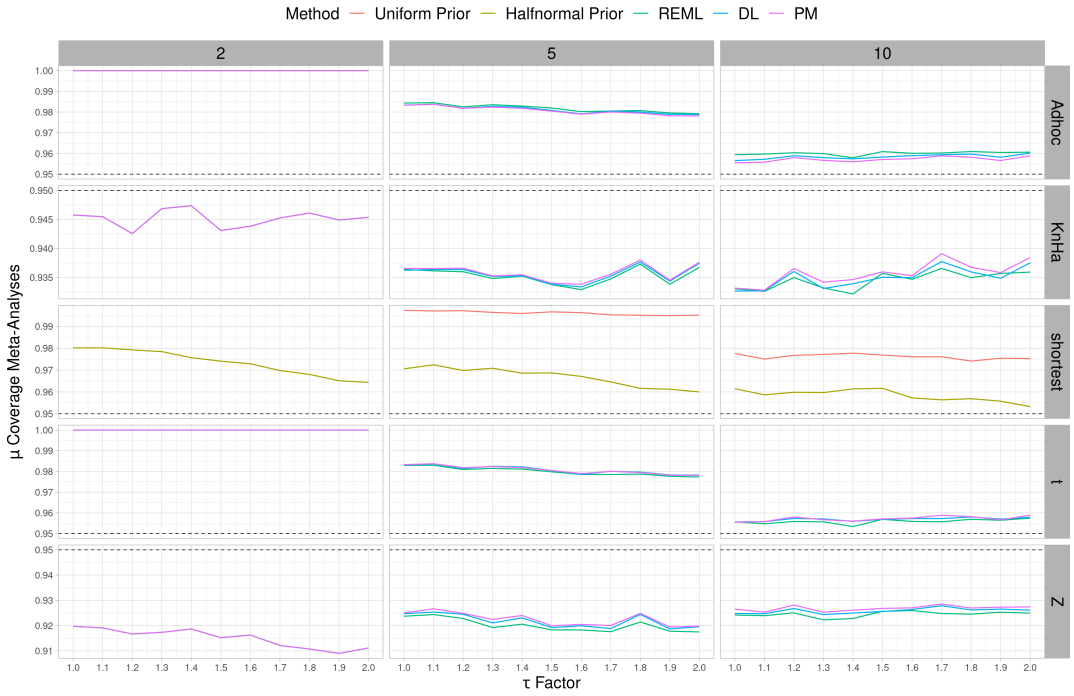

FIGURE 4.11: Coverage of CI for μ in individual meta-analyses for varying sample sizes, factor of τ variation. The shown y-axis values multiplied by 100 represent percentages.

The results of the meta-regression are shown in Figure 4.12. The coverage probability seems to show a slight downwards slope. This behavior is to be expected since the assumptions for large τ are severely violated. We, however, found that the decrease in coverage probability in comparison to the severity of the violation is very moderate. Moreover, we found that using meta-regression leads to coverage probabilities of conservative CIs closer to the 95% level. In the case of liberal CIs, e.g., Z, we found that for a small variation of τ across subgroups the meta-regression still leads to better coverage probabilities.

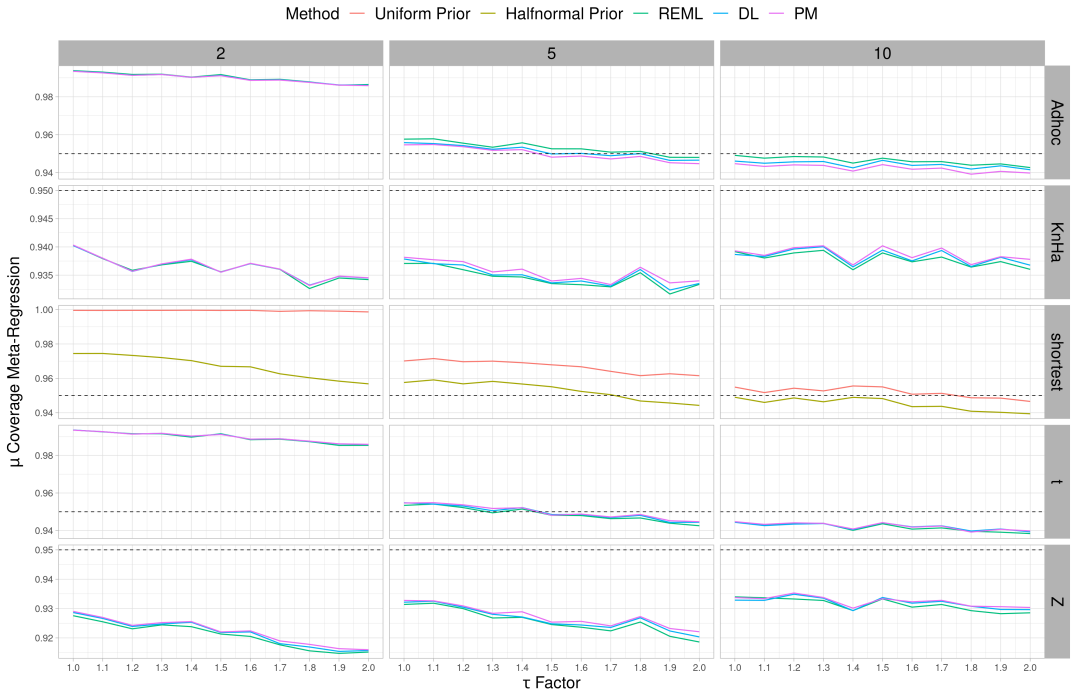

FIGURE 4.12: μ CI Coverage in meta-regression for varying sample sizes and factor of τ variation. The shown y-axis values multiplied by 100 represent percentages.

In Figure 4.13 we plotted the ratio of the coverage probabilities for individual meta-analysis and meta-regression. We found that the ratio of the coverages shows an increase if the factor of τ is increased. However, if we look at the scale of the y-axis, we see that this effect is fairly small. The Z-CI, for individual meta-analysis, has a worse coverage for all used factors of τ and sample sizes. We would conclude, for our scenario, that using meta-regression with a Z-CI is beneficial in all cases. In the case of the KnHa-CI, we found that for large sample sizes, the coverage of the meta-regression is larger compared to individual meta-analysis. The same can be found for a sample size of five per subgroup and a factor of τ up to 1.7. In the case of conservative CIs, a ratio of larger than one does not necessarily indicate an advantage. We would expect for conservative CIs that the respective CIs are wider as well and, therefore, indicate less precision.

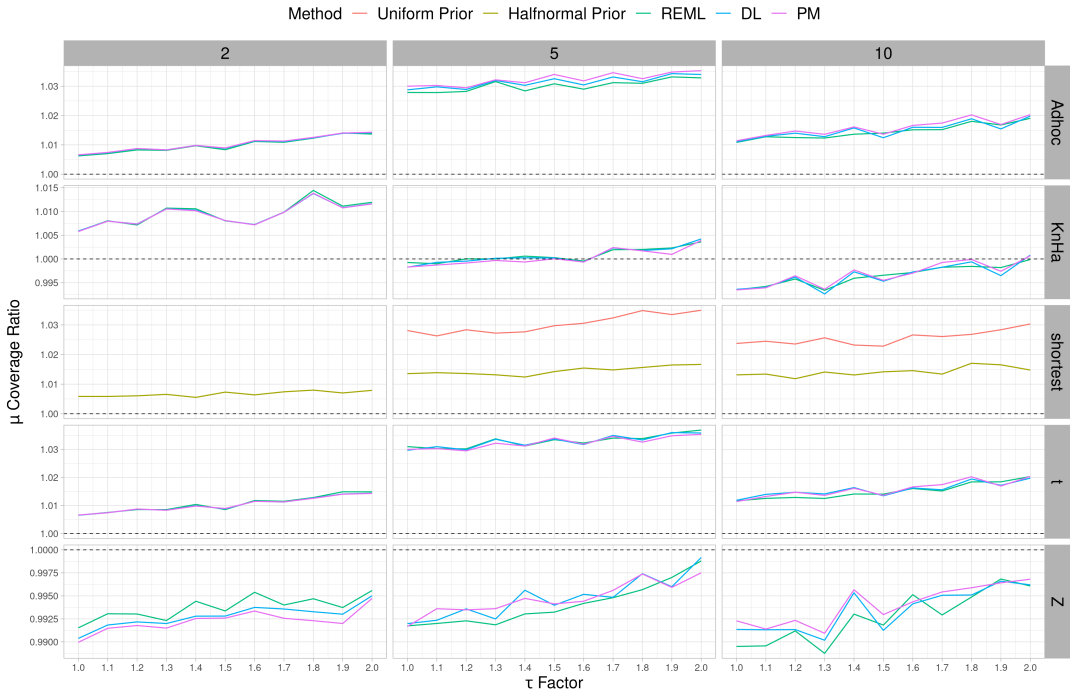

FIGURE 4.13: Coverage ratio of μ meta-analysis and meta-regression for varying τ per subgroup. A ratio of < 1 indicates that the coverage probability of individual meta-analyses is smaller compared to meta-regression and viceversa.

The ratio of the coverage probability of τ is shown in Figure 4.14. We found that in the case of individual meta-analyses, the CI for τ covers the targeted 95%. In the case of meta-regression, we observed a clear trend that the coverage for an increased difference in τ tends to be less and less good. This is to be expected since the assumptions were violated.

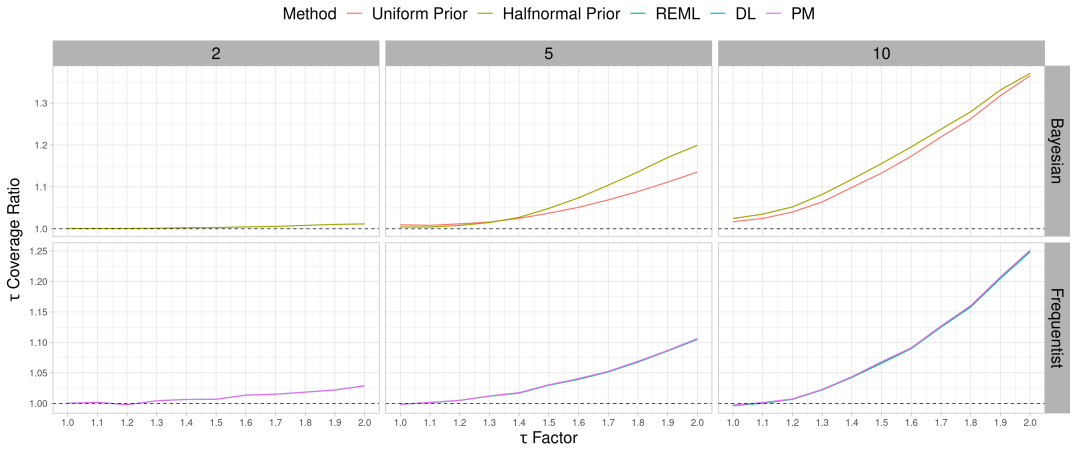

FIGURE 4.14: Coverage ratio of τ for varying sample sizes, factor of τ variation and CIs in meta-analysis and meta-regression. A ratio of < 1 indicates that the coverage probability of individual meta-analyses is smaller compared to meta-regression and vice versa.

