Meta-Analysis

Bayesian Hierarchical Meta-Analysis: Theoretical Approach and Current Applications

Youhee Kil r0768512
Daria Dementeva r0771521

Spring 2020

Contents

1 Introduction

Meta-analysis is a set of quantitative procedures that are employed to perform statistical analyses on the sample of studies (Cooper, 2016, 9). Consequently, in meta-analysis the studies are regarded as complex observations, which cannot be studies in isolation from each other (Cooper, 2016, 13). All things considered, the prime task of meta-analysis is to integrate relatively similar but at the same time distinct studies in order to get the cumulative quantitative evidence on subject matter (Cooper, 2016, 23). As a method for combing information from several parallel data sources, meta-analysis is enormously connected to hierarchical modeling as stated in Gelman et al. (2014). Meta-analytic problems and data do have a natural hierarchical structure per se. Additionally, it might be unrealistic to assume the homogeneity of studies in meta-analysis and, thus, hierarchical approach seems more plausible to implement.

In general, Bayesian modeling offers more opportunities in hierarchical meta-analysis than a classical hierarchical approach Lesaffre and Lawson (2012). Thus, the Bayesian approach to hierarchical modeling in meta-analysis will be elucidated in this paper.

In Bayesian hierarchical modeling applied to meta-analysis one does not assume that several parameters from all investigated studies are completely independent from each other. This idea had been implicitly introduced by the founding father of meta-analysis Gene Glass. As McElreath (2015) points out, Bayesian hierarchical modeling, regardless whether it is applied in meta-analysis or in any other research field, is not memoryless. In fact, one can learn about the parameter in one study, based on the parameter derived from another study. Furthermore, one can learn not only about the individual parameters of investigated studies, but also the parameters of the whole population of studies.

In essence, the prime task of this paper is to disentangle both theoretically and practically how Bayesian hierarchical modeling is implemented in meta-analysis.

This paper contributes to the general meta-analytic scholarship by focusing specifically on Bayesian hierarchical modeling and its current applications in meta-analysis. Furthermore, we attempt to re-analyze the data on conscientiousness and medication adherence and rare events (Section 4) using Bayesian hierarchical meta-analysis. Moreover, we implement our analyses in newly released package *MetaStan*.

This paper is organized as follows. In Section 2 we briefly discuss hierarchical meta-analysis in order to elucidate the hierarchical nature of meta-analytic problems and sketch the potential difficulties with traditional hierarchical approach in meta-analysis. In Section 3 we present Bayesian approach to hierarchical modeling and the application to meta-analysis. In Section 4 we present the current application of Bayesian Hierarchical meta-analysis. and the re-analyses of the available meta-data using Bayesian hierarchical meta-analysis, followed by a short conclusion.

2 Hierarchical Meta-Analysis: Definitional Aspects, Applications and Complications

In this Section we briefly discuss hierarchical meta-analysis in order to elucidate the hierarchical nature of meta-analytic problems and sketch the potential difficulties with traditional hierarchical approach in meta-analysis.

The hierarchical models is such a natural technical tool for implementing the meta-analysis that it may be said to have been waiting for hierarchical models to come along (Draper (1995)). For instance, Higgins and Green (2011) addressed when random-effects are used to represent unexplained variation

in effect estimates among studies, especially such hierarchical framework is applicable. However, there is no common agreement concerning the way to label the analyses, in which the data of multilevel, hierarchical, nested or clustered nature is used as stated in Van Den Noortgate and Onghena (2003), and further supported in Lesaffre and Lawson (2012), McElreath (2015). For example, the notions of "mixed effects model", "multilevel model", "frequentist hierarchical model" are treated as synonymous since in these models there is a discrimination between fixed and random effects (Lesaffre & Lawson, 2012, 227). Eventually, Lesaffre and Lawson (2012) proceed with the term "hierarchical model". Same is concluded in McElreath (2015), but he opts for the term of "multilevel model" in the end. However, a different perspective is presented in Van Den Noortgate and Onghena (2003). Hierarchical model in meta-analysis is considered to be different from a typical multilevel - also known as hierarchical, random effects, varying effects, or mixed model analysis since it does not use the raw data. Further, we would like to use the terms interchangeably in our paper.

In the following context, hierarchical models in meta-analysis that account for non-independent sampling errors and/or true effects are more advantageous as claimed in Higgins and Green (2011):

- The imprecision of the variance estimates of treatment effects within studies is allowed.
- The imprecision in the estimated between-study variance estimate and τ^2 is allowed

For the sake of simplicity, we present the following notation for hierarchical model in meta-analysis. In hierarchical meta-analysis, it all starts from the random-effects model as mentioned in Harrer, Cuijpers, and Furukawa (2010)

$$\hat{\theta}_k = \mu + \epsilon_k + \zeta_k \tag{1}$$

with θ_k the true effect size in study k, μ the mean true effect size, and ϵ_k sampling error, ζ_k is the between-study heterogeneity. Equation (2) is level 1 and equation (3) is level 2.

$$\hat{\theta}_k = \theta_k + \epsilon_k, \hat{\theta}_k \sim N(\theta_k, \sigma_k^2) \tag{2}$$

$$\theta_k = \mu + \zeta_k, \theta_k \sim N(\mu, \tau_k^2) \tag{3}$$

Different estimation procedures such as ML, REML are commonly used for analyses of hierarchical models to estimate the above mentioned parameters. We prefer to skip many methodological, theoretical and technical peculiarities of hierarchical modeling in meta-analysis given that it is beyond the scope of this paper.

We are generally interested making inferences on many parameter θ_1, ... , θ_i measured on N 'units' in hierarchical models. There are three possible assumptions about the θ's as elucidated in Albert, Spiegelhalter, and Best (2005):

- the θ's are *identical*
- the θ's are *independent* and entirely unrelated
- the θ's are *exchangeable* or similar which is a compromise between independent and identical relationships.

Hierarchical modeling in meta-analysis requires implementing complicated computational commands in sophisticated software. The bulk of current methodological evidence in meta-analysis focuses on

hierarchical modeling often with Bayesian implementation (Higgins, 2011). Van Den Noortgate and Onghena (2003) stated that the hierarchical approach does not necessarily provide better results than the the traditional conditional or random-effects procedures, but, for the large flexibility, still many meta-analysts feel more confident and comfortable using hierarchical analysis as claimed in Van Den Noortgate and Onghena (2003). In addition, there are many shortcomings of hierarchical approach mentioned in Van Den Noortgate and Onghena (2003). We believe Bayesian estimation procedures will solve the shortcomings the approach briefly described above as Lesaffre and Lawson (2012) argue that for Bayesian estimation it is not required to distinguish between random and fixed effects. Bayesian approach assumes that all parameters are random (Lesaffre & Lawson, 2012, 227).

