

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Two Examples of Bayesian Evidence Synthesis with the Hierarchical Meta-Regression Approach

Pablo Emilio Verde

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.70231>

Abstract

This is the Information Age. We can expect that for a particular research question that is empirically testable, we should have a collection of evidence which indicates the best way to proceed. Unfortunately, this is not the case in several areas of empirical research and decision making. Instead, when researchers and policy makers ask a specific question, such as “What is the effectiveness of a new treatment?”, the structure of the evidence available to answer this question may be complex and fragmented (e.g. published experiments may have different grades of quality, observational data, subjective judgments, etc.).

Meta-analysis is a branch of statistical techniques that helps researchers to combine evidence from a multiplicity of indirect sources. A main hurdle in meta-analysis is that we not only combine results from a diversity of sources but we also combine their multiplicity of biases. Therefore, commonly applied meta-analysis methods, e.g. random-effects models, could be misleading.

In this chapter we present a new method for meta-analysis that we have called: the “Hierarchical Meta-Regression” (HMR). The HMR is an integrated approach for evidence synthesis when a multiplicity of bias, coming from indirect and disparate evidence, has to be incorporated in a meta-analysis.

Keywords: Bayesian hierarchical models, meta-analysis, multi-parameters evidence synthesis, conflict of evidence, randomized control trials, retrospective studies

1. Introduction

In today’s information age one can expect that the digital revolution can create a knowledge-based society surrounded by global communications that influence our world in an efficient and convenient way. It is recognized that never in human history we have accumulated such

an astronomical amount of data, and we keep on generating data at in an alarming rate. A new term, “big data,” was coined to indicate the existence of “oceans of data” where we may expect to extract useful information for any problem of interest.

In this technological society, one could expect that for a particular research question we should have a collection of high quality evidence which indicates the best way to proceed. Paradoxically, this is not the case in several areas of empirical research and decision making. Instead, when researchers and policy makers ask a specific and important question, such as “What is the effectiveness of a new treatment?”, the structure of the evidence available to answer this question may be complex and fragmented (e.g., published experiments may have different grades of quality, observational data, subjective judgments, etc.). The way how researchers interpret this multiplicity of evidence will be the basis for their understanding of reality and it will determine their future decisions.

Bayesian meta-analysis, which has its roots in the work of Eddy et al. [1], is a branch of statistical techniques for interpreting and displaying results of different sources of evidence, exploring the effects of biases and assessing the propagation of uncertainty into a coherent statistical model. A gentle introduction of this area can be found in Chap. 8 of Spiegelhalter et al. [2] and a recent review in Verde and Ohmann [3].

In this chapter we present a new method for meta-analysis that we have called: the “Hierarchical Meta-Regression” (HMR). The aim of HMR is to have an integrated approach for bias modeling when disparate pieces of evidence are combined in meta-analysis, for instance randomized and non-randomized studies or studies with different qualities. This is a different application of Bayesian inference than those applications with which we could be familiar, for instance an intricate regression model, where the available data bear directly upon the question of interest.

We are going to discuss two recent meta-analyses in clinical research. The reason for highlighting these two cases is that they illustrate a main problem in evidence synthesis, which is the presence of a multiplicity of bias in systematic reviews.

1.1. An example of meta-analysis of therapeutic trials

The first example, is a meta-analysis of 31 randomized controlled trials (RCTs) of two treatment groups of heart disease patients, where the treatment group received bone marrow stem cells and the control group a placebo treatment, Nowbar et al. [4]. The data of this meta-analysis appear in the Appendix, see **Table 1**. **Figure 1** presents the forest plot of these 31 trials, where the treatment effect is measured as the difference of the ejection fraction between groups, which measures the improvement of left ventricular function in the heart.

At the bottom of **Figure 1** we see average summaries represented by two diamonds: the first one corresponds to the fixed effect meta-analysis model. This model is based under the assumption that studies are identical and the between study variability is zero. The widest diamond represents the results of a random effects meta-analysis model, which assume a substantial heterogeneity between studies. In this meta-analysis both models confirmed a positive treatment of effect of a mean difference 3.95 95% CI [3.43; 4.47] and 2.92 and a 95% CI of [1.47, 4.36], respectively.

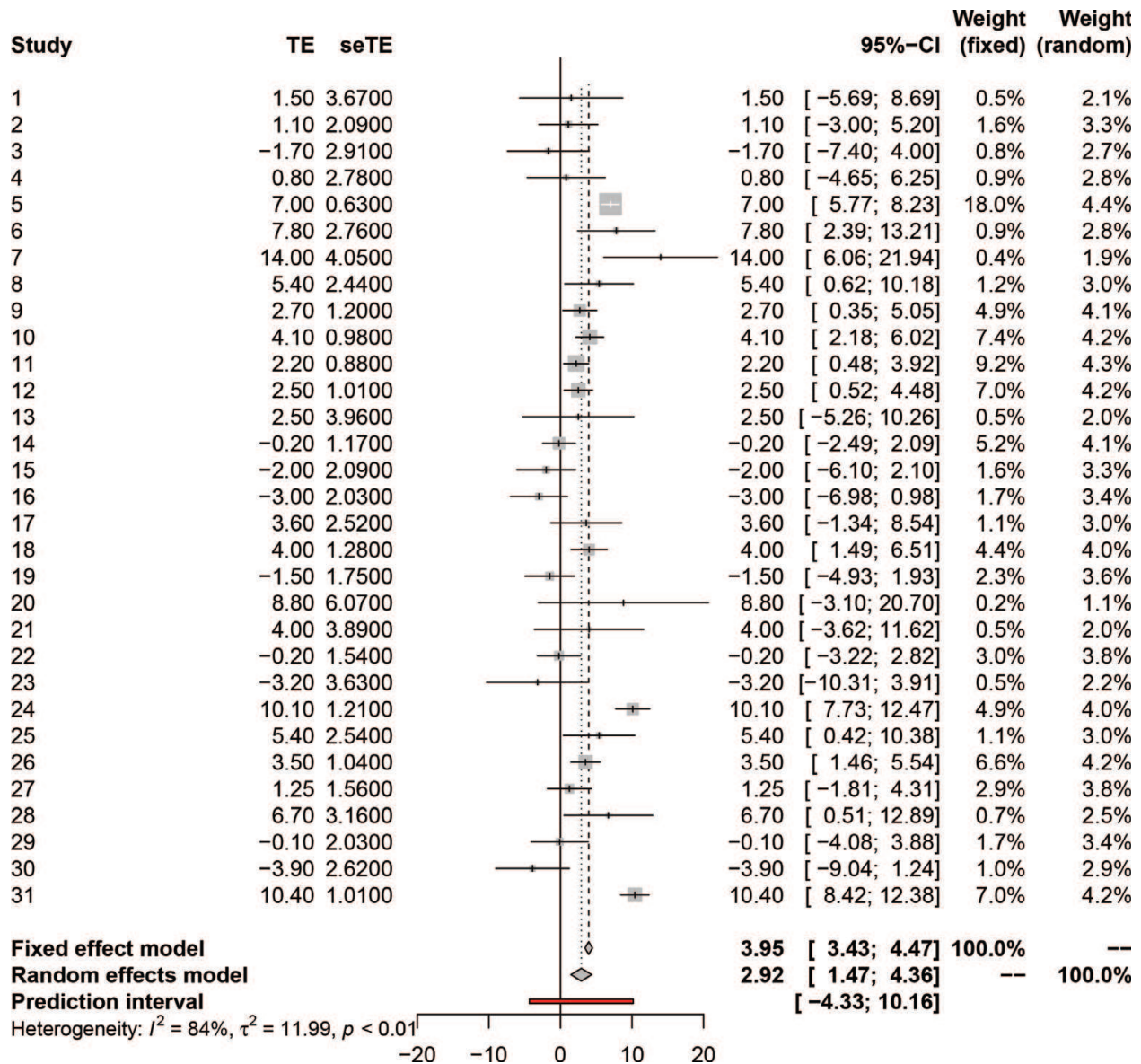


Figure 1. Meta-analysis results of studies applying treatments based on bone marrow stem cells to improve the left ventricular function.