Next, we consider the ratio of the average CI lengths for individual meta-analysis and meta-regression, as shown in Figure 4.15. We found that the use of meta-regression in nearly all cases leads to shorter CIs on average. We only found that for factors lager than 1.9 the Z-CIs for DL, PM and REML are wider. Especially for smaller sample sizes, the CI using e.g., KnHa, Adhoc and t shrunk substantially. The larger the sample size, the smaller the shrinkage. We observed the same in the Bayesian case with the use of a UP. If a HNP is used, the CIs are always shorter, but we cannot see that the sample size would affect this. The precision gain by shorter CIs combined with a similar, if not better coverage indicates that the use of meta-regression is beneficial even for moderately different τ.

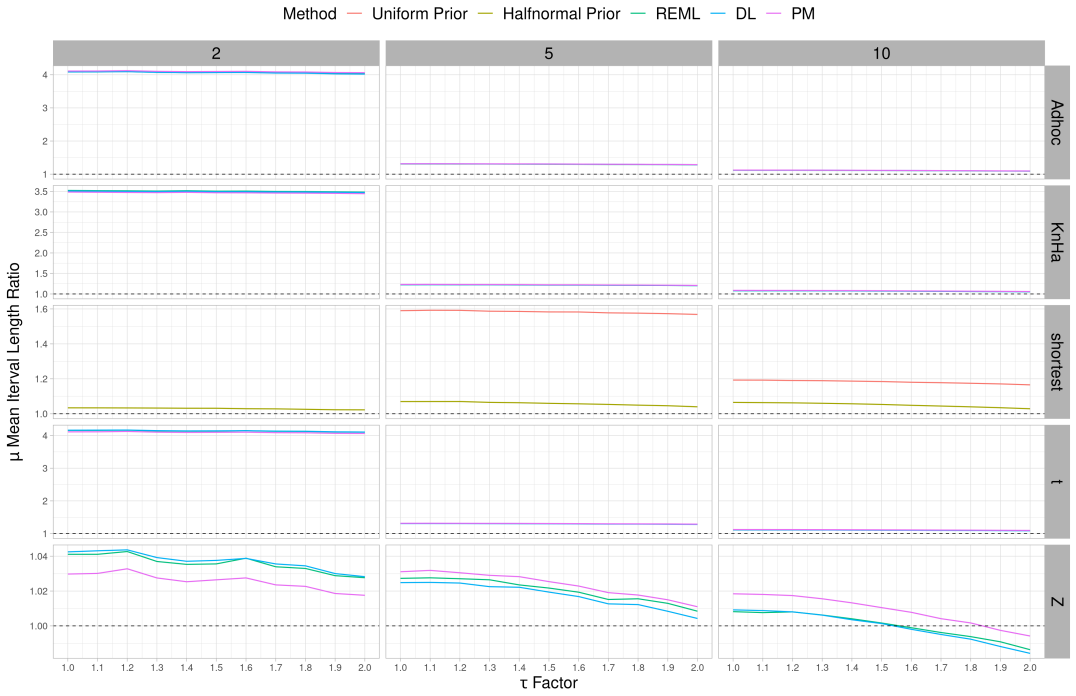

FIGURE 4.15: 95% CI length ratio of μ for varying sample sizes, factor of τ variation and CIs in meta-analysis and meta-regression. A ratio larger than 1 indicates that the CIs of individual meta-analyses are wider on average compared to meta-regression

In Figure 4.16 we show the ratio of MSE for μ. We found that only for large sample sizes and large differences in τ the frequentist setting leads to a larger MSE in meta-regression. In all other cases, meta-regression also has a beneficial effect on the MSE of μ. In the Bayesian setting we found that for small differences in τ the MSE is fairly equal, whereas for large differences in τ across subgroups the MSE of the individual meta-analyses is smaller. For a HNP and a small sample size, we would say that the MSE of meta-regression and meta-analysis is similar.

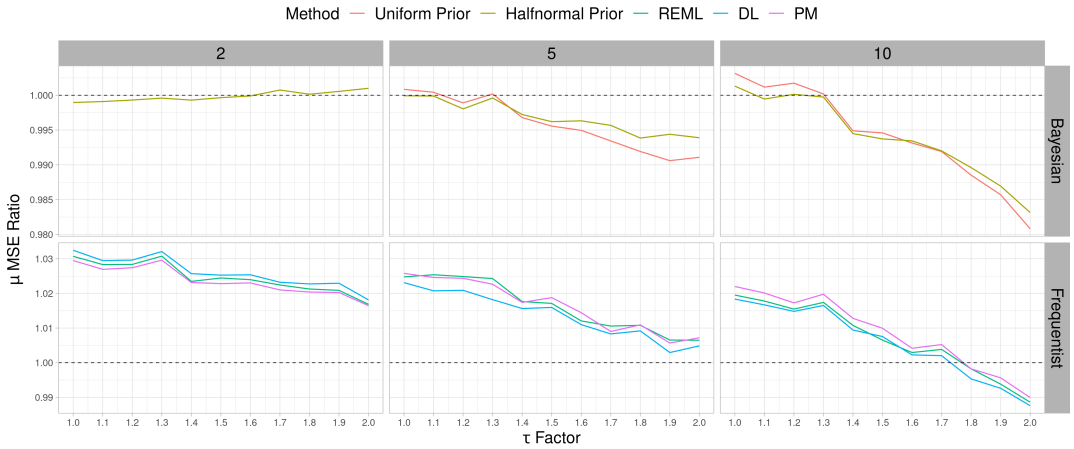

FIGURE 4.16: Mean squared error ratio of μ for varying sample sizes, factor of τ variation and CIs in meta-analysis and meta-regression for varying τ. A ratio of larger than 1 indicates that the MSE for individual meta-analyses is larger compared to meta-regression.

We would conclude that the use of meta-regression even for differing τ is overall beneficial across all CIs and methods.

4.4 Scenario III: Calibration

In Scenario III, we sampled τ from a half-normal distribution with a scale parameter of 0.5. Analogous to Röver (2020) Appendix E we want to demonstrate that the model yields consistent results. The distribution we use here is the same as the HNP used for analyzing the data. The resulting distribution of point estimates is shown in Figure 4.17. We know that for small sample sizes, the Bayesian point estimates tend to overestimate τ, since the point estimate is corrected more towards the prior median the less information is available. In fact, Bayesian methods estimate fewer τ to be close to zero for small sample sizes compared to larger sample sizes.

The same can be found for the comparison of meta-analyses and meta-regression. The effect of larger sample sizes is more evident in the UP case. For large sample sizes, the point estimates should be close to the true value and therefore the distribution should be close to $HN(0.5)$. The larger sample size in the meta-regression leads to a better representation of the half-normal distribution. The frequentist methods tend to estimate more τ to be zero for smaller sample sizes and when individual meta-analysis is used instead of meta-regression.

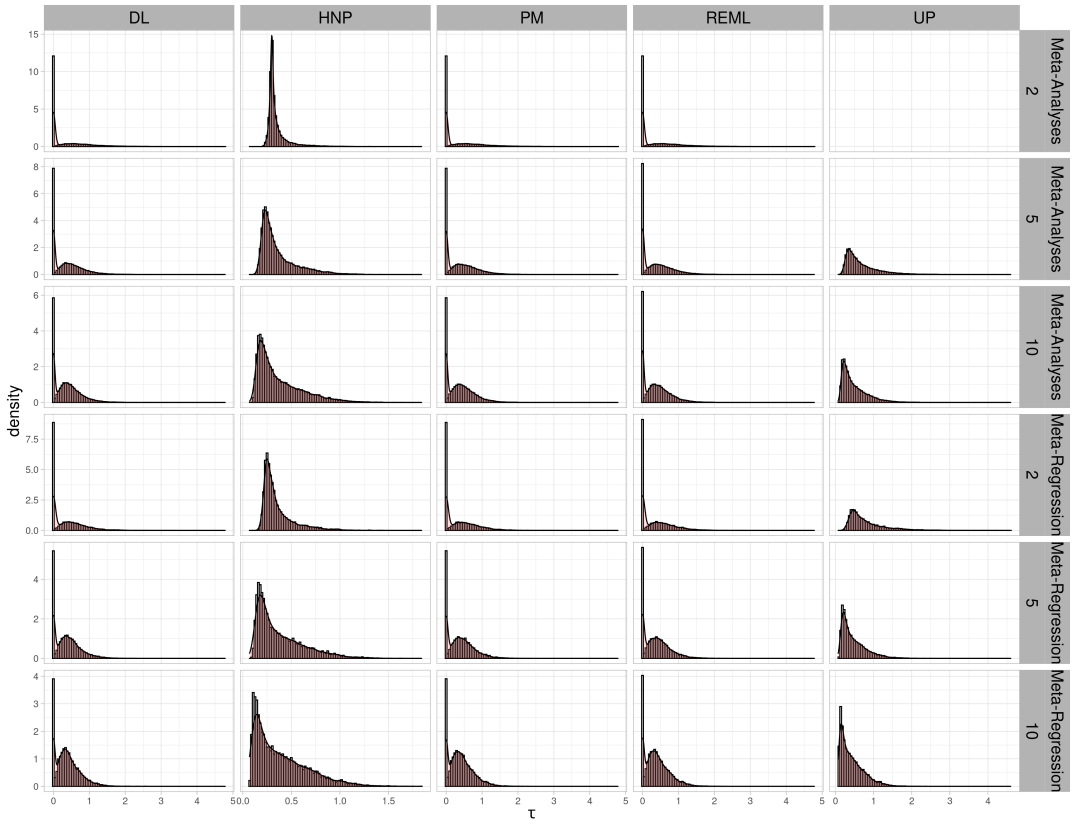

FIGURE 4.17: Distribution of τ point estimates if τ is sampled from $HN(0.5)$. The gray bar on the right indicates the sample size per subgroup and if meta-analyses or meta-regression were used.

In the Bayesian case, we can look at the distribution of the true τ quantiles. These quantiles can be calculated from the posterior distribution. For this, we use so-called probability integral transform (PIT) values (Gneiting et al., 2007):

$$p_t = F_t(x_t),$$

where $F_t(\cdot)$ in our case is the cumulative distribution function of τ from the posterior and x_t is the true value of τ used in the simulations. p_t is then nothing else than the quantile from the posterior distribution of the true τ used in the simulations.