In the next Section we present the general notation for Bayesian hierarchical model. Further in next Section we elaborate on how it is theoretically employed in the context of meta-analysis. We illustrate our elaboration with the example adopted from McElreath (2015).

3 Bayesian Hierarchical Model in Meta-Analysis: Explanation of Inference Procedure

3.1 Background of Bayesian Hierarchical Modeling

In the previous Section, we established how classical hierarchical models are conceptualized and further applied in the context of meta-analytic problems.

In this Section we reconstruct and elucidate the Bayesian approach to hierarchical modeling and further expand it to meta-analysis. We primarily base our reconstruction on the theoretical approaches proposed by Gelman et. al (2014), Kruschke (2015), McElreath (2015). We adopt the mathematical notation presented in Gelman et. al (2014).

As in hierarchical modeling, we assume that each group j with a set of subjects i holds a parameter of interest θ_j. In Bayesian approach, it is essential to assume that each θ_j is dependent on another θ_j. Kruschke (2015, p. 221) denotes it as the chain of dependencies among the parameters. From Bayesian perspective, we view each θ_j as a derived sample from a general population distribution (Gelman et. al, 2014, p.101). Krushke (2015, p. 222) argues that the advantage of Bayesian hierarchical model is that the estimates of individual parameters are communicated by the data from all other items. This makes the parameters dependent on each other. At the same time, the better estimates of all parameters are produced when there are dependencies among the parameters (Krushke, 2015, p. 222).

We consider the following formulation of Bayesian hierarchical model (Gelman et. al, 2014, p.104):

- a set of $j = 1, ..., J$;
- each j corresponds to observed data y_j;
- each j probability is described by θ_j (Kruschke, 2015);
- likelihood is given by $p(y_j|\theta_j)$.

Let us present the example to illustrate the above mentioned from McElreath (2015). McElreath (2015, p. 355) presents the following. Suppose, we have a robot that would estimate the waiting time (y_j) for a cup of coffee to be served at two coffee shops ($j = 1, 2$). When it enters the first coffee shop, it uses weak prior for mean and standard deviation of waiting time. When coffee arrives, the robot estimates

the observed serving time and, subsequently, updates the prior distribution. When a robot enters the second coffee shop, it might use the information derived from the first one to evaluate the mean and standard deviation of waiting time at the second place. However, all coffee shops are not fully alike, but do have some resemblances. In this case, a robot reports the parameter for each coffee shop and, simultaneously, the mean and standard deviation of the population of coffee shops. When more waiting times are observed, a robot upgrades both individual and population parameters. This example exhibits a hierarchical structure as well: (1) general (perhaps, infinitely large) population of coffee shops, (2) the number of coffee shops robot actually visited.

Thus, based on the observed data, we can estimate the population distributions of all θ_j, although, as argued, not all θ_j can be observed simultaneously (Gelman et. al, 2014, p.101). Thus, through dependency among the parameters we can estimate unobserved θ_j. Given the hierarchical constitution of the problem, observed outcomes are modelled conditionally on parameters of interest, which, in turn, are expressed in probabilistic fashion in terms of the other parameters, hyperparameters (Gelman et. al, 2014, p.101). In the presented example, hyperparameters are the population parameters of waiting time (mean and standrard devitation) of coffee shops that govern the individual parameters, but, at the same time, are updated when more data are observed. Approaching the hierarchical problem from the Bayesian perspective requires the elaboration on how to construct a prior distribution for a) point estimates of population parameters and respective hyperparameters, b)a joint posterior distribution.

We might unrealistically presume that prior distribution is fixed in the sense that θ is distributed with known mean and standard deviation. For instance, all coffee shops have identical mean and standard deviation of waiting time.

On the contrary, we may use approximations from available previous evidence to construct a prior distribution (Gelman et. al, 2014, p.101). For example, there might be historical evidence from operations research on mean and standard deviation of waiting time in coffee shops. However, the historical studies and the data that are used for constructing likelihood are not compatible in terms of conditions under which the studies had been conducted (e.g, the designs of the studies, time-specific conditions). Thus, the correction for systematic differences between studies should be employed (Gelman et. al, 2014, p.102).

Gelman et. al (2014) give the following reasons why, in general, there are problems with straight estimation of prior distribution from available data (p. 103-104). First, we encounter the overestimation of precision since the data are utilized twice in estimation process: 1) calculation of prior distribution, 2) estimation of each θ_j. For instance, we use the parameters of waiting time in each coffee shop to estimate 1) individual $\theta_{coffeeshop}$, 2) updated population parameters. In turn, both are combined to produce prior distribution. Second, the point estimates for prior distribution parameters are subjective in the sense that the posterior uncertainty can be ignored.

Still, as Gelman et. al (2014, p. 104) argue, the logical grounds of information combination are still satisfied, regardless of the above mentioned problems. The existing data are used to estimate the parameters of prior distribution, but we assign the probabilities on all prior parameters and θ_j. After that we apply Bayesian estimation on the joint distribution of all included parameters. Joint distribution of θ_j requires the implementation of exchangeability. For instance, we treat each $\theta_{coffeeshop}$ exchangeably in the sense that all coffee shops are not fully alike, but do have some resemblances between each other. We discuss the notion exchangeability in great detail in the next section.

3.2 The Concept of Exchangeability

The concept of exchangeability is one of the most crucial in the hierarchical Bayesian models applied to the problems of Meta-Analysis. We will expand on it later in this paper. In general, exchangeability is the

symmetry between the parameters in prior distribution expressed in a probabilistic manner (Gelman et. al, 2014, p. 104). This symmetry is assumed when it is almost impossible or unnecessary to discriminate between any of θ_j. In particular, no ordering or classification in groups of θ_j can be performed. When this is the case, the parameters are said to be exchangeable in their joint distribution if the probabilities of θ_j, $p(\theta_1, ..., \theta_J)$, are unalterable to any discrimination. Simply put, the indices of θ_j, whatever value j takes, are just chosen labels, and no plausible grouping can be employed (McElreath, 2015, p.387). All coffee shops are not fully alike, but do have some resemblances between each other. To illustrate, a robot cannot differentiate between coffee shop it visits. We cannot group coffee shops plausibly and, therefore, the grouping of $\theta_{coffeeshop}$ is implausible, too.