Could we conclude that we have enough evidence to demonstrate the efficacy of the treatment? Unfortunately, these apparently confirming results are completely misleading. The problem is that these 31 studies are very heterogeneous, which resulted in a wide 95% prediction interval $[-4.33; 10.16]$ covering the no treatment effect, and a large number of contradictory evidence displayed in Figure 1.

In order to explain the sources of heterogeneity in this area Nowbar et al. [4] investigated whether detected discrepancies in published trials, might account for the variation in reported effect sizes. They define a discrepancy in a trial as two or more reported facts that cannot both be true because they are logically or mathematically incompatible. In other words, the term discrepancies is a polite way to indicate that a published study suffers from poor reporting, could be implausible or its results have been manipulated. For example, as we see at the

bottom of **Table 1** in the appendix, it would be difficult to believe in the results of a study with 55 discrepancies. In Section 2 we present a HMR model to analyze a possible link between the risk of bias results and the amount of discrepancies.

1.2. An example of meta-analysis of diagnostic trials

The topic of Section 3 is the meta-analysis of diagnostic trials. These trials play a central role in personalized medicine, policy making, healthcare and health economics. **Figure 2** presents our example in this area. The scatter plot shows the diagnostic summaries of a meta-analysis investigating the diagnostic accuracy of computer tomography scans in the diagnostic of appendicitis [5]. Each circle identifies the true positive rate vs. the false positive rate of each study, where the different circles' sizes indicate different sample sizes. One characteristic of this meta-analysis is the combination of disparate data. From 51 studies 22 were retrospective and 29 were prospective, which is indicated by the different grey scale of the circles.

The main problem in this area is the multiple sources of variability behind those diagnostic results. Diagnostic studies are usually performed under different diagnostic setups and patients' populations. For a particular diagnostic technique we may have a small number of studies which may differ in their statistical design, their quality, etc. Therefore, the main question in meta-analysis of diagnostic test is: How can we combine the multiplicity of diagnostic accuracy rates

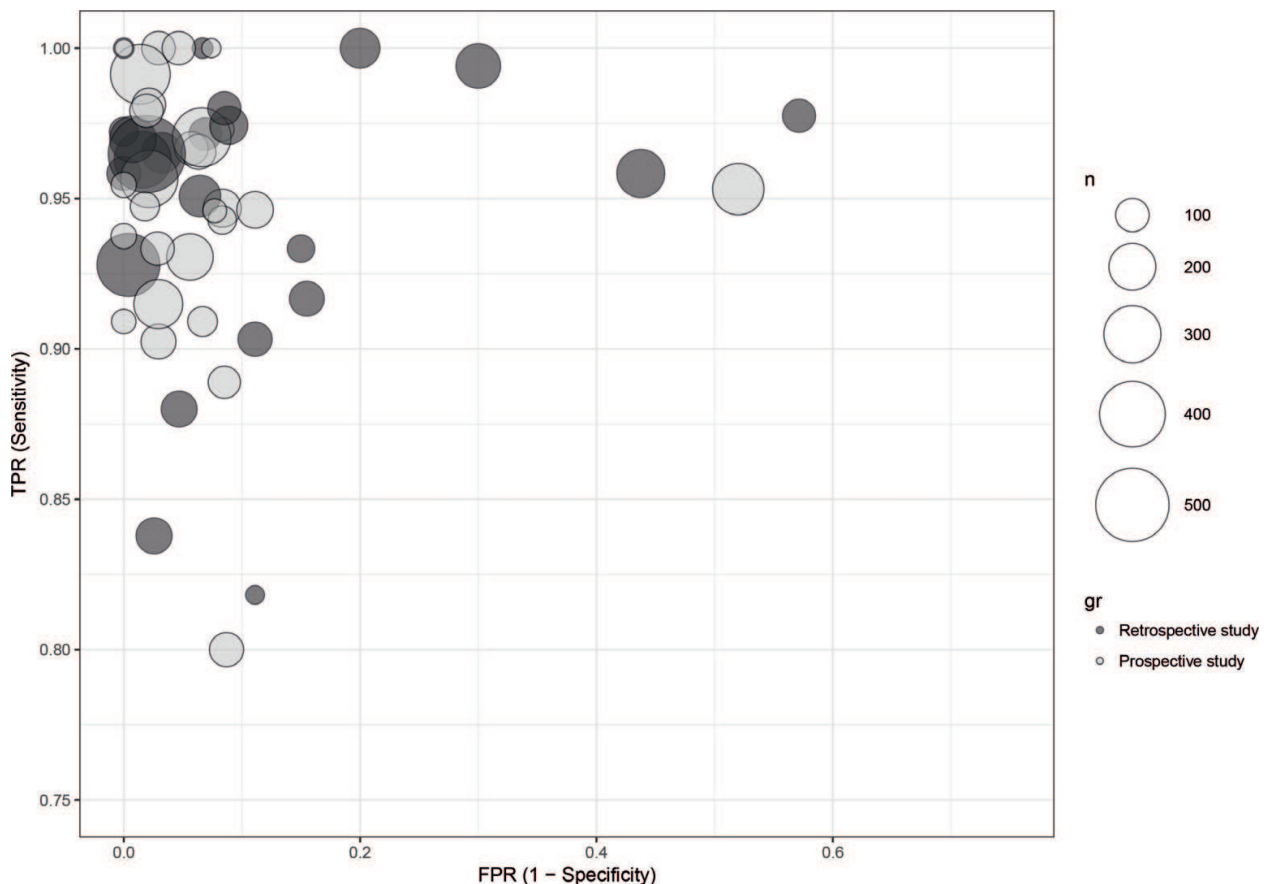


Figure 2. Display of the meta-analysis results of studies performing computer tomography scans in the diagnostic of appendicitis. Each circle identifies the true positive rate vs. the false positive rate of each study. Different colors are used for different study designs and different diameters for sample sizes.

in a single coherent model? A possible answer to this question is a HMR presented in Section 3. This model has been introduced by Verde [5] and it is available in the R's package **bamdit** [6].

2. A Hierarchical Meta-Regression model to assess reported bias

Figure 3 shows the reported effect size and the 95% confidence intervals of 31 trials from [4] against the number of discrepancies (in logarithmic scale). The authors reported a positive statistical significant correlation between the size effect and the number of discrepancies detected in the papers. However, a direct correlation analysis of aggregated results is threatened by ecological bias and it may lead to misleading conclusions. The amount of variability presented by the 95% confidence intervals is very big to accept a positive correlation at face value. In this section we are going to present a HMR model to link the risk of reporting bias with the amount of reported discrepancies. This model assumes that the connection between discrepancies and size effect could be much more subtle.