In the case of the HNP, we would expect a uniform distribution if the prior and the true distribution are the same for both μ and τ. Since we set $\mu = 0$, we expect only roughly a uniform distribution for the PIT values of τ. The PIT values are shown for the Bayesian methods in Figure 4.18. We can see that for the HNP, the distribution is roughly a uniform distribution for both individual meta-analysis and meta-regression. In the case of the UP, we found a

skewed distribution, indicating that the UP overestimates τ to a degree. This overestimation of τ can be regarded as a conservative form of bias. Further on this topic is written in Röver (2020) Appendix B. The skewness is less apparent if the sample size is increased, either by more studies per subgroup or using meta-regression.

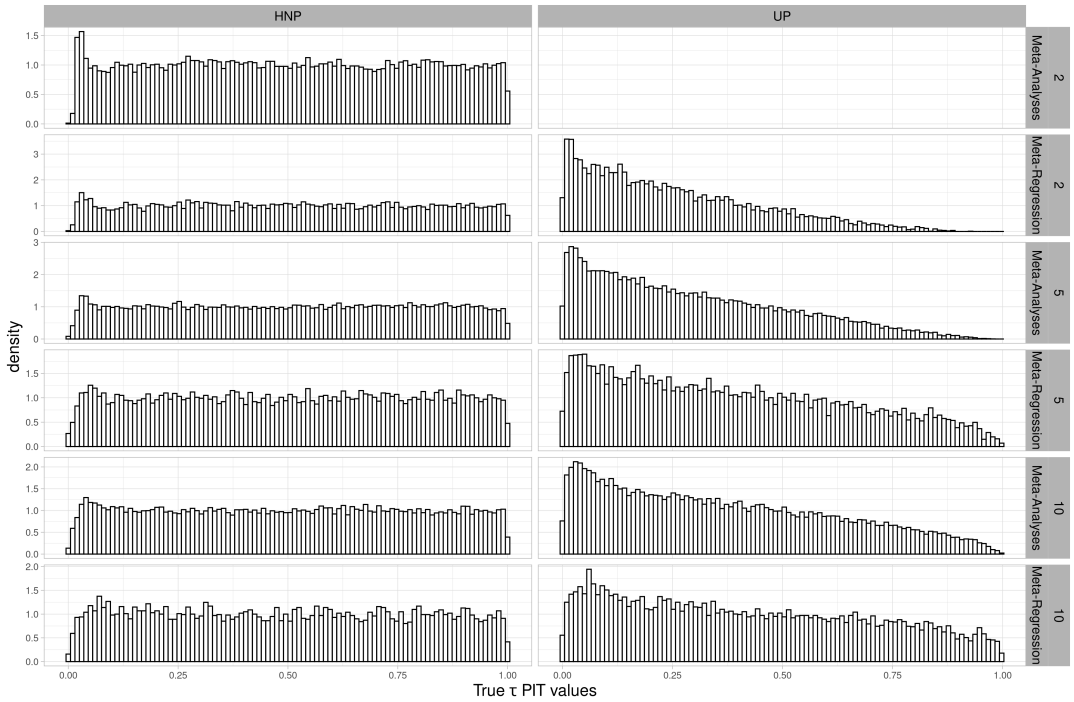

FIGURE 4.18: PIT values (or quantile) of true τ parameter for Bayesian methods and sample sizes per subgroup when sampling τ from $HN(0.5)$ in individual meta-analyses and meta-regression. The gray bar on the right indicates the sample size per subgroup and if meta-analyses or meta-regression were used.

We further found that the use of a HNP leads to a 95% coverage probability of μ for all sample sizes and methods. The use of a UP however, leads to coverage probabilities of μ larger than 95%. The coverage probabilities of μ for UP, HNP and REML are shown in Table 4.1. For REML we found that the Z-CI is liberal. The Adhoc- and t-CI tend to be conservative, while for large sample sizes k and meta-regression covering the 95%. The KnHa-CI for covers at least 94%, while for large sample sizes and meta-regression also covers the 95%. The coverage of μ for DL and PM is given in Table A.7 (Appendix). They show a similar picture as REML. Friede et al. (2017a) also concluded that the τ estimator is not so important, what is more important is whether the τ estimation uncertainty is taken into account.

	Interval:	shortest		Z	KnHa	Adhoc	t
k	Type	UP	HNP	REML	REML	REML	REML
2	Meta-Analyses	.	0.95	0.90	0.94	1	1
2	Meta-Regression	0.99	0.95	0.92	0.94	0.99	0.99
5	Meta-Analyses	0.99	0.95	0.92	0.94	0.98	0.97
5	Meta-Regression	0.97	0.95	0.93	0.94	0.96	0.95
10	Meta-Analyses	0.97	0.95	0.93	0.94	0.96	0.96
10	Meta-Regression	0.96	0.95	0.94	0.95	0.95	0.95

TABLE 4.1: Coverage of μ for τ sampled from $HN(0.5)$ for UP, HNP and REML. The shown values multiplied by 100 represent percentages.

Table A.8 (Appendix) shows the mean μ CI lengths for UP, HNP and REML. DL and PM are shown in Table A.9 (Appendix). The HNP leads to the shortest CIs on average that maintain the 95% level. The Z-CI is found to have shorter CIs than the HNP. However, it does not maintain the 95% level for μ. The coverage probability is overall never lower than 90%. For small sample sizes, KnHa-, Adhoc- and the t-CI have very wide CIs on average. The widest CI originate from the UP. KnHa-, Adhoc- and the t-CIs produce wide CIs for small sample sizes. The usage of meta-regression with small sample sizes drastically shrinks the average CI length. All CI are shorter when either using meta-regression or with larger sample sizes. This is analogous to the previous findings.

4.5 Scenario VI: Location invariance

As a last scenario, we want to show that the variation of μ in the subgroups does not influence the results from the other scenarios, aside from the location of the μ point estimates. In Figure 4.19 we show the distribution of μ point estimates for $\tau = 0.3$. All models were able to differentiate between the true parameters of μ across the subgroups. The true values are indicated by the dashed lines. Each subgroup μ is represented by one color. We can see that each subgroup point estimate distribution of μ has a mean around the true μ (indicated by the dashes line) per subgroup. This holds also for the other τ parameters, which are shown in Figures A.8 and A.9 (Appendix).

We found that the coverage probabilities of μ from a simulation using $\mu = 0$ and parameters from simulations using varying μ are very similar. The coverage probabilities for μ given $\tau = 0.3$ are shown in Table A.10 (Appendix).

The coverage probabilities for Scenario I are shown in Figure 4.8. In Scenario I, we can see that in Figure 4.1 the point estimates are centered around zero, whereas the point estimates are distributed around the true μ for each subgroup with approximately equal variances.

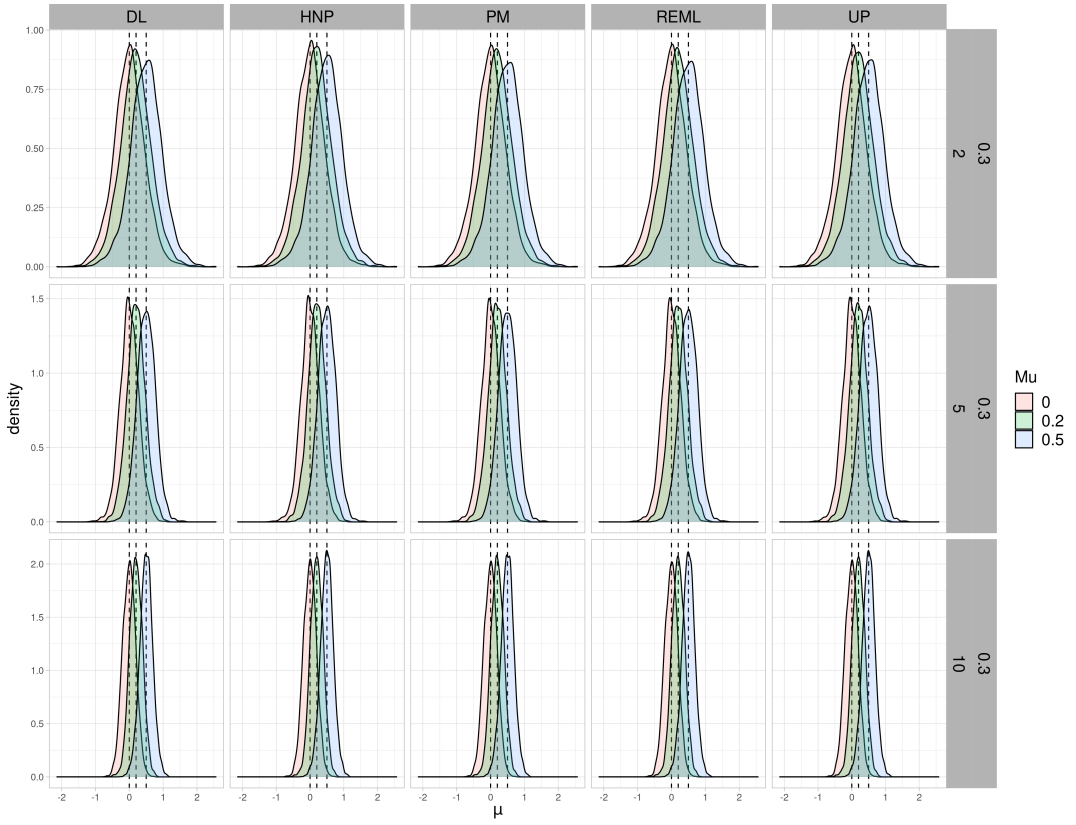

FIGURE 4.19: Distribution of point estimates of μ split by subgroup for $\tau = 0.3$ in meta-regression. Bayesian models use the posterior median. The true value of μ per subgroup is indicated by color. The gray bar on the right indicates the sample size per subgroup and the respective true τ

Conclusions and discussion

Our expectations were that meta-regression should improve the estimation of τ , since more data is available. Associated with the improved estimation for τ, we also expected an improved estimation of μ, i.e., more precise CIs. Especially, the tendency of frequentist methods to estimate τ to be equal to zero should be reduced as well. This approach was mentioned, e.g., in Dias et al. (2013) and Donegan et al. (2015). However, it is not often used in practice.

We found that when subgroups are present, meta-regression seems to have an advantage over individual meta-analyses. To study meta-regression under realistic conditions, we used data from the CDSR. We demonstrated that when meta-regression is used on the empirical dataset from the CDSR, the mean CI lengths shrunk when using meta-regression instead of individual meta-analyses. One exception to this finding is the CI based on normal approximation (Z). We also found that using meta-regression in a frequentist setting reduces the number of estimated zeros for τ.