The exchangeable distribution is controlled by the unknown vector of hyperparameters ϕ. Going back to our example, this unknown vector of hyperparameters ϕ is the population parameters of waiting time (mean and standard deviation). By exchangeable distribution we consider the distribution that assumes that each θ_j is independent identically distributed sample (Gelman et. al, 2014, p. 105). Thus, the probability of θ given hyperparameter ϕ is the product of all θ_j given ϕ. For example, the probability of $\theta_{coffeeshop}$ given the population parameters of coffee shops $(\mu_{waitingtime}, \sigma_{waitingtime})$ is the product of all θ from all coffee shops given $\mu_{waitingtime}$ and $\sigma_{waitingtime}$. Mathematically, it is summarized as follows:

$$p(\theta|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi) \tag{4}$$

Since hyperparameter ϕ is unknown, we must integrate it out to derive the distribution for θ (Gelman et. al, 2014, p. 105):

$$p(\theta) = \int \left(\prod_{j=1}^{J} p(\theta_j|\phi) \right) p(\phi) d(\phi) \tag{5}$$

In practice, however, full exchangeability is rarely encountered. Normally, we suppose, if necessary, secondary information on observations can be derived by researcher. Gelman et. al (2014) claim that when additional information is available and the grouping can be accomplished, hierarchical model is still a valid choice. In general, the characteristics of the categories or groups are unknown. In that regard, group features become exchangeable, and the prior distribution can be derived (Gelman et. al, 2014, p. 106). In this regard, we are able to classify coffee shops (for example, price level, location, number of employed personnel), but these features, therefore, become exchangeable, and the prior distribution can be derived.

Second, the observed data for each observed unit, y_i might exhibit the extra information expressed in the variable, or some covariate, x_i. Given what has been stated, (y_i,x_i) are considered exchangeable and it is possible to estimate (y_i,x_i) jointly or conditionally (Gelman et. al, 2014, p. 106). These assumptions seem especially valid for meta-analytic problems in the presence of moderator variables as we explain later in the paper.

3.3 Full Bayesian Hierarchical Model

The inferential peculiarity of the Bayesian hierarchical model, as we stated previously, is that θ is controlled by the unknown vector of hyperparameters ϕ, which has its own hyperprior distribution. Given what is stated above, the observed data is affected by ϕ through θ, because θ is supervised by ϕ (Gelman et. al, 2014, p. 107). Further, we obtain the posterior distribution for (ϕ, θ). As we noted earlier, when more waiting times are observed, a robot upgrades both individual and population

parameters. Given that, we obtain the joint posterior distribution for the parameters on the coffee shop level as well as on the population level.

First, mathematically, the joint prior distribution is given by (Gelman et. al, 2014, p. 107):

$$p(\phi, \theta) = p(\phi)p(\theta|\phi) \tag{6}$$

further, the joint posterior distribution is obtained by:

$$p(\phi, \theta|y) = p(\phi, \theta)p(y|\theta) \tag{7}$$

Usually hyperparameters ϕ are not known, so the hyperprior distribution in most cases is vague or weakly informative. The specification of hyperprior distribution depends on what prior distribution θ is assigned.

Once posterior predictive distribution is obtained, it is possible to obtain two posterior predictive distributions, depending what research question is of prime interest. Gelman et. al (2014, p. 108) claim that it is possible to obtain $\tilde{y}$ given θ_j. Secondly, the distribution of observations $\tilde{y}$ given the future θ_j can be obtained as well.

In general, Gelman et. al (2014) suggest using the inferential strategy for hierarchical models when conjugate distributions can be produced. It facilitates the derivation of approximate estimates and sets the ground for more complicated Bayesian analyses when more involved analytical problems are of interest.

3.4 Bayesian Hierarchical Model in Meta-Analysis

As we outlined earlier, hierarchical modeling is enormously related to the problems considered in meta-analysis. Hierarchical Bayesian modeling in Meta-Analysis is implemented in order to combine empirical evidence from a particular set of studies and improve on the estimation of the parameters of interest, θ_j. Likewise, the integral part is to correct the estimation of mean and variance of the effects over the set of studies. That is, we are interested in the update of parameters in the whole population of studies. In the section "Background of Bayesian Hierarchical Modeling" we established the notation of hierarchical problem. Let us adapt it to the context of meta-analysis:

- a set of studies $j = 1, ..., J$;
- each study, j, has a study-specific number of observed y_j from units $i = 1, ..., I$;
- each j probability is described by θ_j;
- the likelihood function is defined by $p(y_j|\theta_j)$

For the sake of simplicity, we discriminate between four key steps in hierarchical Bayesian modeling in Meta-Analysis. First, one needs to define the parameters of interest, θ_j, for each study, j, in order to determine the prior distribution. Second, Determination of weakly informative priors for the unknown vector of hyperparameters ϕ is essential to determine joint prior distribution $p(\phi, \theta) = p(\phi)p(\theta|\phi)$. Third, it is necessary to define the likelihood estimation. Fourth, the application of exchangeability is deemed crucial to proceed with construction of prior distribution for the parameters.

Let us know expand on each step we listed above.

Defining θ_j. Typically, one is interested in the estimation of mean and variance of posterior distribution of effect sizes, the averaged effect over all studies (Gelman et. al, 2014, p.126). Given that, we define θ_j as the effect size, obtained from study j. It is also possible to obtain each of θ_j from either included observed study j or unobserved study j. Therefore, the parameters to determine the prior distribution are the mean and the variance of effect sizes. Taking into account the hierarchical nature of the problem, one needs to set weakly informative priors on hyperparameters, the paramaters of population of studies, which, in turn, govern the parameters of prior distribution, mean and the variance of effect sizes.

Likelihood function. It is argued that Bayesian hierarchical models in meta-analysis can incorporate on the systematic basis the evidence from studies with distinct initial designs (Schmitz et al. (2020)). Henceforth, we believe that the construction of likelihood function strongly depends on the nature of collected sample of studies and the prime parameter of interest θ_j. For example, Schmitz et al. (2020) assumed normal likelihood model for health-related quality of life, where the parameter of interest, θ_j, is the slope of health-related quality of life given a set of multiple covariates.