The starting point of any meta-analytic model is the description of a model for the pieces of evidence at face value. In statistical terms, this means the likelihood of the parameter of interest. Let y_1, \dots, y_N and SE_1, \dots, SE_N be the reported effect sizes and their corresponding standard errors, we assume a normal likelihood of θ_i the treatment effect of study i :

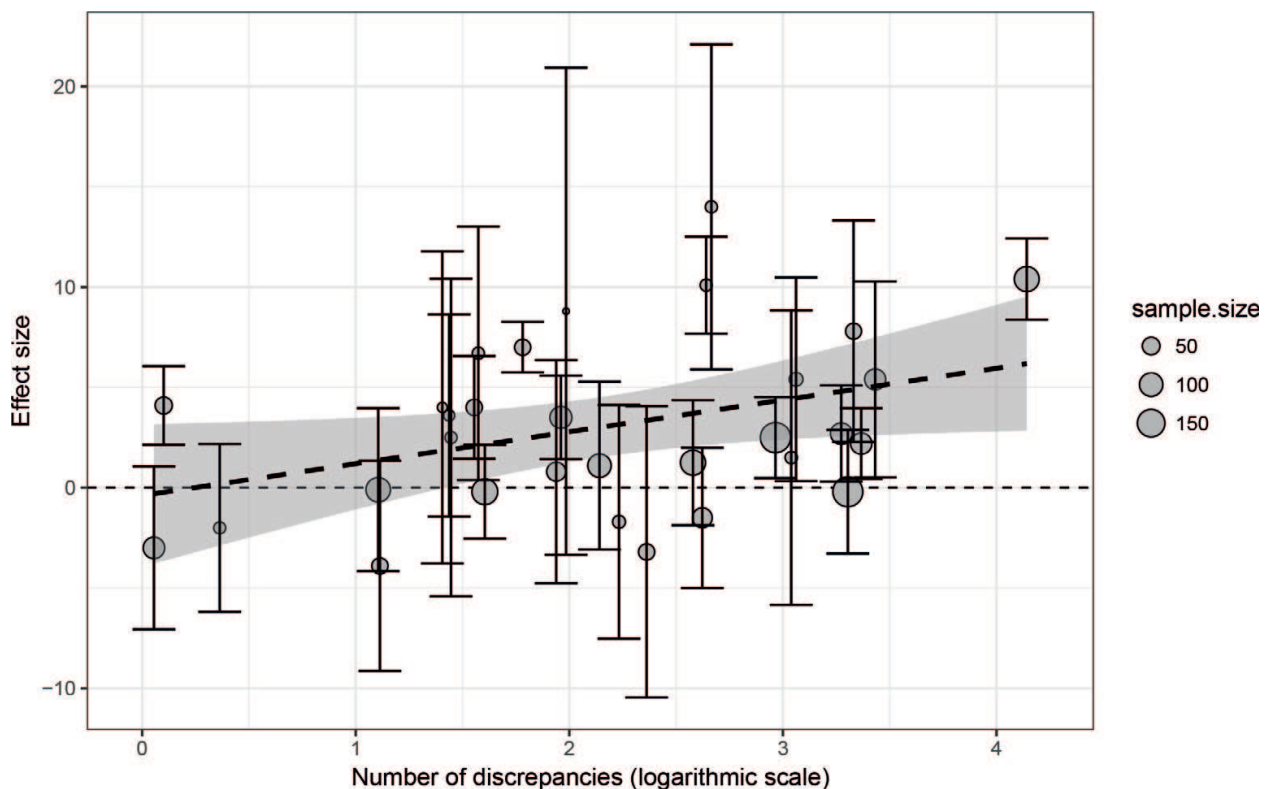


Figure 3. Relationship between effect size and number of discrepancies. The vertical axis corresponds to the effect size, the treatment group received a treatment based on bone marrow stem cells and the control group a placebo treatment. The horizontal axis corresponds to the number of discrepancies (in the logarithmic scale) found in the publication.

$$y_i | \theta_i \sim N(\theta_i, SE_i^2), \quad i = 1, \dots, N. \quad (1)$$

If a prior assumption of exchangeability was considered reasonable, a random effects Bayesian model incorporates all the studies into a single model, where the $\theta_1, \dots, \theta_N$ are assumed to be a random sample from a prior distribution with unknown parameters, which is known as a hierarchical model.

In this section we assume that exchangeability is unrealistic and we wish to learn how the unobserved treatment effects $\theta_1, \dots, \theta_N$ are linked with some observed covariate x_i .

Let x_i be the number of observed discrepancies in the logarithmic scale. We propose to model the association between the treatment effect θ_i and the observed discrepancies x_i with the following HMR model:

$$\theta_i | I_i, x_i \sim I_i N(\mu_{\text{biased}}, \tau^2) + (1 - I_i) N(\mu, \tau^2), \quad (2)$$

where the non-observable variable I_i indicates if study i is at risk of bias:

$$I_i | x_i = \begin{cases} 1 & \text{if study } i \text{ is biased} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The parameter μ corresponds to the mean treatment effect of studies with low risk of bias. We assume that in our context of application biased studies could report higher effect sizes and the biased mean μ_{biased} can be expressed as:

$$\mu_{\text{biased}} = \mu + K, \text{ with } K > 0. \quad (4)$$

In this way, K measures the average amount of bias with respect to the mean effect μ . Eq. (4) also ensures that μ and μ_{biased} are identifiable parameters in this model. The parameter τ measures the between-studies variability in both components of the mixture distributions.

We model the probability that a study is biased as a function of x_i as follows:

$$\text{logit}(\Pr(I_i = 1 | x_i)) = \alpha_0 + \alpha_1 x_i. \quad (5)$$

In Eq. (5) positive values of α_1 indicate that an increase in the number of discrepancies is associated with an increased risk of study bias.

In this HMR model the conditional mean is given by

$$E(\theta | x_i) = \Pr(I_i = 1 | x_i) \mu_{\text{biased}} + (1 - \Pr(I_i = 1 | x_i)) \mu. \quad (6)$$

Eqs. (5) and (6) can be calculated as functional parameters for a grid of values of x . Their posteriors intervals are calculated at each value of x .

This HMR not only quantifies the average bias K and the relationship between bias and discrepancies in Eq. (5), but also allows to correct the treatment effect θ_i by its propensity of being biased:

$$\theta_i^{\text{corrected}} = (\theta_i - K)\Pr(I_i = 1|x_i) + \theta_i(1 - \Pr(I_i = 1|x_i)), \quad (7)$$

where the amount $(\theta_i - K)$ measures the bias of study i and $\Pr(I_i = 1|x_i)$ its propensity of being biased.

The HMR model presented above is completed by the following vague hyper-priors: For the regression parameters $\alpha_0, \alpha_1 \sim N(0, 100)$. We give to the mean $\mu_1 \sim N(0, 100)$ and for the bias parameter $K \sim \text{Uniform}(0, 50)$. Finally, for the variability between studies we use $\tau \sim \text{Uniform}(0, 100)$, which represent a vague prior within the range of possible study deviations.

The model presented in this section is mathematically non-tractable. We approximated the posterior distributions of the model parameters with Markov Chain Monte Carlo (MCMC) techniques implemented in *OpenBUGS*.

BUGS stands for *Bayesian Analysis Using Gibbs Sampling*, the *OpenBUGS* software constructs a Directed Acyclic Graph (DAG) representation of the posterior distribution of all model's parameters. This representation allows to automatically factorize the DAG as a product of each node (parameters or data) conditionally on its parents and children. The software scans each node and proposes a method of sampling. The kernel of the Gibbs sampling is built upon this algorithm.

Computations were performed with the statistical language *R* and MCMC computations were linked to *R* with the package **R2OpenBUGS**. We used two chains of 20,000 iterations and we discarded the first 5000 for the burn-in period. Convergence was assessed visually by using the *R* package **coda**.

The diagonal panels of **Figure 4** summarize the resulting posterior distributions for μ, K, τ, α_0 and α_1 . The posterior of μ clearly covers the zero indicating that the stem cells treatment is not effective. The bias parameter K indicates a considerable over-estimation of treatment effects reported for some trials. The posterior of α_1 is concentrated in positive values, which indicates that an increase in discrepancies is associated with an increase of the risk of reporting bias. The posteriors of α_0 and α_1 also present a large variability, which is expected when a hidden effect is modeled.