After applying meta-regression on the empirical dataset, we conducted a simulation study to systematically check the properties, where we can control the parameters and certain boundary cases. First, we used fixed τ (and μ). We again found that using a meta-regression, rather than individual meta-analyses, shrinks the CI lengths of μ. In the same moment the coverage probability of μ for liberal CIs is increased, while for conservative CIs meta-regression brings the coverage probability closer to the 95% level.

In a second scenario, we showed that meta-regression is also robust when the assumptions of equal τ values is violated to a degree. Here, the coverage probability for μ does not severely drop while the CI lengths still shrink even with severe differences in τ. In the third scenario, we showed what happens if the true τ in the simulations is the same as the HNP prior used in the

analysis. The distribution we sample from is here equivalent to the prior HNP for τ. This was done to check the calibration of our model. In the short fourth scenario, we showed that the choice of μ in each subgroup does not influence the other scenarios, aside from the location of μ.

One has to keep in mind, however, when investigating differences between subgroups, the credibility of this effect modification should be assessed first (Schandelmaier et al., 2020). Once the analysis was found to be credible, we found that using a meta-regression was beneficial compared with individual meta-analyses. The credibility assessment needs to be done even if only individual subgroup meta-analyses are conducted.

Meta-regression promises advantages over individual meta-analyses in the context of subgroups, as shown in an application on CDSR dataset and simulations. Therefore, meta-regression should be considered more often. Especially for small sample sizes, one should consider using a Bayesian framework for the analysis, if an educated guess on the distribution of τ can be made. The `bayesmeta` package provides a straightforward and easy way to do so.

Bibliography

Bodnar, O., Link, A., Arendacká, B., Possolo, A., & Elster, C. (2017). Bayesian estimation in random effects meta-analysis using a non-informative prior. *Statistics in Medicine*, *36*(2), 378–399. https://doi.org/10.1002/sim.7156

Böhning, D., Malzahn, U., Dietz, E., Schlattmann, P., Viwatwongkasem, C., & Biggeri, A. (2002). Some general points in estimating heterogeneity variance with the DerSimonian–Laird estimator. *Biostatistics*, *3*(4), 445–457. https://doi.org/10.1093/biostatistics/3.4.445

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Bowden, J., Tierney, J. F., Copas, A. J., & Burdett, S. (2011). Quantifying, displaying and accounting for heterogeneity in the meta-analysis of RCTs using standard and generalised Q statistics. *BMC Medical Research Methodology*, *11*(1), 41. https://doi.org/10.1186/1471-2288-11-41

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A non-degenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685–709. https://doi.org/10.1007/s11336-013-9328-2

Dersimonian, R., & Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials*. https://doi.org/10.1016/J.CCT.2006.04.004

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Dias, S., Sutton, A. J., Welton, N. J., & Ades, A. E. (2013). Evidence Synthesis for Decision Making 3: Heterogeneity—Subgroups, Meta-Regression,

Bias, and Bias-Adjustment. *Medical Decision Making*, *33*(5), 618–640. https://doi.org/10.1177/0272989X13485157

Donegan, S., Williams, L., Dias, S., Tudur-Smith, C., & Welton, N. (2015). Exploring treatment by covariate interactions using subgroup analysis and meta-regression in cochrane reviews: A review of recent practice. *PloS One*, *10*(6), e0128804. https://doi.org/10.1371/journal.pone.0128804

Follmann, D. A., & Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics*, *55*(3), 732–737. https://doi.org/10.1111/j.0006-341x.1999.00732.x

Friede, T., Röver, C., Wandel, S., & Neuenschwander, B. (2017a). Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods*, *8*(1), 79–91. https://doi.org/10.1002/jrsm.1217

Friede, T., Röver, C., Wandel, S., & Neuenschwander, B. (2017b). Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical Journal*, *59*(4), 658–671. https://doi.org/10.1002/bimj.201500236

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015, July). *Bayesian Data Analysis* (3rd ed.). Chapman; Hall/CRC. https://doi.org/10.1201/b16018

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, *5*(10), 3–8. https://doi.org/10.2307/1174772

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(2), 243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Goodwin, A., Parker, S., Ghersi, D., & Wilcken, N. (2013). Post-operative radiotherapy for ductal carcinoma in situ of the breast. *Cochrane Database of Systematic Reviews*, (11). https://doi.org/10.1002/14651858.CD000563.pub7

Günhan, B. K., Röver, C., & Friede, T. (2020). Meta-analysis of few studies involving rare events [arXiv: 1809.04407]. *Research Synthesis Methods*, *11*(1), 74–90. https://doi.org/10.1002/jrsm.1370

Harrer, M., Cuijpers, P., A, F. T., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st). Chapman & Hall/CRC Press. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/

Hartung, J., & Makambi, K. H. (2002). Positive estimation of the between-study variance in meta-analysis : Theory and methods. *South African Statistical Journal*, *36*(1), 55–76. https://doi.org/10.10520/EJC99050

Hartung, J. (1999). An Alternative Method for Meta-Analysis. *Biometrical Journal*, *41*(8), 901–916. https://doi.org/10.1002/(SICI)1521-4036(199912)41:8<901::AID-BIMJ901>3.0.CO;2-W

Hartung, J., & Knapp, G. (2001a). On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine*, *20*(12), 1771–1782. https://doi.org/10.1002/sim.791

Hartung, J., & Knapp, G. (2001b). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, *20*(24), 3875–3889. https://doi.org/10.1002/sim.1009

Hartung, J., & Makambi, K. H. (2003). Reducing the Number of Unjustified Significant Results in Meta-analysis. *Communications in Statistics - Simulation and Computation*, *32*(4), 1179–1190. https://doi.org/10.1081/SAC-120023884

Hedges, L., & Olkin, I. (1985, January). *Statistical Methods in Meta-Analysis*.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Jackson, D., Bowden, J., & Baker, R. (2010). How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *Journal of Statistical Planning and Inference*, *140*(4), 961–970. https://doi.org/10.1016/j.jspi.2009.09.017

Jackson, D., Turner, R., Rhodes, K., & Viechtbauer, W. (2014). Methods for calculating confidence and credible intervals for the residual between-study variance in random effects meta-regression models. *BMC Medical Research Methodology*, *14*, 103. https://doi.org/10.1186/1471-2288-14-103

Jackson, D., & Riley, R. D. (2014). A refined method for multivariate meta-analysis and meta-regression. *Statistics in Medicine*, *33*(4), 541–554. https://doi.org/10.1002/sim.5957

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, *22*(17), 2693–2710. https://doi.org/10.1002/sim.1482

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86. Retrieved February 14, 2022, from https://www.jstor.org/stable/2236703

McNeish, D. (2016). On using bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773. https://doi.org/10.1080/10705511.2016.1186549

Murray, A., Nguyen, T. M., Parker, C. E., Feagan, B. G., & MacDonald, J. K. (2020). Oral 5-aminosalicylic acid for induction of remission in ulcerative colitis. *Cochrane Database of Systematic Reviews*, (8). https://doi.org/10.1002/14651858.CD000543.pub5

Novianti, P. W., Roes, K. C. B., & van der Tweel, I. (2014). Estimation of between-trial variance in sequential meta-analyses: A simulation study. *Contemporary Clinical Trials*, *37*(1), 129–138. https://doi.org/10.1016/j.cct.2013.11.012

Panityakul, T., Bumrungsup, C., & Knapp, G. (2013). On Estimating Residual Heterogeneity in Random-Effects Meta-Regression: A Comparative Study. *Journal of Statistical Theory and Applications*, *12*(3), 253–265. https://doi.org/10.2991/jsta.2013.12.3.4

Paule, R. C., & Mandel, J. (1982). Consensus Values and Weighting Factors. *Journal of Research of the National Bureau of Standards (1977)*, *87*(5), 377–385. https://doi.org/10.6028/jres.087.022

Pullenayegum, E. M. (2011). An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Statistics in Medicine*, *30*(26), 3082–3094. https://doi.org/10.1002/sim.4326

Rhodes, K. M., Turner, R. M., & Higgins, J. P. T. (2015). Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology*, *68*(1), 52–60. https://doi.org/10.1016/j.jclinepi.2014.08.012

Rhodes, K. M., Turner, R. M., White, I. R., Jackson, D., Spiegelhalter, D. J., & Higgins, J. P. T. (2016). Implementing informative priors for heterogeneity in meta-analysis using meta-regression and pseudo data. *Statistics in Medicine*, *35*(29), 5495–5511. https://doi.org/10.1002/sim.7090

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ*, *342*, d549. https://doi.org/10.1136/bmj.d549

Röver, C. (2020). Bayesian random-effects meta-analysis using the bayesmeta R package [arXiv: 1711.08683]. *Journal of Statistical Software*, *93*(6). https://doi.org/10.18637/jss.v093.i06

Röver, C., & Friede, T. (2021). Bounds for the weight of external data in shrinkage estimation. *Biometrical Journal*, *63*(5), 1131–1143. https://doi.org/10.1002/bimj.202000227

Röver, C., & Friede, T. (2017). Discrete Approximation of a Mixture Distribution via Restricted Divergence. *Journal of Computational and Graphical Statistics*, *26*(1), 217–222. https://doi.org/10.1080/10618600.2016.1276840

Röver, C., & Friede, T. (2020). Dynamically borrowing strength from another study through shrinkage estimation. *Statistical Methods in Medical Research*, *29*(1), 293–308. https://doi.org/10.1177/0962280219833079

Röver, C., Knapp, G., & Friede, T. (2015). Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Medical Research Methodology*, *15*(1), 99. https://doi.org/10.1186/s12874-015-0091-1

Röver, C., Wandel, S., & Friede, T. (2019). Model averaging for robust extrapolation in evidence synthesis. *Statistics in Medicine*, *38*(4), 674–694. https://doi.org/10.1002/sim.7991

Rukhin, A. L. (2013). Estimating heterogeneity variance in meta-analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(3), 451–469. https://doi.org/10.1111/j.1467-9868.2012.01047.x

Rukhin, A. L., Biggerstaff, B. J., & Vangel, M. G. (2000). Restricted maximum likelihood estimation of a common mean and the Mandel–Paule algorithm. *Journal of Statistical Planning and Inference*, *83*(2), 319–330. https://doi.org/10.1016/S0378-3758(99)00098-1

Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods*, *13*(1), 31–48. https://doi.org/10.1037/1082-989X.13.1.31