As a second example, Pibouleau and Chevret (2013) assumed a binomial likelihood model for implantable medical devices evaluation, where the parameter of interest, θ_j, is study-specific success probability across the sample on the logit scale. Further in this paper we show in detail how likelihood function was constructed for two applications of Bayesian hierarchical meta-analysis.

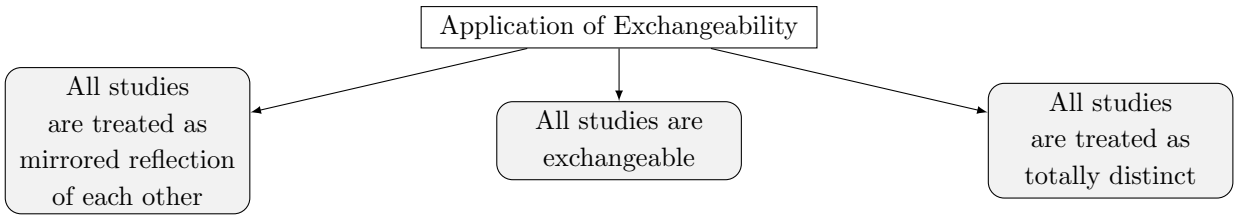
Application of Exchangeability. The chief task is to test the null hypothesis of no effect within the studies under investigation. Gelman et. al (2014, p.125) argue that sensible estimation of parameters is possible under the assumption that all included studies are somewhat akin to each other. In general, we can distinguish between three ways of treating studies under meta-analytic inquiry (Gelman et. al, 2014, p.125-126).

One can think of the included studies as of mirrored reflection of each other. Simply, studies are identical. This approach presumes that all units i are from exactly the same population and exhibit identical y_i. Given that, one might unrealistically presume that prior distribution is fixed in the sense that θ_j is distributed with known mean and standard deviation. We think that this approach is unrealistically simplified.

On the contrary, one can relate to the included studies as totally distinct and unrelated to each other. By following this approach, we reject the possibility to make any inferences about particular θ_j of study j, based on other θ_j, estimated from other studies j. Given that, there might be unrealistic assumption that prior distribution is not fixed at all and it should be as noninformative as possible.

Third approach is the compromise between the two above mentioned and is originated from the concept of exchangeability. To put it simply, the differences between study-to-study are possible. Given that the studies are neither completely the same nor completely unalike. Like in previously mentioned example, all coffee shops are not fully alike, but do have some resemblances between each other. In meta-analysis, one does assume exchangeability in the sense that no ordering or grouping of θ_j, or size effects, derived from studies, can be employed. The parameters are said to be exchangeable in their joint distribution if the probabilities of size effects, θ_j, $p(\theta_1, ..., \theta_J)$, are unalterable to any grouping. However, the hierarchical Bayesian modeling is possible even if the observed data for each observed unit, y_i might exhibit the extra information expressed in the variable, or some covariate, x_i. Is the notion of exchangeability still applicable? We suppose, that, given what has been stated earlier in this paper, (y_i,x_i) are considered exchangeable, too. For example, this extra information can be reflected in moderator variable (e.g., design of studies, year of publication), which further can be used as a covariate in hierarchical Bayesian regression model.

The taxonomy of approaches is graphically summarized below.



Further in this paper, we present the methodological evidence of how Bayesian hierarchical modeling has been applied in Meta-Analysis in practice.

4 Applications

4.1 Conscientiousness and Medication Adherence: A Meta-Analysis from Molloy et al. (2014)

The first application of Bayesian hierarchical meta analysis we want to present is based on the data introduced in Molloy et al. (2014). It has been revealed that medication adherence is positively related to the conscientiousness. Overall, there are 16 studies in the sample size with total of 3,476 observation. It should be noted that the random effects meta-regression has been implemented to estimate the size effects. In this Section we present our re-analysis attempt, using Bayesian hierarchical meta-analysis employed in *metafor, brms* from open-source software R.

When the studies report data as correlations, we usually do not use the correlation coefficient itself as the effect size. Given that the sample size is relatively small (16 studies), sampling distribution is skewed. Furthermore, we consider the sampling distribution only approximately normal. Thus, we transform the correlation using the Fisher's Z_r transformation. Then we convert the summary values back to correlations for presentation as suggested in Borenstein, Hedges, Higgins, and Rothstein (2011). In this case, we used the *escalc* function in *metafor* package which calculates various effect sizes or outcome measures including sampling variances.

The transformation from sample correlation r to Fisher's Z_r is given below:

$$z = 0.5 \times ln(\frac{1+r}{1-r}) \tag{8}$$

The conversion of the Fisher's Z_r value back to correlation is given as outlined below:

$$r = \frac{e^{2z}-1}{e^{2x}+1} \tag{9}$$

Netx, the standard error was computed as follows:

$$SE_z = \sqrt{V_z} \tag{10}$$

The data, which is further used for Bayesian hierarchical meta-analysis is presented in Table 1. It should be mentioned, that the data are already transformed as outlined above.

	Study	Yi	SEi
1	Axelsson et al. (2009)	0.18922664	0.09712859
2	Axelsson et al. (2011)	0.16343992	0.03661260
3	Bruce et al. (2010)	0.35409253	0.13867505
4	Christensen et al. (1999)	0.33164711	0.09805807
5	Christensen Smith (1995)	0.27686382	0.12038585
6	Cohen et al. (2004)	0.00000000	0.12700013
7	Dobbels et al. (2005)	0.17682002	0.07647191
8	Ediger et al. (2007)	0.05004173	0.05564149
9	Insel et al. (2006)	0.26610841	0.13483997
10	Jerant et al. (2011)	0.01000033	0.03608439
11	Moran et al. (1997)	-0.09024419	0.13736056
12	O'Cleirigh et al. (2007)	0.38842310	0.10660036
13	Penedo et al. (2003)	0.00000000	0.09407209
14	Quine et al. (2012)	0.15114044	0.04327423
15	Stilley et al. (2004)	0.24477411	0.08032193
16	Wiebe and Christensen (1997)	0.04002135	0.12700013

Table 1: Data set: Medication Adherence Study

As we indicated in theoretical section, the hierarchical model can be formalized as follows below:

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

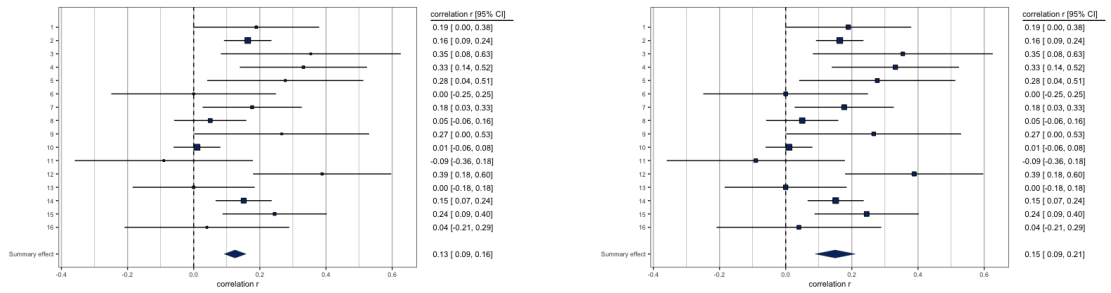
In hierarchical meta-analysis we usually assume that the true effects are normally distributed. Therefore, in this case, y_i is drawn from normal distribution which is centered on that study's effect size θ_i and has standard deviation equal to the study's observed standard error σ_i as claimed by Vuorre (2016). In addition, the parameters θ_i are drawn from a normal distribution with hyperparameters (μ, τ) in this case. In Gelman et al. (2013) it is stated that θ_i's are conditionally independent given (μ, τ) and hierarchical model also permits the interpretation of the θ_i's as a random sample from a shared population distribution.

Bayesian Hierarchical Meta-Analysis (BHMA) will be conducted with *Stan, brms* packages. The Bayesian model in current meta-analytic problem can be generally summarized as:

$$y_i \sim N(\theta_i, \sigma^2)$$
$$\theta_i \sim N(\mu, \tau^2)$$
$$(\mu, \tau^2) \sim p(.), \tau^2 > 0$$

which is similar to hierarchical model but, there is a remarkable difference between two models such as $(\mu, \tau^2) \sim p(.), \tau^2 > 0$. Assigning prior distribution (p(.)) for μ and τ^2 and restricting τ^2 to positive is the first step of implementing Bayesian hierarchical modeling.

For the prior distribution, we assigned a prior distribution to τ. According to the information from the meta-analysis random fixed model, We set μ and τ as follows:



(a) Fixed Effects Model (b) Random Effects Model

Figure 1: Forest plots of the associations between conscientiousness and medication adherence

$$\mu \sim N(0, 1)$$
$$\tau \sim HalfCauchy(0, 0.2)$$

Gelman et al. (2013) stated that a weakly informative prior which is a middle ground between a fully informative and uninformative (flat) prior distribution. A prior does not necessarily have to reflect historical or empirical data, but can be chosen based on desirable mathematical properties as suggested by Williams, Rast, and Bürkner (2018). However, finding a suitable prior distribution for τ^2 is important for several reasons which are elucidated in Williams et al. (2018). As Mila and Ngugi (2011) claimed the results of the Bayesian meta-analysis will be similar to the results of classical model when we assume random-effect model with a normal distribution and non-informative prior (i.e., an uniform distribution). In this case, the reason why we assumed $\tau \sim$ Half Cauchy(0, 0.2) is that average effect size from Figure 1 is close to 0.2 (to be exact, this is 0.144) and half-Cauchy prior is used with the Gelman's support of use of the half-Cauchy prior for the hierarchical model. The half-Caucy prior is a sensible default choice for hierarchical model and it performs a great deal of shrinkage as claimed in Gelman et al. (2013).

The Bayesian approach provides a mechanism for incorporating prior knowledge into an analysis (Enders (2010)), therefore, in this case, we formulated a prior distribution by using information such as the average effect size from the previous hierarchical analysis.

The Figure 2 displays the each θ_i for posterior distribution. The mean and 95% CrI limits are displayed at the right side of the plot. One of advantages of performing Bayesian method on meta-analysis is that uncertainty of the results can be easily measured with the 95% CrI as recommended in Mila and Ngugi (2011). As we can compare with Figure 1 (Forest plot of FE and RE), the resulting posterior distribution is not equivalent to the Figure 1, for example, p-value and range of credible intervals of individual studies are different when we compare.

Model	**Estimate**	**se**	**zval**	**pval**	**ci.lb**	**ci.ub**
Fixed Effect	0.125	0.017	7.364	<.001	0.092	0.158
Random Effect	0.150	0.032	4.750	<.001	0.088	0.212
Multilevel (w/ rma.mv)	0.144	0.069	2.076	0.038	0.008	0.280
BHMA	**Estimate**	**Est.Error**	**l-95% CI**	**u-95% CI**		
Group-Level Effects	0.099	0.038	0.039	0.183		
Population-Level Effects	0.149	0.035	0.081	0.220		