Further results of the Hierarchical Meta-Regression model appears in **Figure 5**, where posteriors 95% intervals are plotted against the number of discrepancies. On the left panel, we can see the relationship between the number of discrepancies and the probability that a study is biased. We can observe an increase of probability with an increase of the number of discrepancies, but also a large amount of variability. On the right panel appears the conditional mean of effect size as a function of the number of discrepancies, which corresponds to Eq. (6). Our analysis shows that the 95% posterior intervals of the conditional mean covers the zero effect in most of the range of discrepancies. Only for studies with more than 33 ($\exp(3.5)$) discrepancies

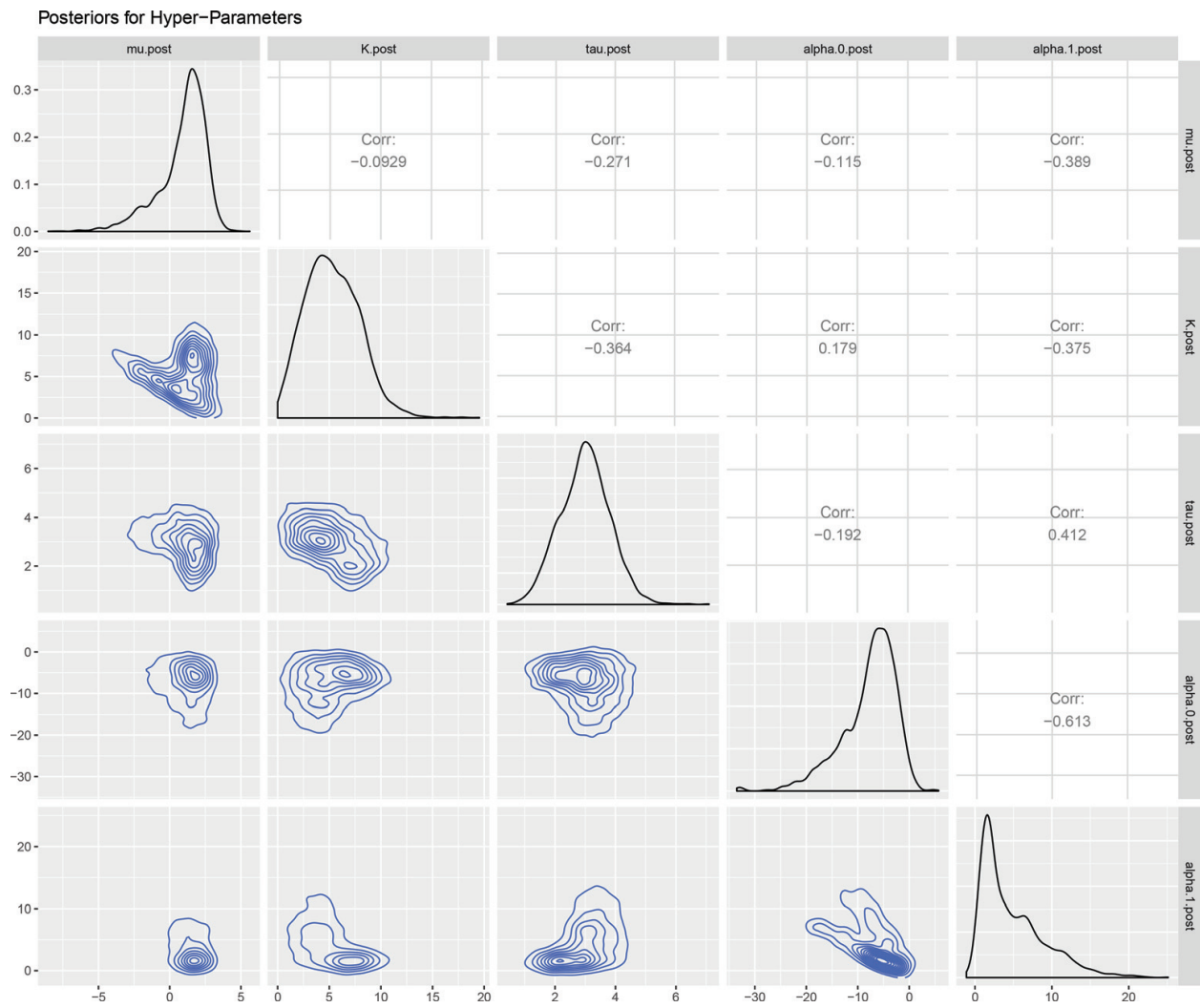


Figure 4. Posterior distributions for the hyper-parameters of the HRM model. The diagonal displays the posterior distributions, the upper panels the pairwise correlations and the lower panels the pairwise posterior densities.

the model predicts a positive effect. One interesting result of this analysis is, that a horizontal line which may represent a zero correlation is also predicted by the model. This means that the regression calculated directly from the aggregated data contains an ecological bias and it is misleading. We have added this regression line to the plot to highlight this issue.

The results presented so far indicate that increases in the amount of discrepancies increases the propensity of bias. The question is: How can we correct a particular study for its bias? Eq. (7) gives the bias correction of treatment effect in this HMR model.

In **Figure 6** we can see HMR bias correction in action. We display two studies which have 21 and 18 discrepancies respectively. The solid lines correspond to the likelihood functions of these studies. These likelihoods represent the information of the effect size at face value. The dashed lines correspond to the posterior treatment effects after bias correction. Clearly, we can see a strong bias correction with the conclusion of no treatment effect.

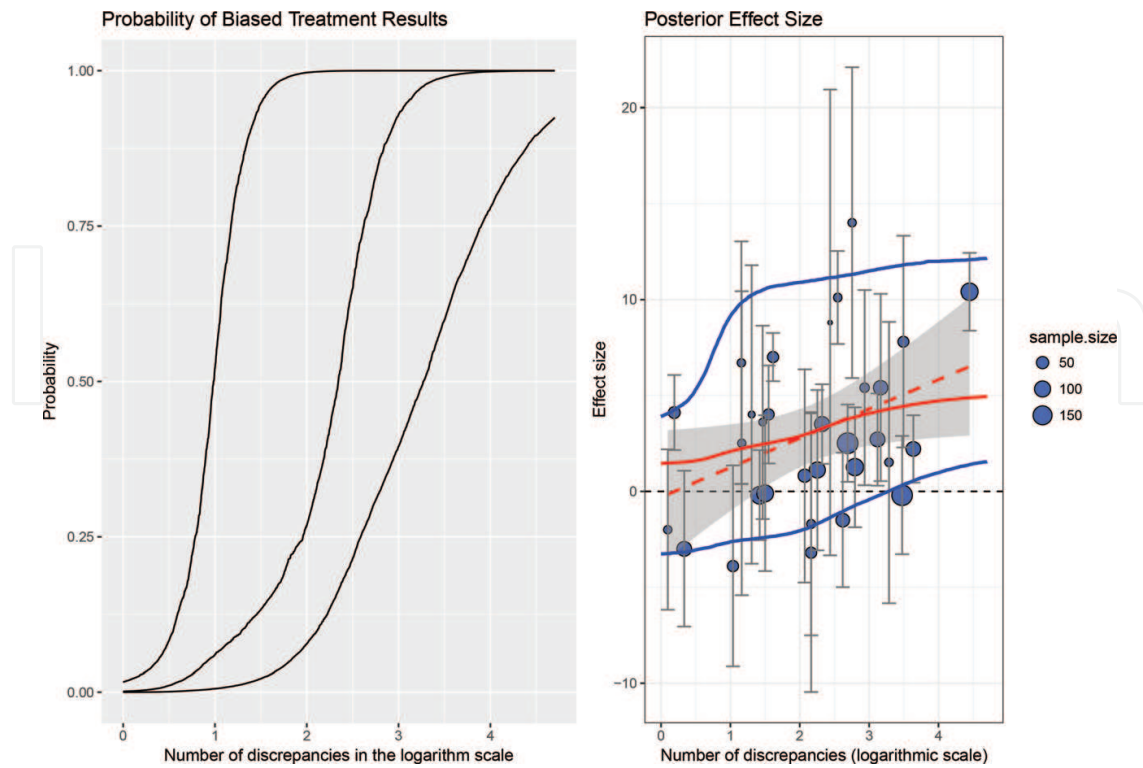


Figure 5. Results of the Hierarchical Meta-Regression model. The posterior median and 95% intervals are displayed as solid lines. Left panel: relationship between the number of discrepancies and probability that a study is biased. Right panel: conditional mean of effect size as a function of the number of discrepancies.