Schandelmaier, S., Briel, M., Varadhan, R., Schmid, C. H., Devasenapathy, N., Hayward, R. A., Gagnier, J., Borenstein, M., van der Heijden, G. J. M. G., Dahabreh, I. J., Sun, X., Sauerbrei, W., Walsh, M., Ioannidis, J. P. A., Thabane, L., & Guyatt, G. H. (2020). Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN)

in randomized controlled trials and meta-analyses. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, *192*(32), E901–E906. https://doi.org/10.1503/cmaj.200077

Sidik, K., & Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine*, *26*(9), 1964–1981. https://doi.org/10.1002/sim.2688

Sidik, K., & Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Statistics in Medicine*, *21*(21), 3153–3159. https://doi.org/10.1002/sim.1262

Sweeting, M. J., Sutton, A. J., & Lambert, P. C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, *23*(9), 1351–1375. https://doi.org/10.1002/sim.1761

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, *21*(11), 1559–1573. https://doi.org/10.1002/sim.1187

Tufanaru, C., Munn, Z., Stephenson, M., & Aromataris, E. (2015). Fixed or random effects meta-analysis? Common methodological issues in systematic reviews of effectiveness. *JBI Evidence Implementation*, *13*(3), 196–207. https://doi.org/10.1097/XEB.0000000000000065

Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., & Higgins, J. P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology*, *41*(3), 818–827. https://doi.org/10.1093/ije/dys041

Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., & Higgins, J. P. T. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, *34*(6), 984–998. https://doi.org/10.1002/sim.6381

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How Many Studies Do You Need?: A Primer on Statistical Power for Meta-Analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. https://doi.org/10.3102/1076998609346961

van Aert, R. C. M., & Jackson, D. (2018). Multistep estimators of the between-study variance: The relationship with the Paule-Mandel estimator. *Statistics in Medicine*, *37*(17), 2616–2629. https://doi.org/10.1002/sim.7665

van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine*, *21*(4), 589–624. https://doi.org/10.1002/sim.1040

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*, 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine*, *26*(1), 37–52. https://doi.org/10.1002/sim.2514

Williams, D. R., Rast, P., & Bürkner, P.-C. (2018, January). *Bayesian Meta-Analysis with Weakly Informative Prior Distributions* (tech. rep.) [type: article]. PsyArXiv. https://doi.org/10.31234/osf.io/7tbrm

Appendix

A.1 Code

The code can be made available on request.

A.2 Computational aspects

The `bayesmeta` package uses the DIRECT (divergence restricting conditional tessellation) algorithm described in Röver and Friede (2017), to approximate the marginal posterior distribution of μ. In our context, it can be seen as a way of doing Bayesian analysis without the need of Markov chain Monte Carlo sampling (Röver, 2020). Alternatives to this approach are probabilistic languages using MCMC methods like Stan and Jags. Since no assessment of convergence is required, the `bayesmeta` package has advantages in the context of simple meta-analysis/-regression compared to approaches using Jags or Stan.

$P(\tau|Y)$ is the marginal posterior distribution of τ. For the application in meta-analysis, it can be derived in closed form (Röver, 2020). The conditional posterior $P(\mu|\tau, Y)$ can also be computed. However, the marginal distribution $P(\mu|Y)$ that is of interest can not be easily computed and needs to be approximated.

$$P(\mu|Y) = \int P(\mu|\tau, Y)P(\tau|Y)d\tau \tag{A.1}$$

We can approximate the marginal distribution of μ through a discrete set of points:

$$P(\mu|Y) \approx \sum_i P(\mu|\tau_i, Y)\pi_i \tag{A.2}$$

The question is how to define the grid points for the approximation. The idea of the DIRECT algorithm is to link the bin spacing to a measure of similarity of the resulting conditional posterior distributions of μ ($P(\mu|\tau_i, Y)$). For this, we use the symmetrized Kullback-Leibler Divergence (D_s). It is further described in Section A.3. We use the symmetrized version to ensure that it does not matter in which order we compare $P(\mu|\tau_i, Y)$. We set the bin margins τ_i such that a predefined maximum divergence δ inside the bins is not exceeded, e.g, $\delta = 0.01$.

The binning procedure works as follows. The left bin margin for the first bin is set to a point of choice, e.g., $\tau = 0$ or τ equal to the $\epsilon/2$ quantile of the marginal posterior distribution of τ. ϵ is the amount of neglected tail probability that we would allow. For example, it can be set to 0.001. For the first bin, we set the reference point ($\tilde{\tau}_1$), that represents the bin, equal to the left bin margin. To find the right margin for the bin, we increase τ until $D_s = \delta$. The resulting value of $\tau_{(1)}$ is then our margin. The right margin of the first bin is also the left margin of the second bin. After the left bin margin for the second bin is known, we search for the reference point of the bin. We increase τ again until $D_s = \delta$. We then set this as the reference point $\tilde{\tau}_2$. From this reference point, we then look for the point where $P(\mu|\tau_i, Y)$ is different from the distribution using $\tilde{\tau}_2$. This point $\tau_{(2)}$ we set to the right margin of the second bin. We do this until, for the last bin margin, $P(T > \tau)$ is less or equal to the remaining predefined amount of neglected tail probability (ϵ_r). ϵ_r is dependent on the first reference point $\tilde{\tau}_1$. It is given by: $\epsilon_r = \epsilon - P(T \leq \tilde{\tau}_1)$. We then use the reference points to approximate the marginal distribution of μ as the weighted sum of $P(\mu|\tau_i, Y)$. The weights π_i for each $\tilde{\tau}_i$ are given by the integral of $P(\tau|Y)$ inside the respective bin margins.

A.3 Kullback–Leibler divergence

The Kullback–Leibler Divergence (also known as relative entropy) is used to quantify how one probability distribution P is different to another probability distribution Q (Kullback & Leibler, 1951).

The discretized version of the Kullback–Leibler Divergence is given by:

$$D(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right) = -\sum_x P(x) \log\left(\frac{Q(x)}{P(x)}\right) \tag{A.3}$$

Kullback–Leibler Divergence is also defined for continuous random variables:

$$D(P||Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{A.4}$$

The Kullback–Leibler Divergence is not symmetric. In case of comparing two probability distributions (P, Q) a symmetrized version can be used:

$$D_s(P||Q) = D(P||Q) + D(Q||P) = D_s(Q||P) \tag{A.5}$$

The Kullback–Leibler Divergence can be used in our application to compare μ and τ estimates at the same time. It can be used to compare the estimation of θ across the used methods. In the Bayesian case, we can integrate the distribution of interest out of the posterior distribution. Then the resulting distribution can be compared with the true distribution. The discrete Kullback–Leibler Divergence is based on a fine grid to represent the distributions. In order to avoid point densities and errors in the used R method, we set the heterogeneity

$$\hat{\tau} = \begin{cases} 0.01 & \text{if } \hat{\tau} < 0.01 \\ \hat{\tau} & \text{else.} \end{cases} \tag{A.6}$$

The same holds for the true values of τ.

Meta-regression provides only a single estimate for τ. Thus, in case of differing τ_i across subgroups the Kullback–Leibler Divergence can not easily be compared between individual meta-analysis and meta-regression. Moreover, the assumptions of the meta-regression are violated. For example, one could take a pooled τ_{pooled} of the true τ_i and then use this in the given functions or compute one Kullback–Leibler Divergence estimate for each τ_i while always using the estimated $\hat{\tau}$ from the meta-regression.

A.4 Additional plots and tables

A.4.1 Example meta-analyses and meta-regression

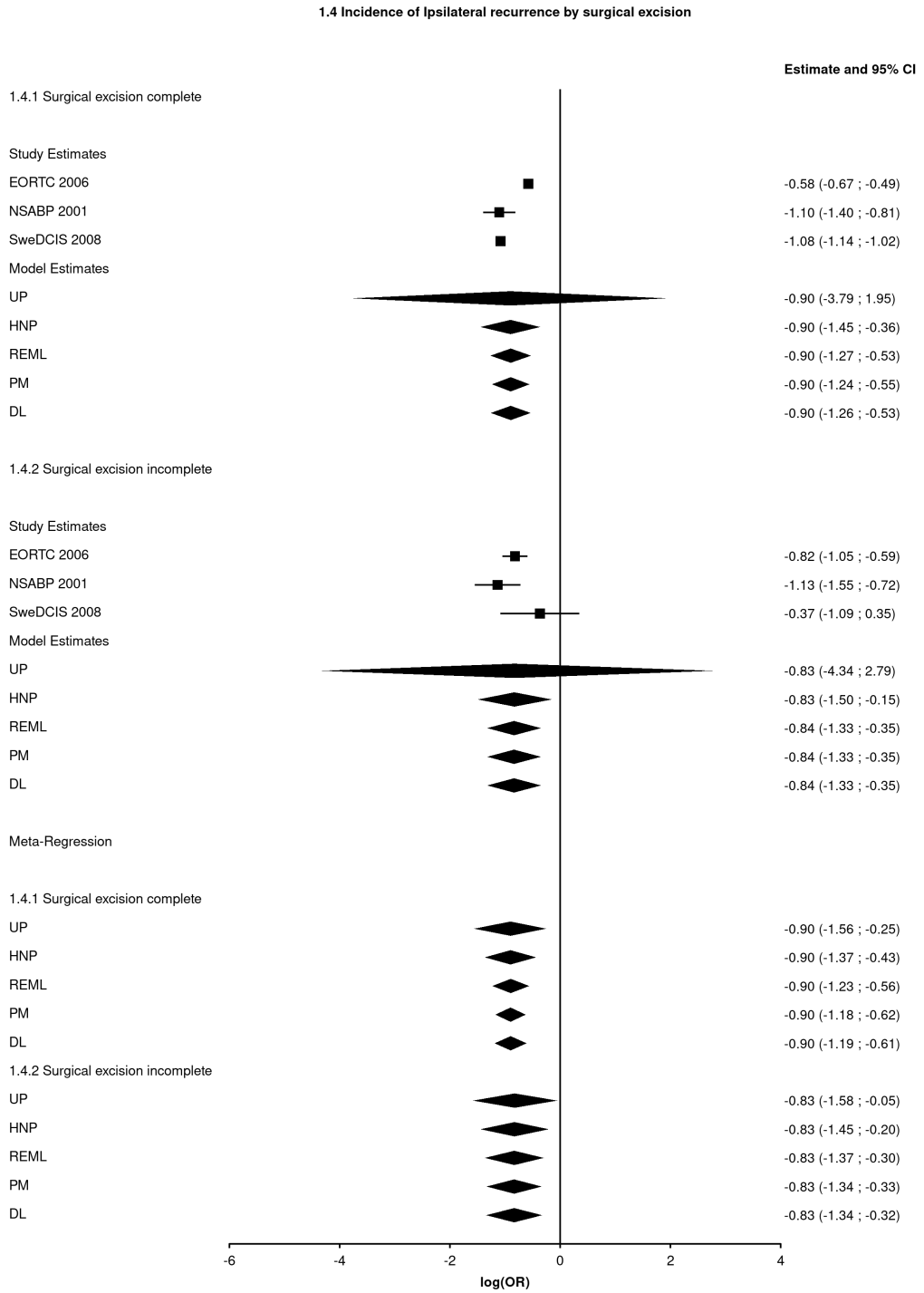

FIGURE A.1: Example forest plot for first comparison and fourth outcome in Goodwin et al. (2013). Shown are the μ point estimates with CIs for subgroup specific meta-analyses and meta-regression. This was done for the methods UP, HNP, REML, DL and PM. In the frequentist case, the Z CI was used.