Table 2: Summary of Models

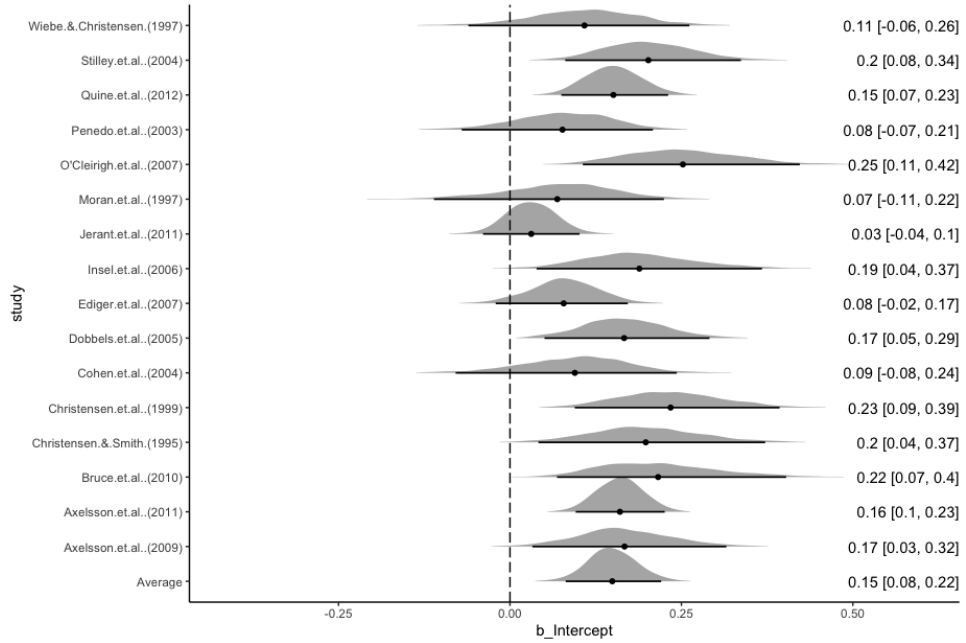


Figure 2: Forest Plot for Bayesian Hierarchical Meta-analysis model

Table 2 represents Bayesian hierarchical model with displayed group-level effects and population-level effects. Every parameter is summarized using the mean ('Estimate') and the standard deviation ('Est.Error') of the posterior distribution as well as two-sided 95% credible intervals ('l-95% CI' and 'u-95% CI'). For hyperparameters (population-level effects) the 95% credibility interval for the average size effect is between 0.081 and 0.220 (Borenstein et al. (2011)). To sum up, it means that we are 95% that the average size effect of conscientiousness on medical adherence is between 0.081 and 0.220 for the whole population of studies. For individual parameters (group-level effects, θ_j for study j), we are 95% certain that the the average size effect of conscientiousness on medical adherence is between 0.039 and 0.183. It needs to be mentioned that additional computational procedures should be employed to check for convergence. As for now, this goes beyond the scope of this paper.

4.2 Random-Effects Meta-Analysis of Few Studies Involving Rare Events: Günhan et al. (2020)

The aim of the second application is to present a model of different distribution. That is, a Binomial-Normal Hierarchical Model applied in meta-analytic research. Moreover, we utilize new package called *MetaStan* in R published last month for Bayesian hierarchical modeling in meta-analysis.

In this application, we are using a binomial-normal hierarchical model(BNHM) with Bayesian approach. The data set is presented in Table 3. Data contain the results from 13 trials examining the efficacy of the BCG-vaccine against tuberculosis. from a package *MetaStan* in R and is referred to as *"dat.Berkey1995"* The package *MetaStan* will be used to conduct Bayesian hierarchical meta-analysis via programming language Stan. The package includes computations capabilities to estimate binomial-normal hierarchical models and option to use weakly informative priors for the heterogeneity and the treatment effect parameters as can found in Günhan et al. (2020).

	Trial	nt	rt	rc	nc	Latitude
1	1	4	123	11	139	44
2	2	6	306	29	303	55
3	3	3	231	11	220	42
4	4	62	13598	248	12867	52
5	5	33	5069	47	5808	13
6	6	180	1541	372	1451	44
7	7	8	2545	10	629	19
8	8	505	88391	499	88391	13
9	9	29	7499	45	7277	27
10	10	17	1716	65	1665	42
11	11	186	50634	141	27338	18
12	12	5	2498	3	2341	33
13	13	27	16913	29	17854	33

Table 3: Application 2: Data

Data from multiple comparative trials involve binary outcomes, which, in turn, are very often pooled in systematic reviews and meta-analyses. Predominantly, this type of data is presented as the series of 2-by-2 tables (Günhan et al. (2020)). In this application, the primary interest is the pooled odds ratio which is very popular measure of size effect (or treatment effect) for meta-analysis when we consider binomial outcome variable. In medical field, treatment effect is often referred to the effect size in meta analysis and often it assumed to refer to odds ratio, risk ratios, or risk differences as can be found in Borenstein et al. (2011). Besides, we put existing methods with binary outcomes into the general framework of multilevel modeling which is accepted statistical analysis tool of hierarchical data as claimed by Turner, Omar, Yang, Goldstein, and Thompson (2000). Furthermore, Bayesian methods will be presented as one possible approach.

To deal with data sparsity present in the meta-analysis of few studies with rare events, Gunhan (2020) suggests the use of weakly informative prior for the treatment effect parameter in a fully Bayesian. Additionally, it has been noted that it is a contrast-based model meaning that relative treatment effects are assumed to be exchangeable across trials. This assumption originates from the notion of exchangeability, presented in Section 3

Y_i is defined as the odds ratio for the ith study:

$$Y_i = \frac{trt_i(ctrlN_j - ctrl_j)}{ctrl_j(trtNj - trt_j)} \tag{11}$$

For s_j^2, we can get the approximate sampling variance of θ_j:

$$s_j^2 = \frac{1}{trt_j} + \frac{1}{trtN_j - trt_j} + \frac{1}{ctrl_j} + \frac{1}{ctrlN_j - ctrl_j} \tag{12}$$

In the binomial normal hierarchical model, the treatment for each trial is θ_i, $i = 1, ...13$ and treatment arm is $j = 0, 1$. The even counts r_{ij} are modeled using binomial distribution.

$$\begin{aligned} r_{ij} &\sim Bin(\pi_{ij}, n_{ij}) \\ logit(\pi_{ij} &= \mu_i + \pi_i x_{ij}) \\ \theta_i &\sim N(\theta, \tau^2) \end{aligned}$$

with the μ_i is fixed effect in each study i, θ is the mean treatment effect, and τ is the heterogeneity in treatment effects between trials. Bayesian methods are commonly used to fit the BNHM. In Bayesian approach, prior distributions for parameters θ, μ_i, and τ are needed to be specified. In this case, the a normal prior for the log-odds ratio is suggested as mean 0 and standard deviation 2.82. Further details of getting a normal prior can be checked in Günhan et al. (2020). In addition, we will use half-normal prior with scale 0.5 for τ. Günhan et al. (2020) stated that a half-normal (0.5) prior captures heterogeneity values for log odds ratio will therefore be a sensible choice in many applications.

Therefore, our BNHM model in this case :

$$\theta \sim N(0, 2.82^2)$$
$$\tau \sim HalfNormal(0.5)$$

with a method = *BNHM1* as Model 4 from Jackson el al (2018). In *MetaStan* there is a function called *'meta_stan'* which is used to fit a meta-analytic model using programming language Stan. One of arguments constructed in the function *'meta_stan'* is 'model', and it has the following available options: 1) 'FE' (fixed-effect model using binomial likelihood), 2) 'BNHM1' (Model 4 from Jackson et al (2018)), 3) 'BNHM2' (Model 2 from Jackson et al (2018)), 4) 'Beta-binomial'. Default is 'BNHM1'. The 'BNHM2' and 'BNHM1' appear very similar, but the difference between these models is most clearly seen from the bivariate representation of the log odds of event in the control and treatment groups of the i_th study as evidenced by Jackson, Law, Stijnen, Viechtbauer, and White (2018)).