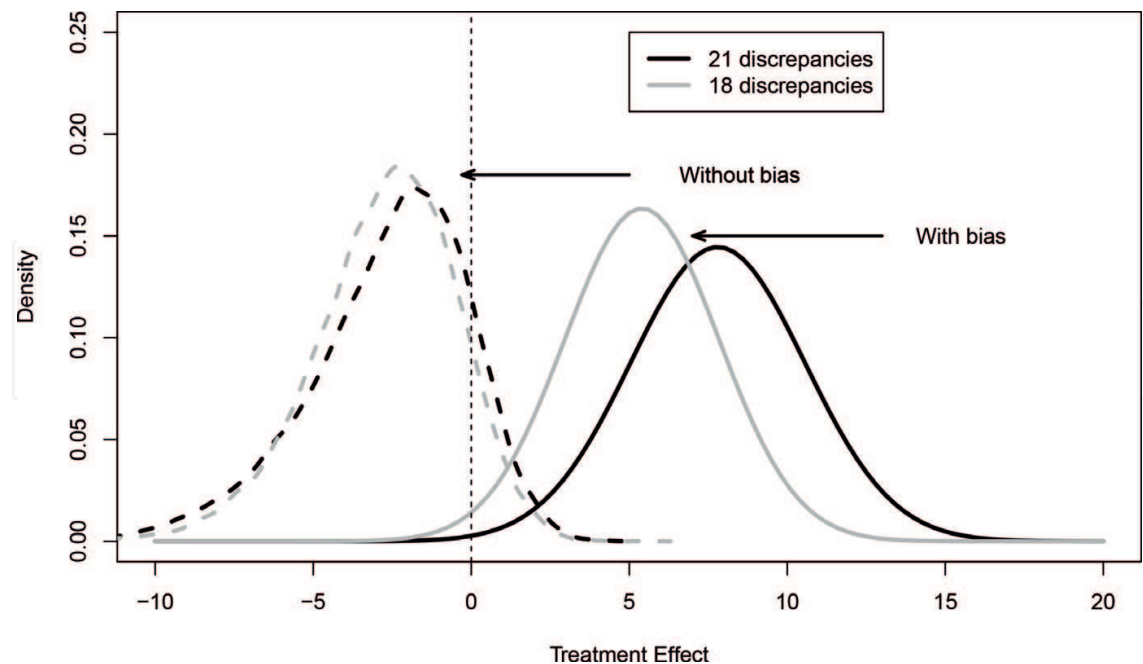


Figure 6. Bias correction for two studies with 21 and 18 discrepancies respectively. The solid lines correspond to the likelihood functions of effect sizes. The dashed lines represent the posteriors for treatment effect after bias correction.

3. Hierarchical Meta-Regression analysis for diagnostic test data

In meta-analysis of diagnostic test data, the pieces of evidence that we aim to combine are the results of N diagnostic studies, where results of the i th study ($i = 1, \dots, N$) are summarized in a 2×2 table as follows:

| | | Patient status | |
|---------|---|----------------|-----------------|
| | | With disease | Without disease |
| Test | + | tp_i | fp_i |
| Outcome | - | fn_i | tn_i |
| Sum: | | $n_{i,1}$ | $n_{i,2}$ |

where tp_i and fn_i are the number of patients with positive and negative diagnostic results from $n_{i,1}$ patients with disease, and fp_i and tn_i are the positive and negative diagnostic results from $n_{i,2}$ patients without disease.

Assuming that $n_{i,1}$ and $n_{i,2}$ have been fixed by design, we model the tp_i and fp_i outcomes with two independent Binomial distributions:

$$tp_i \sim B(\text{TPR}_i, n_{i,1}) \quad \text{and} \quad fp_i \sim B(\text{FPR}_i, n_{i,2}), \quad (8)$$

where TPR_i is the true positive rate or sensitivity, Se_i , of study i and FPR_i is the false positive rate or complementary specificity, i.e., $1 - \text{Sp}_i$.

At face value, diagnostic performance of each study is summarized by the empirical true positive rate and true negative rate or specificity

$$\widehat{\text{TPR}}_i = \frac{tp_i}{n_{i,1}} \quad \text{and} \quad \widehat{\text{TNR}}_i = \frac{tn_i}{n_{i,2}} \quad (9)$$

and the complementary empirical rates of false positive rate and false negative diagnostic results,

$$\widehat{\text{FPR}}_i = \frac{fp_i}{n_{i,2}} \quad \text{and} \quad \widehat{\text{FNR}}_i = \frac{fn_i}{n_{i,1}}. \quad (10)$$

In this type of meta-analysis we could separately model TPR_i and FPR_i (or Sp_i), but this approach ignores that these rates could be correlated by design. Therefore, it is more sensible to handle TPR_i and FPR_i jointly.

We define the random effect D_i which represents the study effect associated with the diagnostic discriminatory power:

$$D_i = \log\left(\frac{\text{TPR}_i}{1 - \text{TPR}_i}\right) - \log\left(\frac{\text{FPR}_i}{1 - \text{FPR}_i}\right). \quad (11)$$

However, diagnostic results are sensitive to diagnostic settings (e.g., the use of different thresholds) and to populations where the diagnostic procedure under investigation is applied. These issues are associated with the *external validity* of diagnostic results. To model external validity bias we introduce the random effect S_i :

$$S_i = \log\left(\frac{\text{TPR}_i}{1 - \text{TPR}_i}\right) + \log\left(\frac{\text{FPR}_i}{1 - \text{FPR}_i}\right). \quad (12)$$

This random effect quantifies variability produced by patients' characteristics and diagnostic setup, that may produce a correlation between the observed $\widehat{\text{TPR}}$ s and $\widehat{\text{FPR}}$ s. In short, we called S_i **the threshold effect** of study i and it represents an adjustment of external validity in the meta-analysis.

We could assume exchangeability of pairs (D_i, S_i) , but study's quality is known to be an issue in diagnostic studies. For this reason we model the *internal validity* of a study by introducing random weights w_1, \dots, w_N . Conditionally to a study weight w_i , the study effects D_i and S_i are modeled as exchangeable between studies and they follow a *scale-mixture of bivariate Normal* distributions with the following mean and variance:

$$E\left[\begin{pmatrix} D_i \\ S_i \end{pmatrix} \middle| w_i\right] = \begin{pmatrix} \mu_D \\ \mu_S \end{pmatrix} \quad \text{and} \quad \text{var}\left[\begin{pmatrix} D_i \\ S_i \end{pmatrix} \middle| w_i\right] = \frac{1}{w_i} \begin{pmatrix} \sigma_D^2 & \rho\sigma_D\sigma_S \\ \rho\sigma_D\sigma_S & \sigma_S^2 \end{pmatrix} = \Sigma_i, \quad (13)$$

and scale mixing density

$$w_i \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \quad (14)$$

The inclusion of the random weights w_i into the model was proposed by [5]. This approach was generalized in [6] in two ways: firstly, by splitting w_i in two weights $w_{1,i}$ and $w_{2,i}$ corresponding to each component D_i and S_i respectively. Secondly, by putting a prior on the degrees of freedom parameter ν , which corresponds to an adaptive robust distribution of the random-effects.