A.4.2 Estimation of meta-regressions and individual meta-analyses using the empirical data

Type	UP	HNP	REML	PM	DL
Meta-Analysis	0	0	0.54	0.61	0.61
Meta-Regression	0	0	0.34	0.45	0.45

TABLE A.1: Proportion of zero point estimates for τ where the number of studies per subgroup is at least two. The shown values multiplied by 100 represent percentages.

	UP	HNP	REML	PM	DL
shortest	1.84	1.05	.	.	.
Adhoc	.	.	1.62	1.56	1.55
KnHa	.	.	1.53	1.47	1.47
t	.	.	1.63	1.56	1.56
Z	.	.	1.04	1.02	1.02

TABLE A.2: Mean of μ interval length ratio, where the number of studies per subgroup is at least two. The ratio is computed as the interval length of individual meta-analyses divided by the interval length of meta-regression. A ratio of< 1 indicates that the interval of individual meta-analyses are shorter and vice versa.

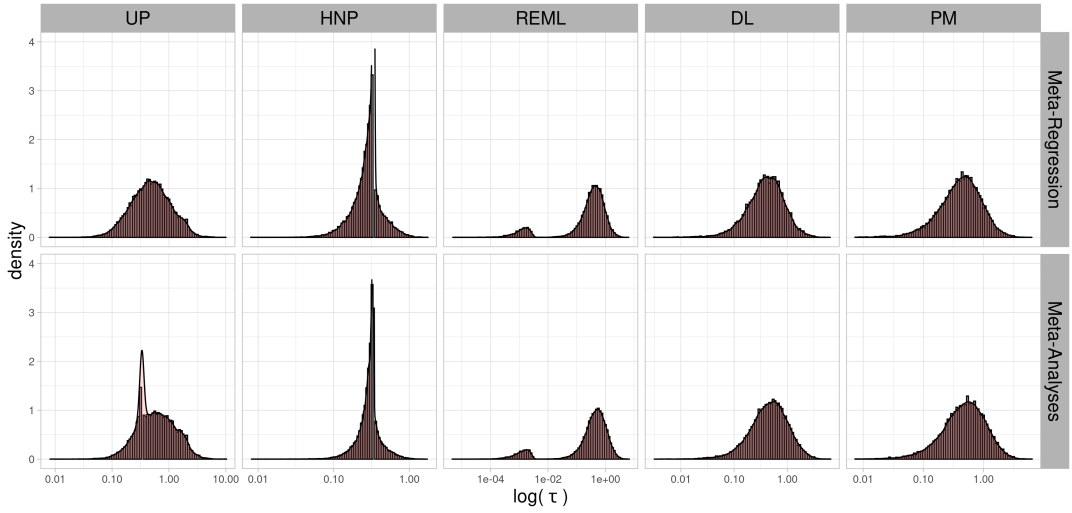

FIGURE A.2: Distribution of log-transformed point estimates for τ in meta-regression and meta-analysis. The log results in zero point estimates for τ not being displayed. The relative share of zero point estimates is shown in Table 3.3

A.4.3 Estimation of meta-regressions and individual meta-analyses using the empirical data with subgroups, where overall meta-analyses per subgroup were carried out

Thus far, we have considered subgroup analyses that did not necessarily involve an overall meta-analysis. An overall analysis means that we combine the study estimates from the subgroups and analyze them together in a single meta-analysis. Here, we consider only the cases where such an overall analysis was conducted. Since the authors of the studies used an overall meta-analysis, they considered a common between-trial heterogeneity as plausible. Therefore, the assumption of a common between-trial heterogeneity made by meta-regression should be plausible. The results show a similar behavior compared to an analysis containing all data with subgroups.

Type	Count
Reviews	2107
Comparisons	4621
Outcomes	16978
Subgroups	44721
Studies	166506

TABLE A.3: Summary of Cochrane database for subgroups subset with binary data where overall analyses were carried out

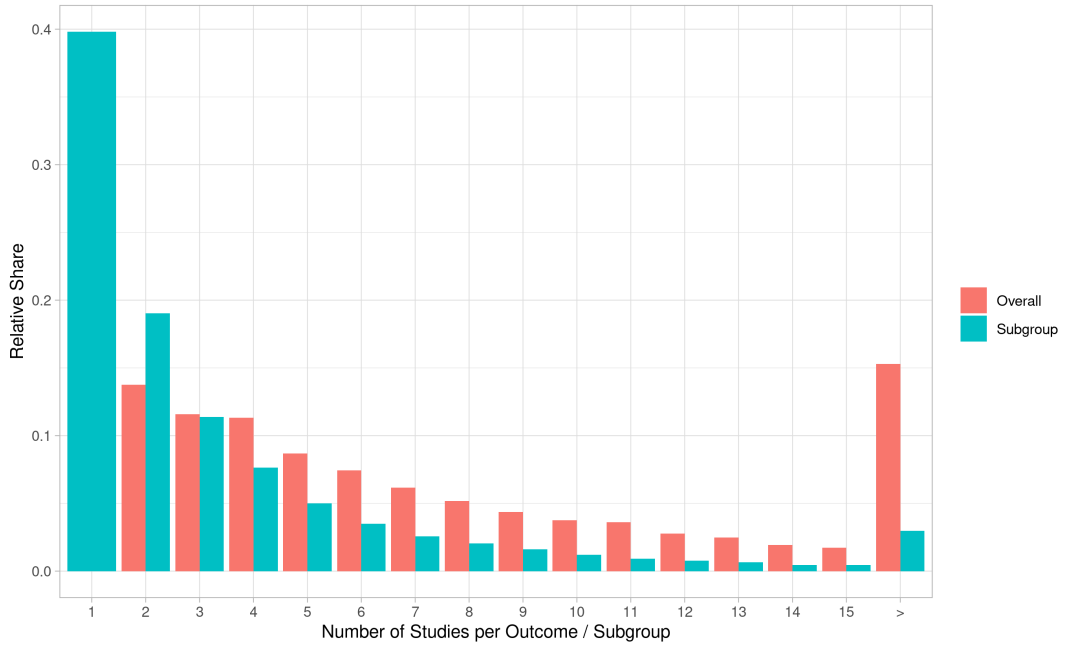

FIGURE A.3: Relative share of study counts and subgroups per analysis where overall analyses were carried out

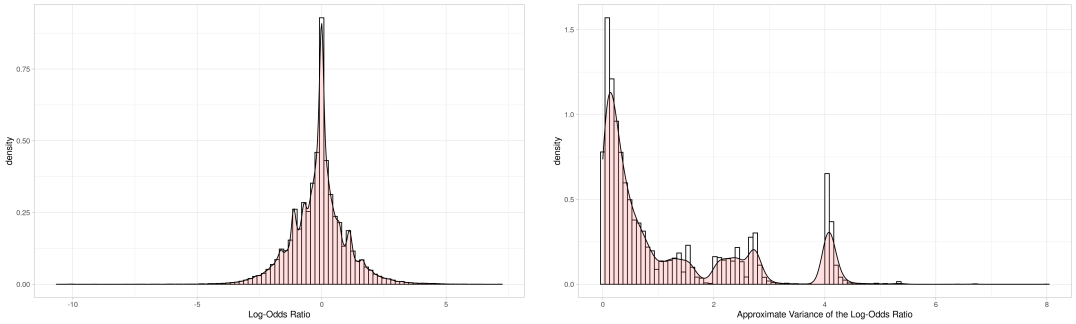

FIGURE A.4: Log-Odds ratios and their approximate variances where overall analyses were carried out

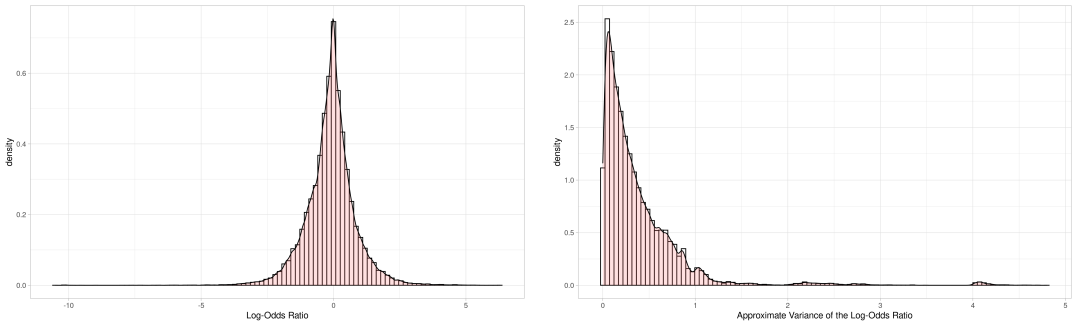

FIGURE A.5: Log-Odds ratios and their approximate variances if "single & double zero" studies are removed where overall analyses were carried out

Type	UP	HNP	REML	PM	DL
Meta-Analysis	0	0	0.51	0.60	0.60
Meta-Regression	0	0	0.39	0.52	0.52

TABLE A.4: Relative share of zero point estimates for τ where overall analyses were carried out. The shown values multiplied by 100 represent percentages.

	UP	HNP	REML	PM	DL
shortest	1.64	1.05	.	.	.
Adhoc	.	.	1.46	1.45	1.45
KnHa	.	.	1.39	1.38	1.38
t	.	.	1.47	1.45	1.46
Z	.	.	1.02	1.02	1.02

TABLE A.5: Mean of μ interval length ratio where overall analyses were carried out. The ratio is computed as the interval length of individual meta-analyses divided by the interval length of meta-regression. A ratio of< 1 indicates that the interval of individual meta-analyses are shorter and vice versa.