Parameter	**Rhat**	**n_eff**	**mean**	**sd**	**2.5%**	**50%**	**97.5%**
mu[1]	1.0	5616	-3.0	0.3	-3.6	-2.9	-2.4
mu[2]	1.0	3723	-3.0	0.2	-3.4	-3.0	-2.6
mu[3]	1.0	4346	-3.6	0.3	-4.2	-3.6	-3.1
mu[4]	1.0	5788	-4.6	0.1	-4.8	-4.6	-4.5
mu[5]	1.0	6082	-4.9	0.1	-5.2	-4.9	-4.7
mu[6]	1.0	7216	-1.5	0.1	-1.6	-1.5	-1.4
mu[7]	1.0	4913	-5.0	0.2	-5.5	-5.0	-4.6
mu[8]	1.0	5685	-5.2	0.0	-5.2	-5.2	-5.1
mu[9]	1.0	5711	-5.3	0.1	-5.6	-5.3	-5.1
mu[10]	1.0	4933	-3.9	0.1	-4.2	-3.9	-3.6
mu[11]	1.0	5351	-5.4	0.1	-5.5	-5.4	-5.3
mu[12]	1.0	4804	-6.5	0.4	-7.3	-6.5	-5.8
mu[13]	1.0	6527	-6.4	0.1	-6.7	-6.4	-6.2
theta	1.0	1185	-0.7	0.2	-1.1	-0.7	-0.4
tau	1.0	1354	0.6	0.1	0.4	0.6	0.9

Table 4: Posterior Summary Statistics

	Mean	**2.5%**	**50%**	**97.5%**
Mean treatment effect	-0.75	-1.15	-0.75	-0.36
Heterogeneity stdev	0.60	0.37	0.58	0.93

Table 5: Summary of Meta-analysis using MetaStan

The posterior summary statistics for each parameter is presented in Table 4. In Table 5, the estimates of the mean treatment effect parameter θ is -0.75, with 95% credibility interval (-1.15 and -0.36). Mean of heterogeneity in treatment effects between trials τ is 0.60 with 95% credibility interval (0.37 and 0.93).

Assuming homogeneity in biomedical sciences is often unrealistic for meta-analyses, so that we used hierarchical meta-analysis. However, additional issue with binary outcomes is that sometimes there are only few or no events are observed. To deal with this case in the meta-analysis, it is proposed to use Bayesian estimation of hierarchical model with weakly informative priors for the treatment effect parameter θ (Günhan et al. (2020)). The application we presented is aimed at showing how the Bayesian approach can be easily incorporated. Undoubtedly, we skipped some technicalities for the sake of clarity of explanation and space. Additional elaborations are beyond the primary subject matter of this paper.

5 Conclusion

In meta-analysis the studies are regarded as complex observations, which cannot be studies in isolation from each other (Cooper, 2017 p. 13). Accordingly, the prime task of meta-analysis is to integrate relatively similar but at the same time distinct studies in order to get the cumulative quantitative evidence on subject matter (Cooper, 2017 p. 23). In this paper we elucidated the theoretical and methodological aspects of the Bayesian approach to hierarchical modeling in meta-analysis.

In general, we attempted to contribute to the general meta-analytic scholarship by disentangling Bayesian hierarchical modeling and its current applications in meta-analysis. Additionally, we performed the re-analysis of existing meta-analytic data such as (1) conscientiousness and medication adherence, (2) the effectiveness of the Bacillus Calmette-Guerin (BCG) vaccine against tuberculosis (TB) using Bayesian hierarchical meta-analysis via well-established meta-analytic package in R and a brand new one.

In conclusion, Bayesian approach to hierarchical modeling in meta-analysis offers more flexible computational and methodological opportunities.