The Hierarchical Meta-Regression representation of the model introduced above is the model based on the conditional distribution of $(D_i | S_i = x)$ and the marginal distribution of S_i . This HMR model was introduced by [7], who followed the stepping stones of the classical Summary Receiving Operating Characteristic (SROC) [8].

The conditional mean of $(D_i | S_i = x)$ is given by:

$$E(D_i | S_i = x) = A + Bx \quad (15)$$

where the functional parameters A and B are

$$A = \mu_D, \quad \text{and} \quad B = \rho \frac{\sigma_D}{\sigma_S}. \quad (16)$$

We define the *Bayesian SROC Curve* (BSROC) by transforming back results from (S, D) to (FPR, TPR) with

$$\text{BSROC}(\text{FPR}) = g^{-1} \left[\frac{A}{(1-B)} + \frac{B+1}{(1-B)} g(\text{FPR}) \right], \quad (17)$$

where $g(p)$ is the logit(p) transformation, i.e. $\text{logit}(p) = \log(p/(1-p))$.

The BSROC curve is obtained by calculating TPR in a grid of values of FPR which gives a posterior conditionally on each value of FPR. Therefore, it is straightforward to give credibility intervals for the BSROC for each value of FPR.

One important aspect of the BSROC is that it incorporates the variability of the model's parameters, which influences the width of its credibility intervals. In addition, given that FPR is modeled as a random variable, the curve is corrected by measurement error bias in FPR.

Finally, we can define a *Bayesian Area Under the SROC Curve* (BAUC) by numerically integrating the BSROC for a range of values of the FPR:

$$\text{BAUC} = \int_0^1 \text{BSROC}(x) dx. \quad (18)$$

In some applications it is recommend to use the limits of integration within the observed values of $\widehat{\text{FPR}}$ s.

In order to make this complex HMR model applicable in practice, we have implemented the model in the R's package **bamdit**, which uses the following set of hyper-priors:

$$\mu_D \sim \text{Logistic}(m_1, v_1), \quad \mu_S \sim \text{Logistic}(m_2, v_2) \quad (19)$$

and

$$\sigma_D \sim \text{Uniform}(0, u_1), \quad \sigma_S \sim \text{Uniform}(0, u_2). \quad (20)$$

The correlation parameter ρ is transformed by using the Fisher transformation,

$$z = \text{logit} \left(\frac{\rho + 1}{2} \right) \quad (21)$$

and a Normal prior is used for z :

$$z \sim N(m_r, v_r). \quad (22)$$

Modeling priors in this way guarantees that in each MCMC iteration the variance-covariance matrix of the random effects θ_1 and θ_2 is positive definite. The values of the constants $m_1, v_1, m_2, v_2, u_1, u_2, m_r$ and v_r have to be given. They can be used to include valid prior information which might be empirically available or they could be the result of expert elicitation. If such information is not available, we recommend setting these parameters to values

that represent weakly informative priors. In this work, we use $m_1 = m_2 = m_r = 0$, $v_1 = v_2 = 1$, $u_1 = u_2 = 5$ and $v_r = \sqrt{1.7}$ as weakly informative prior setup.

These values are fairly conservative, in the sense that they induce prior uniform distributions for TPR_i and FPR_i . They give locally uniform distributions for μ_1 and μ_2 ; uniforms for σ_1 and σ_2 ; and a symmetric distribution for ρ centered at 0.

Figure 7 summarizes the meta-analysis results of fitting the bivariate random-effect model to the computer tomography diagnostic data. The Bayesian Predictive Surface are presented by contours at different credibility levels and compare these curves with the observed data represented by the circles with varying diameters according to the sample size of each study. The scattered points are samples from the predictive posteriors and the histograms correspond to the posterior predictive marginals. This result was generated by using the functions `metadiag()` and `plot` in the R package **bamdit**.

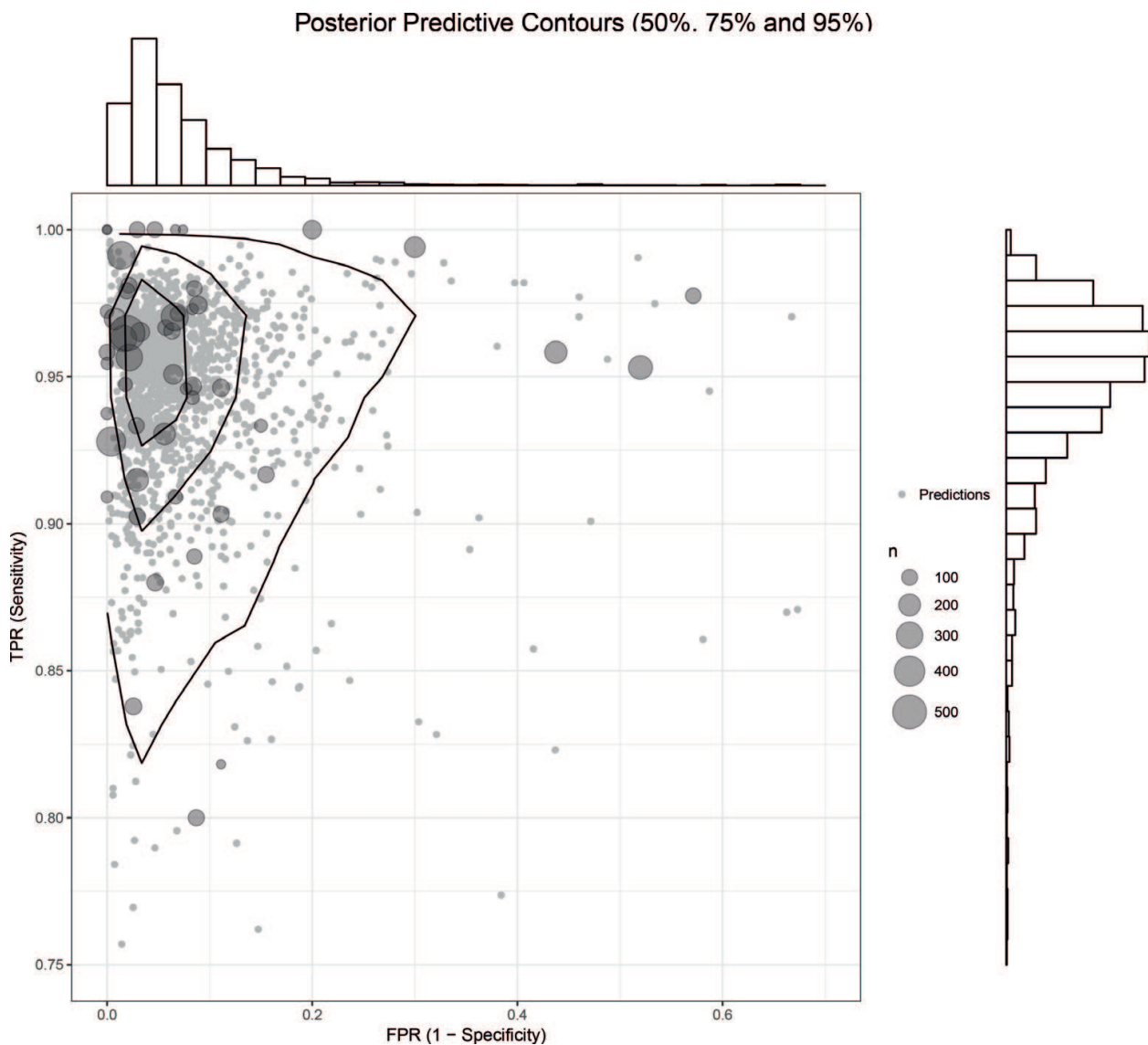


Figure 7. Results of the meta-analysis: Bayesian Predictive Surface by contours at different credibility levels.