	UP	HNP	REML	PM	DL
τ_{diff}	0.15	0.02	-0.01	-0.01	-0.003

TABLE A.6: Mean difference of τ meta-analysis and τ meta-regression, where overall analyses were carried out. A difference < 0 indicates that on average, the point estimate of τ is larger when using meta-regression

A.4.4 Scenario I

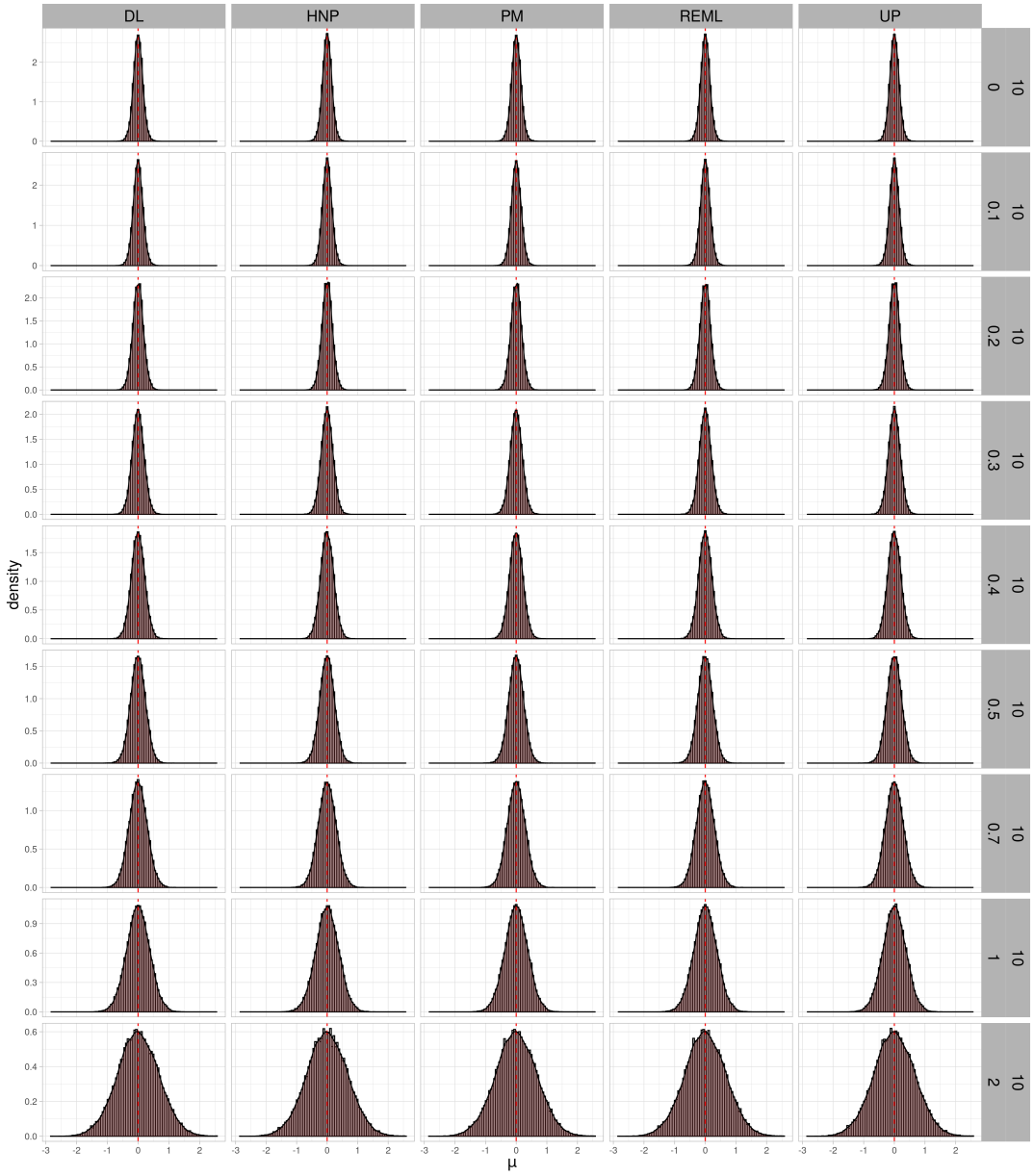

FIGURE A.6: Point estimates of μ for varying τ values, a sample size per subgroup of k=10 and a true $\mu = 0$ for individual meta-analyses. The gray bar on the right indicates the sample size per subgroup and the respective true τ.

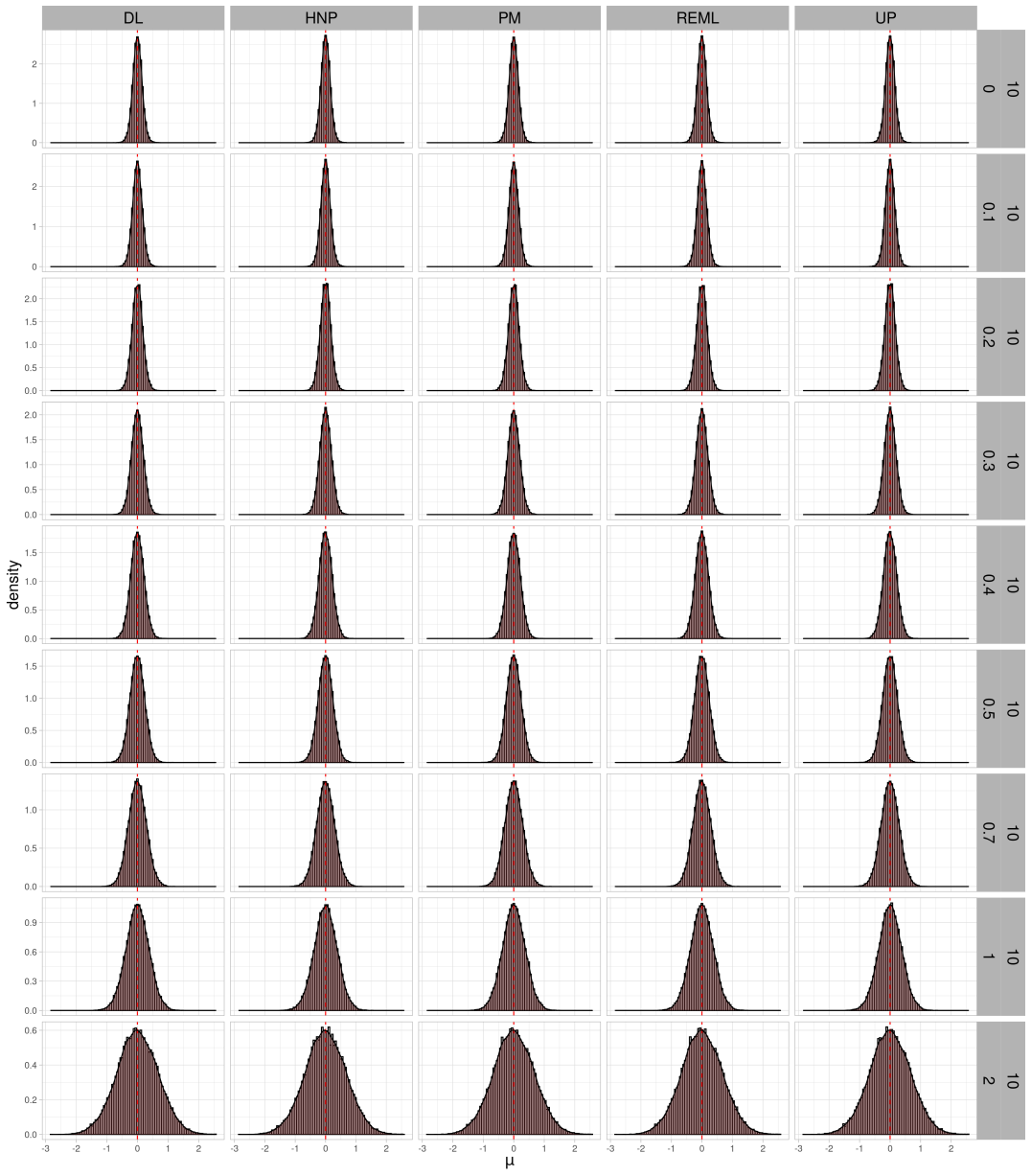

FIGURE A.7: Point estimates of μ for varying τ values, a sample size per subgroup of k=10 and a true $\mu = 0$ for meta-regression. The gray bar on the right indicates the sample size per subgroup and the respective true τ.

A.4.5 Scenario II

A.4.6 Scenario III

	Interval:	Z	KnHa	Adhoc	t
k	Type	DL	DL	DL	DL
2	Meta-Analyses	0.90	0.94	1	1
2	Meta-Regression	0.92	0.94	0.99	0.99
5	Meta-Analyses	0.92	0.94	0.98	0.97
5	Meta-Regression	0.93	0.94	0.95	0.95
10	Meta-Analyses	0.93	0.94	0.96	0.96
10	Meta-Regression	0.94	0.95	0.95	0.95
		PM	PM	PM	PM
2	Meta-Analyses	0.90	0.94	1	1
2	Meta-Regression	0.92	0.94	0.99	0.99
5	Meta-Analyses	0.92	0.94	0.98	0.98
5	Meta-Regression	0.93	0.94	0.96	0.96
10	Meta-Analyses	0.93	0.94	0.96	0.96
10	Meta-Regression	0.94	0.95	0.95	0.95

TABLE A.7: Coverage of μ for τ sampled from $HN(0.5)$ for DL and PM. The shown values multiplied by 100 represent percentages.