6 Appendix

6.1 Code

```r
library(metafor)
library(dplyr)
library(ggplot2)
library(brms)
library(metaviz)
df <- escalc(measure="ZCOR", ri=ri, ni=ni, data=dat.molloy2014)
data<- df %>%
  mutate(study = paste0(authors, " (", year, ")"), sei = sqrt(vi)) %>%
select(study, yi, sei) %>%
  slice(1:16)


par(mfrow=c(1,1))
ggplot(data, aes(x=yi, y=study)) +
  geom_segment(aes(x=yi-sei*2, xend = yi+sei*2,
                   y = study, yend= study)) +
  geom_point()



par(mfrow=c(1,2))
viz_forest(x = data[1:16, c("yi", "sei")],
           study_labels = data[1:16, c("study")],
           summary_label = "Summary effect",
           xlab = "Correleation Coefficient r", method="FE",
           annotate_CI = TRUE
)

viz_forest(x = data[1:16, c("yi", "sei")] ,
           study_labels = data[1:16, c("study")],
           summary_label = "Summary effect",
           xlab = "Correleation Coefficient r", method="ML",
           annotate_CI = TRUE
)


## Two level model (standard random effects model)

mafixed<- rma( yi=yi, sei=sei,  data= data, method="FE")
print(mafixed, digits=3)
ma<- rma( yi=yi, sei=sei, slab=study, data= data)
print(ma, digits=3)
forest(ma)
ml.ma <- rma.mv(yi, sei, random = ~1 | study, data = data)
print(ml.ma, digits =3)

#BHMA model
set.seed(1004)
prior_c <- c(set_prior("normal(0, 1)", class = "Intercept"),
             set_prior("cauchy(0, 0.2)", class = "sd"))
brm1<- brm(
  yi|se(sei) ~ 1 + (1|study),
  prior = prior_c,
  data = data,
  cores = 2,
  file = NULL)

summary(brm1)
```

```r


#BHMA forest graph
a<-posterior_summary(brm1)
hist(a[,1])
# Study-specific effects are deviations + average
out_r <- spread_draws(brm1, r_study[study,term], b_Intercept) %>%
  mutate(b_Intercept = r_study + b_Intercept)
# Average effect
out_f <- spread_draws(brm1, b_Intercept) %>%
  mutate(study = "Average")
# Combine average and study-specific effects' data frames
out_all <- bind_rows(out_r, out_f) %>%
  ungroup() %>%
  # Ensure that Average effect is on the bottom of the forest plot
  mutate(study = fct_relevel(study, "Average"))
# Data frame of summary numbers
out_all_sum <- group_by(out_all, study) %>%
  mean_qi(b_Intercept)
#> Warning: unnest() has a new interface. See ?unnest for details.
#> Try `cols = c(.lower, .upper)`, with `mutate()` needed
# Draw plot
out_all %>%
  ggplot(aes(b_Intercept, study)) +
  geom_density_ridges(
    rel_min_height = 0.01,
    col = NA,
    scale = 1
  ) +
  geom_pointintervalh(
    data = out_all_sum, size = 1
  ) +
  geom_text(
    data = mutate_if(out_all_sum, is.numeric, round, 2),
    # Use glue package to combine strings
    aes(label = glue::glue("{b_Intercept} [{.lower}, {.upper}]"), x = Inf),
    hjust = "inward"
  ) +
  geom_vline(xintercept = 0, linetype = "longdash") +
  theme_classic()

#probabilty check
brm1 %>%
  plot(
    combo = c("hist", "trace"), widths = c(1, 1.5),
    theme = theme_classic(base_size = 10))
# investigate model fit
mcmc_plot(brm1)


#application 2

#install.packages("MetaStan")
library("MetaStan")
# Loading required package: Rcpp
data("dat.Berkey1995", package = "MetaStan")
data<-(dat.Berkey1995)
library(ggplot2)
# Calculating log odds ratios and variances from data
logodds <- function(x) log((x[1] * (x[4] - x[3]))/((x[2] - x[1]) * x[3]))
stdes   <- function(x) sqrt(1/x[1] + 1/(x[2] - x[1]) + 1/x[3] + 1/(x[4] - x[3]))
r_ind   <- apply(cbind(dat.Berkey1995$rt, dat.Berkey1995$nt,
                       dat.Berkey1995$rc, dat.Berkey1995$nc), 1, logodds)
se_ind  <- apply(cbind(dat.Berkey1995$rt, dat.Berkey1995$nt,
```

```
                        dat.Berkey1995$rc, dat.Berkey1995$nc), 1, stdes)
lower95_ind <- r_ind + qnorm(.025) * se_ind
upper95_ind <- r_ind + qnorm(.975) * se_ind
# Comparison of the results
trials  <- c("1", "2" ,"3", "4", "5", "6", "7", "8", "9", "10", "11", "12", "13")
trials <- ordered(trials, levels = trials)
table(dat.Berkey1995)
d <- data.frame(x = trials,
                y = r_ind,
                sei = se_ind,
                ylo = lower95_ind,
                yhi = upper95_ind)
forest.plot <- ggplot(d, aes(x = x, y = y, ymin = ylo, ymax = yhi)) +
  geom_pointrange() +
  coord_flip() +
  geom_hline(aes(yintercept=0), lty = 2) +
  xlab("Studies") +
  ggtitle("Forest Plot (BCG vaccines)") +
  theme_classic()

plot(forest.plot)


library(rstan)
library(MetaStan)
library("shinystan")
library(shiny)

bnhm  <- meta_stan(ntrt = nt,
                           nctrl = nc,
                           rtrt = rt,
                           rctrl = rc,
                           data = dat.Berkey1995,
                           tau_prior_dist = "half-normal",
                           tau_prior = 0.5,
                           theta_prior = c(0, 2.82),
                           model = "BNHM1",
                           chains = 4,
                           iter = 2000,
                           warmup = 1000)
library("shiny")
bnhm.shinystan = as.shinystan(bnhm$fit)
launch_shinystan(bnhm.shinystan)
# print(bnhm.shinystan)
```

References

Albert, J., Spiegelhalter, D., & Best, N. (2005). *Bayesian analysis for epidemiologists iii: Regression modeling.* Retrieved from `http://www.columbia.edu/~cjd11/charles_dimaggio/DIRE/styled-4/styled-11/code-8/`

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis.* John Wiley & Sons.

Cooper, H. M. (2016). *Research synthesis and meta-analysis: a step-by-step approach* (Fifth Edition ed.). SAGE Publications.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, *20*(2), 115–147.

Enders, C. K. (2010). *Applied missing data analysis.* Guilford press.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2014). *Bayesian data analysis (3d ed.).* Boca Raton: Chapman and Hall/CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis.* CRC press.

Günhan, B. K., Röver, C., & Friede, T. (2020). Random-effects meta-analysis of few studies involving rare events. *Research Synthesis Methods*, *11*(1), 74–90.

Harrer, M., Cuijpers, P., & Furukawa, T. (2010). Doing meta-analysis in r: A hands-on guide. *Journal of Statistical Software*. Retrieved from `https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/`

Higgins, J. P., & Green, S. (2011, Mar). *Cochrane handbook for systematic reviews of interventions ...* Retrieved from `https://training.cochrane.org/handbook`

Jackson, D., Law, M., Stijnen, T., Viechtbauer, W., & White, I. R. (2018). A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Statistics in medicine*, *37*(7), 1059–1085.

Lesaffre, E., & Lawson, A. (2012). *Bayesian biostatistics.*

McElreath, R. (2015). *Statistical rethinking. a bayesian course with examples in r and stan.* Boca Raton: Chapman and Hall/CRC Press.

Mila, A., & Ngugi, H. (2011). A bayesian approach to meta-analysis of plant pathology studies. *Phytopathology*, *101*(1), 42–51.

Molloy, G., O'Carroll, R., & Ferguson, E. (2014). Conscientiousness and medication adherence: a meta-analysis. *Annals of Behavioral Medicine*, *47*(1), 92–101.

Pibouleau, L., & Chevret, S. (2013, 4). Bayesian hierarchical meta-analysis model for medical device evaluation: Application to intracranial stents. *International Journal of Technology Assessment in Health Care*, *29*. doi: 10.1017/s0266462313000093

Schmitz, S., Makovski, T., Adams, R., van den Akker, M., Stranges, S., & Zeegers, M. (2020). ayesian hierarchical models for meta-analysis of quality-of-life outcomes: An application in multimorbidity. *PharmacoEconomics*, *38*. doi: 10.1007/s40273-019-00843-z

Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H., & Thompson, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in medicine*, *19*(24), 3417–3432.

Van Den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and psychological measurement*, *63*(5), 765–790.

Vuorre, M. (2016, Sep). *Meta-analysis is a special case of bayesian multilevel modeling.* Retrieved from `https://mvuorre.github.io/post/2016/09/29/meta-analysis-is-a-special-case-of-bayesian-multilevel-modeling/`

Williams, D. R., Rast, P., & Bürkner, P.-C. (2018). Bayesian meta-analysis with weakly informative prior distributions.