Figure 8 displays the posteriors of each components' weights. The left panel shows that prospective studies number 25 and 33 deviate with respect to the prior mean of 1, while on the right panel we see that a prospective study (number 47) and five retrospective studies (number 1, 3, 4, 8 and 29) have substantial variability.

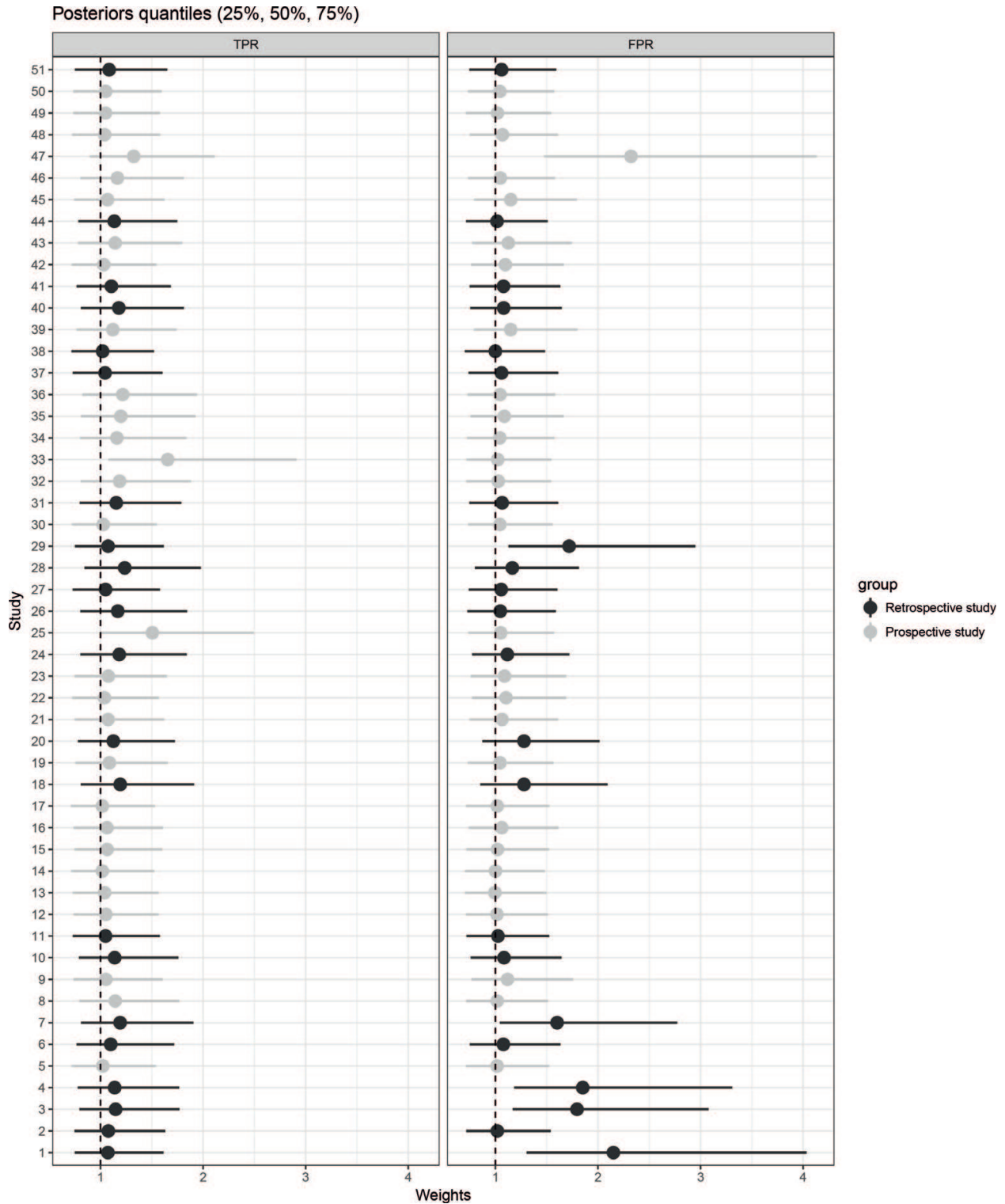


Figure 8. Posterior distributions of the component weights: it is expected that the posterior is centered at 1. Studies with retrospective design tend to present deviations in FPR.

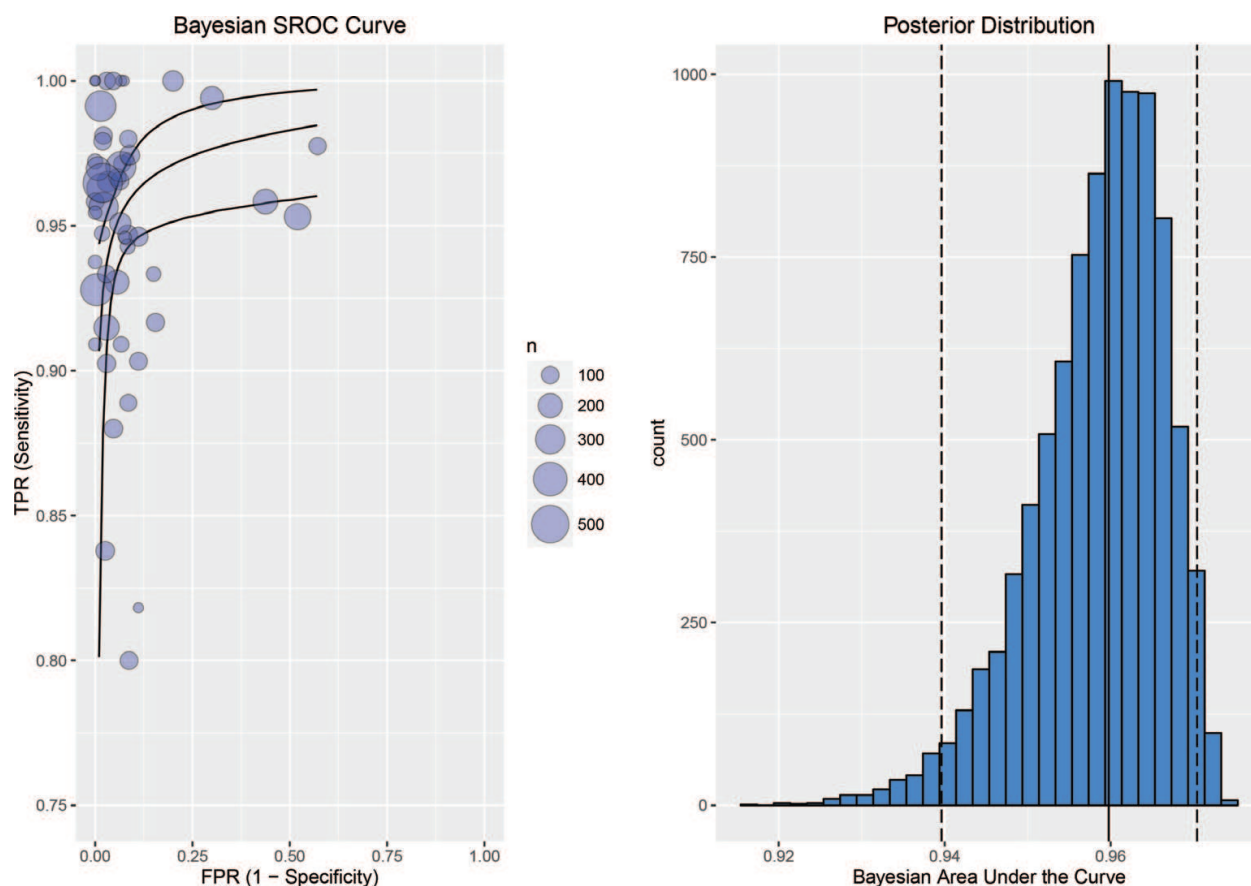


Figure 9. Hierarchical Meta-Regression model: left panel shows the BSROC curve, the central line corresponds to the posterior median and the upper and lower curves correspond to the quantiles of the 2.5 and 97.5%, respectively. The right panel displays the posterior distribution of the area under the BSROC curve.