	Interval:	shortest		Z	KnHa	Adhoc	t
k	Type	UP	HNP	REML	REML	REML	REML
2	Meta-Analyses	.	2.15	2.10	10.55	13.59	13.59
2	Meta-Regression	5.14	2.13	2.06	3.02	3.40	3.35
5	Meta-Analyses	2.29	1.30	1.24	1.64	1.40	1.76
5	Meta-Regression	1.46	1.28	1.24	1.34	0.96	1.37
10	Meta-Analyses	1.09	0.90	0.86	0.96	1.01	0.99
10	Meta-Regression	0.93	0.88	0.86	0.89	0.92	0.90

TABLE A.8: Mean Interval Length of μ for τ sampled from $HN(0.5)$ for UP, HNP and REML

	Interval:	Z	KnHa	Adhoc	t
k	Type	DL	DL	DL	DL
2	Meta-Analyses	2.10	10.55	13.59	13.59
2	Meta-Regression	2.06	3.03	3.42	3.35
5	Meta-Analyses	1.24	1.64	1.82	1.76
5	Meta-Regression	1.24	1.34	1.41	1.37
10	Meta-Analyses	0.86	0.96	1.02	0.99
10	Meta-Regression	0.86	0.89	0.92	0.90
		PM	PM	PM	PM
2	Meta-Analyses	2.10	10.55	13.59	13.59
2	Meta-Regression	2.08	3.05	3.38	3.38
5	Meta-Analyses	1.26	1.66	1.79	1.79
5	Meta-Regression	1.25	1.35	1.39	1.39
10	Meta-Analyses	0.87	0.97	1.01	1.01
10	Meta-Regression	0.87	0.90	0.91	0.91

TABLE A.9: Mean Interval lengths of μ for τ sampled from $HN(0.5)$ for DL and PM

A.4.7 Scenario IV

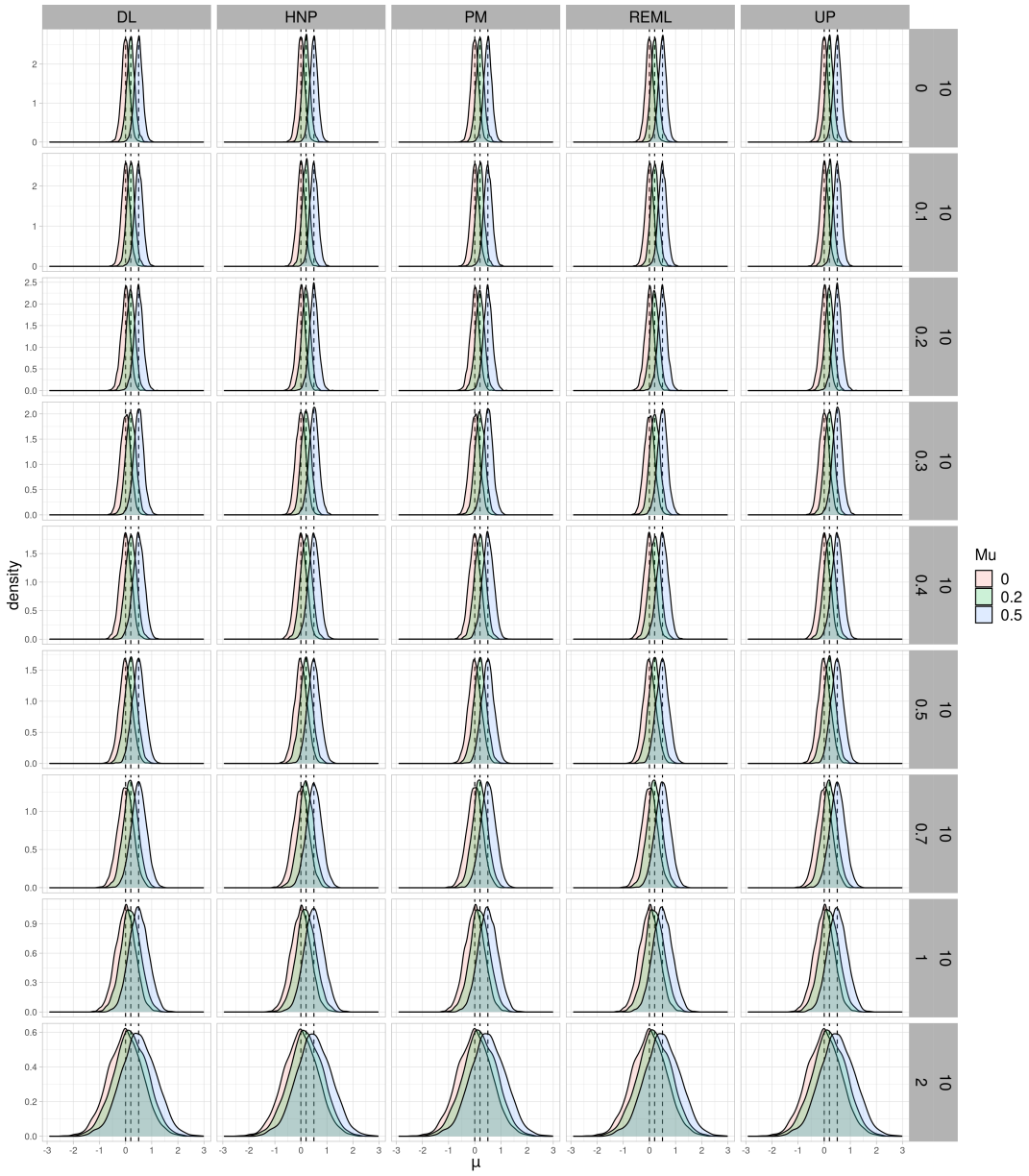

FIGURE A.8: Point estimates of μ split by subgroup for varying τ values, a sample size per subgroup of k=10 and a true varying μ for individual meta-analyses. The true value of μ per subgroup is indicated by color. The gray bar on the right indicates the sample size per subgroup and the respective true τ.

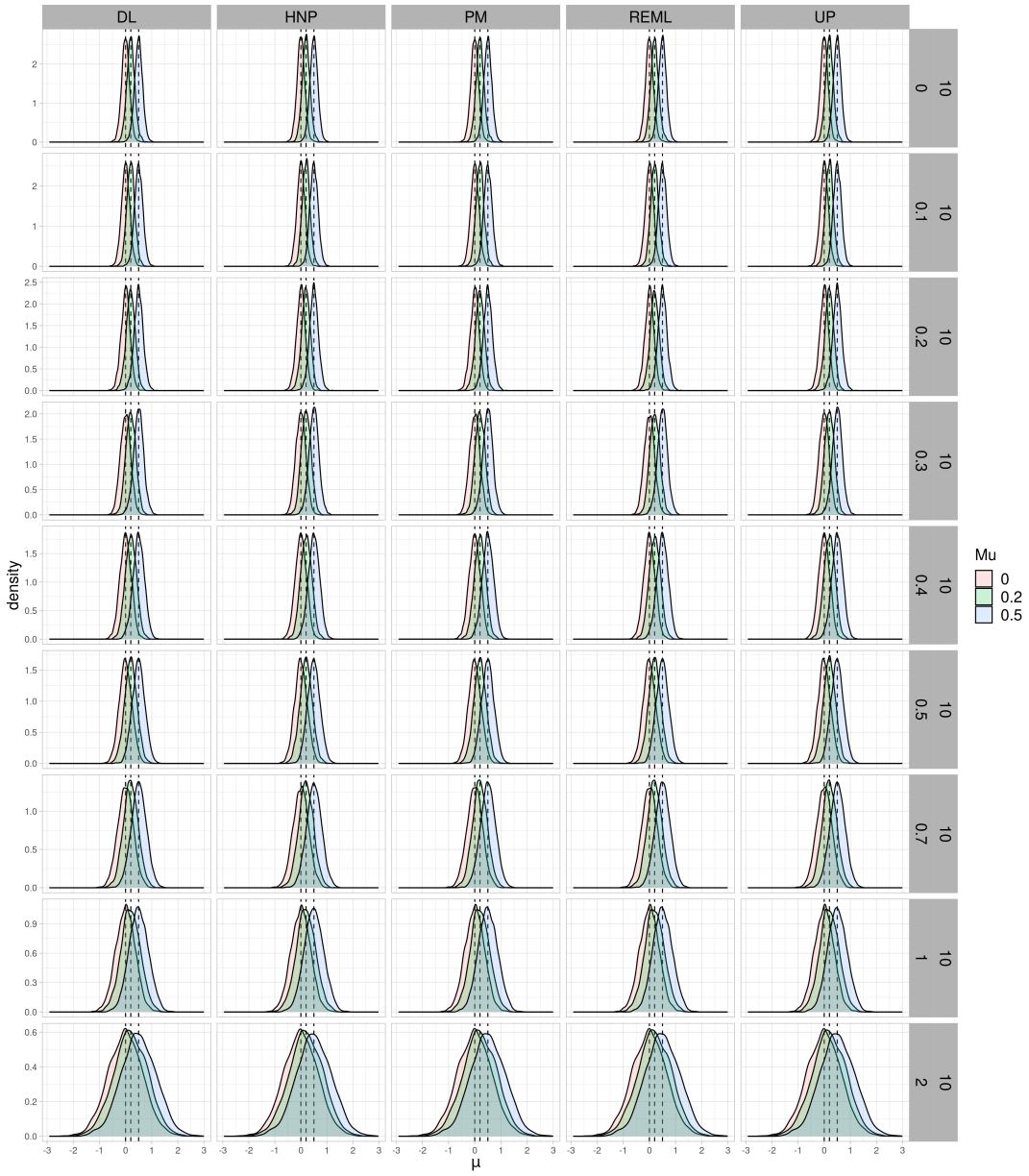

FIGURE A.9: Point estimates of μ split by subgroup for varying τ values, a sample size per subgroup of k=10 and a true varing μ for meta-regression. The true value of μ per subgroup is indicated by color. The gray bar on the right indicates the sample size per subgroup and the respective true τ.

Method	k	τ	UP	HNP	REML	PM	DL
Interval: Shortest							
Meta-Analyses	10	0.3	0.98	0.96			
Meta-Analyses	5	0.3	1.00	0.97			
Meta-Analyses	2	0.3	.	0.98			
Meta-Regression	10	0.3	0.96	0.95			
Meta-Regression	5	0.3	0.97	0.96			
Meta-Regression	2	0.3	1.00	0.97			
Interval: Z							
Meta-Analyses	10	0.3			0.92	0.93	0.93
Meta-Analyses	5	0.3			0.92	0.92	0.92
Meta-Analyses	2	0.3			0.92	0.92	0.92
Meta-Regression	10	0.3			0.93	0.93	0.93
Meta-Regression	5	0.3			0.93	0.93	0.93
Meta-Regression	2	0.3			0.93	0.93	0.93
Interval: KnHa							
Meta-Analyses	10	0.3			0.93	0.93	0.93
Meta-Analyses	5	0.3			0.93	0.93	0.93
Meta-Analyses	2	0.3			0.95	0.95	0.95
Meta-Regression	10	0.3			0.94	0.94	0.94
Meta-Regression	5	0.3			0.94	0.94	0.94
Meta-Regression	2	0.3			0.94	0.94	0.94
Interval: Adhoc							
Meta-Analyses	10	0.3			0.96	0.96	0.96
Meta-Analyses	5	0.3			0.98	0.98	0.98
Meta-Analyses	2	0.3			1.00	1.00	1.00
Meta-Regression	10	0.3			0.95	0.94	0.95
Meta-Regression	5	0.3			0.96	0.95	0.95
Meta-Regression	2	0.3			0.99	0.99	0.99
Interval: t							
Meta-Analyses	10	0.3			0.95	0.96	0.96
Meta-Analyses	5	0.3			0.98	0.98	0.98
Meta-Analyses	2	0.3			1.00	1.00	1.00
Meta-Regression	10	0.3			0.94	0.94	0.94
Meta-Regression	5	0.3			0.95	0.95	0.95
Meta-Regression	2	0.3			0.99	0.99	0.99

TABLE A.10: Coverage probability for μ in the case of varying μ and $\tau = 0.3$. The shown values multiplied by 100 represent percentages.