An important aspect of w_i is its interpretation as **estimated bias correction**. *A priori* all studies included in the review have a mean of $E(w_i) = 1$. We can expect that studies which are unusually heterogeneous will have posteriors substantially greater than 1. Unusual studies' results could be produced by factors that may affect the quality of the study, such as errors in recording diagnostic results, confounding factors, loss to follow-up, etc. For that reason, the studies' weights w_i can be interpreted as an adjustment of studies' **internal validity bias**.

The BSROC curve and its area under the curve are presented in **Figure 9**. The left panel shows this HMR as a meta-analytic summary for this data. On the right panel the posterior distribution of the BAUC show quite a high diagnostic ability for computer tomography scans as diagnostic of appendicitis.

4. Conclusions

In this work we have seen the HMR in action. This approach of meta-analysis is based on a simple strategy: two sub-models are defined in the meta-analysis, one which models the problem of interest, for instance the treatment effect, and one which handles the multiplicity

of bias. The meta-analysis is summarized by understanding how these components interact with each other.

The examples presented in this work have shown that we could have misleading conclusions from indirect evidence, if it were analyzed as directly contributing to the problem of interest.

For instance, in the first example, Section 2, we have seen in **Figure 1** that pooling studies gave a wrong conclusion about the effect of stem cells treatment. The positive correlation between the aggregated effect size and the number of discrepancies exaggerates its relationship.

Actually, in **Figure 5** the HMR has shown that it is possible to simultaneously have a zero correlation between effect size and discrepancies while still having a risk of reporting bias. In addition, the HMR allows to extract the amount of bias in the meta-analysis and to correct the treatment effect at the level of the study (**Figure 6**).

In the second example, Section 3, biases come from the external validity of diagnostic studies and the internal validity due to their quality. In this example the HMR showed that it was possible to simultaneously model these two types of subtle biases.

To account for internal validity bias, the application of a scale mixture of normal distributions allows us to detect conflictive studies, which can be considered as outliers. The Bayesian Summary Receiving Operative Curve accounts for the external validity bias due to changes in factors that affected the diagnostic results. In addition, the posterior for its Area Under the Curve (AUC) summarizes the results of the meta-analysis.

Acknowledgements

This work was supported by the German Research Foundation project DFG VE 896 1/1.

Appendix: Source Data for Sections 1.1 and Section 2

| Trial ID | Effect size | SE (effect size) | Sample size | Number of discrepancies | Author or principal investigator | Year | Country |
|----------|-------------|------------------|-------------|-------------------------|----------------------------------|------|----------------|
| t01 | 1.5 | 3.67 | 21 | 17 | Quyumi | 2011 | USA |
| t02 | 1.1 | 2.09 | 100 | 7 | Lunde | 2007 | Norway |
| t03 | -1.7 | 2.91 | 23 | 7 | Srimahachota | 2011 | Thailand |
| t05 | 0.8 | 2.78 | 60 | 4 | Meyer | 2006 | Germany |
| t06 | 7 | 0.63 | 40 | 4 | Meluzín | 2006 | Czech Republic |
| t09 | 7.8 | 2.76 | 38 | 21 | Piepoli | 2010 | Italy |
| t11 | 14 | 4.05 | 20 | 13 | Suárez de Lezo | 2007 | Spain |
| t12 | 5.4 | 2.44 | 77 | 18 | Huikuri HV | 2008 | Finland |
| t13 | 2.7 | 1.2 | 82 | 16 | Perin | 2012 | USA |
| t15 | 4.1 | 0.98 | 46 | 0 | Assmus | 2006 | Germany |

| Trial ID | Effect size | SE (effect size) | Sample size | Number of discrepancies | Author or principal investigator | Year | Country |
|----------|-------------|------------------|-------------|-------------------------|----------------------------------|------|-----------------|
| t16 | 2.2 | 0.88 | 79 | 27 | Assmus | 2013 | Germany |
| t17 | 2.5 | 1.01 | 187 | 11 | Assmus | 2010 | Germany |
| t18 | 2.5 | 3.96 | 20 | 2 | Hendrikx | 2006 | Belgium |
| t19 | -0.2 | 1.17 | 127 | 3 | Hirsch | 2011 | The Netherlands |
| t20 | -2 | 2.09 | 20 | 0 | Perin | 2012 | USA |
| t23 | -3 | 2.03 | 81 | 0 | Traverse | 2011 | USA |
| t24 | -3.6 | 2.52 | 17 | 2 | Ang | 2008 | UK |
| t25 | 4 | 1.28 | 40 | 3 | Rodrigo | 2012 | The Netherlands |
| t26 | -1.5 | 1.75 | 66 | 8 | Herbolts | 2009 | Belgium |
| t29 | 8.8 | 6.07 | 10 | 6 | Castellani | 2010 | Italy |
| t30 | 4 | 3.89 | 14 | 2 | Maureira | 2012 | France |
| t31 | -0.2 | 1.54 | 183 | 19 | Ribero dos Santos | 2012 | Brazil |
| t32 | -3.2 | 3.63 | 40 | 7 | Traverse | 2010 | USA |
| t35 | 10.1 | 1.21 | 20 | 11 | Patel | 2004 | USA |
| t38 | 5.4 | 2.54 | 27 | 15 | Tse | 2007 | Hong Kong |
| t42 | 3.5 | 1.04 | 86 | 6 | Cao | 2009 | China |
| t45 | 1.25 | 1.56 | 118 | 9 | Sürder | 2013 | Switzerland |
| t46 | 6.7 | 3.16 | 20 | 2 | Ge | 2006 | China |
| t47 | -0.1 | 2.03 | 112 | 2 | Traverse | 2012 | USA |
| t48 | -3.9 | 2.62 | 40 | 1 | Wöhrle | 2010 | Germany |
| t49 | 10.4 | 1.01 | 116 | 55 | Yousef (Strauer) | 2009 | Germany |

Table 1. Results from 31 randomized controlled trials of heart disease patients, where the treatment group received bone marrow stem cells and the control group a placebo treatment. The source of this table is Nowbar et al. [4].

Author details

Pablo Emilio Verde

Address all correspondence to: pabloemilio.verde@hhu.de

Coordination Center for Clinical Trials, University of Duesseldorf, Duesseldorf, Germany

References

- [1] Eddy DM, Hasselblad V, Shachter R. Meta-Analysis by the Confidence Profile Method: The Statistical Synthesis of Evidence. San Diego, CA: Academic Press; 1992

- [2] Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons, Ltd.; 2004
- [3] Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: A review of methods and applications. *Research Synthesis Methods*. Vol. 6. 2014. DOI: 10.1002/jrsm.1122
- [4] Nowbar AN, Mielewczik M, Karavassilis M, Dehbi HM, Shun-Shin MJ, Jones S, Howard JP, Cole GD, Francis DP. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (damascene): Weighted regression and meta-analysis. *BMJ*. 2014;**348**:1-9
- [5] Verde PE. Meta-analysis of diagnostic test data: A bivariate Bayesian modeling approach. *Statistics in Medicine*. 2010;**29**(30):3088-3102
- [6] Verde PE. bamdit: An R package for Bayesian meta-analysis of diagnostic test data. *Journal of Statistical Software*. 2017, in press
- [7] Verde PE. Meta-analysis of diagnostic test data: Modern statistical approaches. PhD Thesis, University of Düsseldorf. Deutsche Nationalbibliothek. July, 2008. Available from: <http://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=8494>
- [8] Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine*. 1993;**12**:1293-1316