Bayesian Hierarchical Meta-Regression Using Study-Informed Priors

Jihyung Song
Jeff Gill
Graham Colditz
Division of Health Sciences, Department of Surgery
School of Medicine
Washington University, St. Louis

ABSTRACT

In this work provide a new approach to Bayesian meta-regression modeling in a hierarchical form with an enhanced ability to add prior information for model choice and coefficients. Our model includes more intra-study information than conventional meta-analysis and current forms of meta-regression. A key feature is simultaneous posterior estimates of standard meta-analysis effect summaries, such as log-relative risks, and model-averaged covariate coefficient estimates. We apply this new technology to....

1 Overview

Widespread policy and prescriptive procedures should be based on a body of studies rather than one single experiment or trial. This is not controversial. However, the manner in which many meta-analyses are performed invites severe criticism in the general statistics literature ( ) and elsewhere (Egger and Smith 1995)). Bluntly put, the heterogeneity of studies included in a meta-analysis is typically understated, and to the detriment of subsequent conclusions. Approaches that recognize unwanted, but present, heterogeneity are more robust describers of the studies effects. The most flexible and principled approach is the fully Bayesian approach of assuming that all unknown quantities should be described with probability statements, and that these probability statements can be parameterized.

In this work we revisit distributional assumptions underlying meta-analysis as normally practiced, and propose an new variant on the Bayesian approach to meta-regression that is more inherently robust than the current version, and demonstrate the inclusion of additional study information provides superior meta-results. The key distinction here is that by moving to a fully-specified meta-regression context, we include the individual covariate effects from the included model rather than just the treatment effect in isolation. Thus our contribution is the joint estimation covariate and treatment effects with priors that are informed by individual model quality as well as qualitatively informed views that can come from previous research, clinical information, or additional model features from these studies.

2 Basics

There are essentially three distributional approaches to meta-analysis. The so-called *fixed effects* approach assumes that the underlying quantity of interest (parameter estimates, relative risk, odds ratio, or general effect size) is fixed by nature and each study included in the analysis simply draws from a sampling distribution around this fixed point. Therefore the interpretation of the summarized results includes the assumption that the square-root of the variance of found quantities in the studies is the standard error of a mean estimate of the true fixed value. The *random effects* approach assumes (Higgins *et al.* 2009) that there is sufficient heterogeneity and uncertainty for the quantity of interest that it really defines a distribution rather than a fixed point in the complete population sense. Therefore study observations of this quantity are simply draws from this unknown distribution. The Bayesian approach uses random effects perspective as a starting point and considers parameterizations of the distribution of random effects rather than just assuming a basic (usual normal) form. Therefore we have escalating levels of importance of the distributional assumption.

Despite routine tests that are made, the fixed effects approach is rarely justified. This is because included studies almost always have different mixes of right-hand side covariates in the original models. It is well-appreciated that minor differences in regression-style GLM specifications can yield immense differences in inference and prediction. Therefore it is a very strict assumption

to say that the studies produce IID draws from a fixed sampling distribution.

The Random effect approach is much more principled since it admits a level of heterogeneity between the included studies such that the quantity of interest needs to be modeled distributionally. That is, there may be a modal region for strong effects, but differing specifications and sampling strategies cause deviations from this point of centrality. The Bayesian approach improves-upon the random effects approach because it attempts to *model* the underlying cause for study heterogeneity.

The key distinction between meta-analysis and meta-regression is the inclusion of the model components in addition to the treatment/control contrast. Therefore each included study contributes covariate information in addition to the key variable of interest. The challenge here, of course, is that each study will specify a different mix of right-hand-side explanatory variables and the meta-regression needs to account for this additional heterogeneity. Our approach is to consider each of these models as a selection from the super-set of possible regression parameters defined by the union of explanatory variables across the included models.

3 Bayesian Meta-Analysis

The Bayesian approach to meta-analysis is increasingly popular since it explicitly considers distributional properties of unknown treatment effects (Higgins and Spiegelhalter 2002). Consider a set of K studies, each of which provide an estimated treatment effect, $\theta_1, \ldots, \theta_K$. This treatment effect is often: a set of regression parameter estimates, a log-odds ratio, log relative risk, or absolute relative risk. Under very general circumstances each of these can be considered normally distributed and operating on a normally distributed outcome, y_k, according to the normal-linear specification:

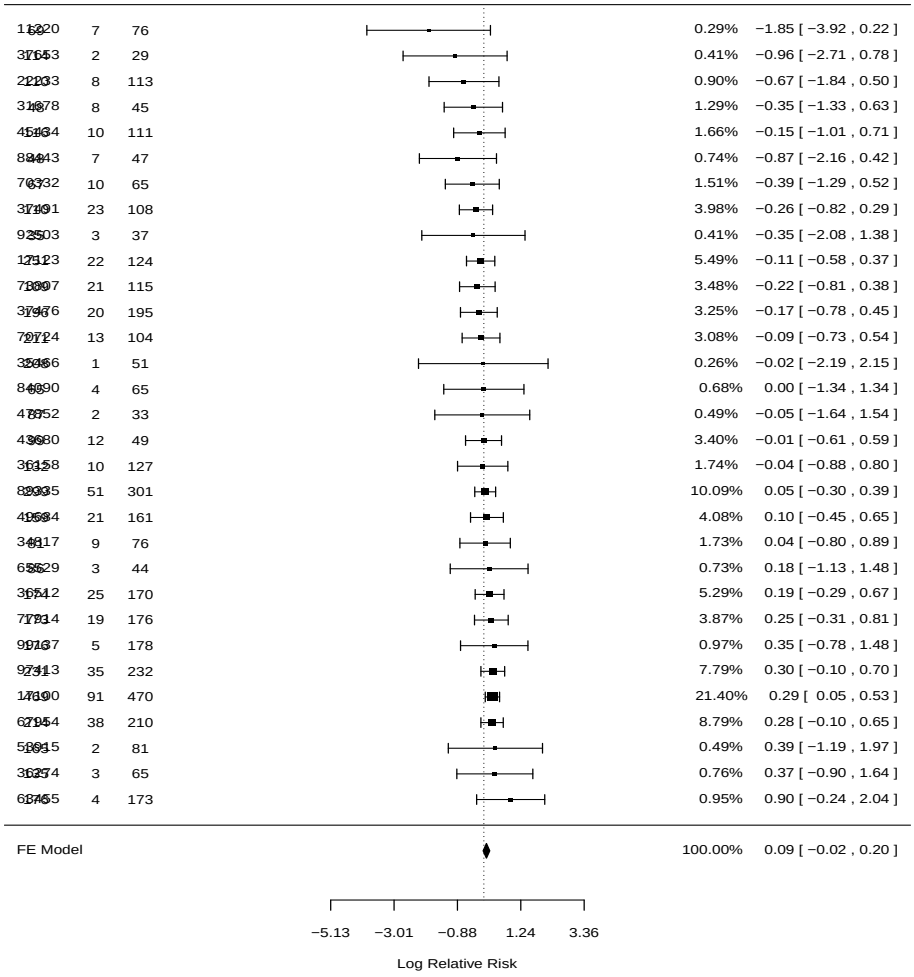
$$\begin{aligned} y_k &\sim N(\theta_k, \tau_k) \\ \theta_k &\sim N(\mu, \tau), \end{aligned} \tag{1}$$

where $\tau_1, \ldots, \tau_K$ are the individual study sample precisions, and μ and τ are the true mean and precision of population distribution. This random effects specification is typical for Bayesian meta-analysis (DerSimonian and Laird 1986), and includes the assumption of exchangeability of the studies. In addition, prior distributions are required for μ and τ, although these are often diffuse forms in practice and can even be mixtures ((Lopes, Müller, and Rosner 2003). A fixed effects specification the reduced form of (1) replacing the second line with the assumption of a true fixed θ.

3.1 Replication of a Non-Bayesian Meta-Analysis

Bohlius *et al.* (2009) are interested in whether erythropoiesis-stimulating agents reduce anaemia in patients with cancer and therefore improve their quality of life. These drugs increase patient haemoglobin concentrations thus reducing the need for red blood cell transfusions that are a standard part of the chemotherapy regimen. Unfortunately there is also evidence that these drugs also might increase mortality. Therefore these investigators perform a meta-analysis of randomized

Figure 1: Bohlius *et al.* Fixed Effects Meta-Analysis (Table 3 in the original)



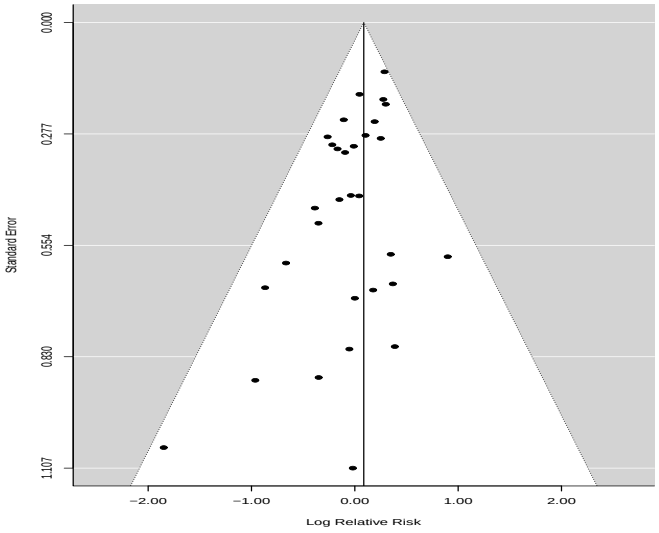
controlled trials in which the treatment is the application these drugs (epoetin alfa, epoetin beta, or darbepoetin alfa) plus red blood cell transfusions, and the control was transfusion alone.

Figure 1 gives the forest plot with study information covering 38 studies and a total of 10,441 cancer patients from January 1, 1985 to January 31, 2008. In the plot r denotes mortality counts during chemotherapy trials (randomization to 28 days after completion) out of n enrollees, for both treatment and control groups The fixed effects meta-analysis provides a relative risk estimate of 1.10 with a confidence interval of [0.98 : 1.22] (the authors have a typographical error on this confidence interval in the original table). This result indicates a increased mortality for the erythropoiesis-stimulating agent treatment. The article provides additional evidence stratified by various study characteristics.

Funnel plots show treatment effects, estimated in the individual studies, on the x-axis plotted against the associated standard error on the y-axis. Other choices for the y-axis can be informative

such as the precision, variance, inverse variance, and sample size (Sterne and Egger 2001). These are used to detect possible publication bias from omitted small sample studies in meta-analysis where an asymmetric pattern around the overall mean line implies either publication bias or overestimation of the treatment effect in poorly designed small sample studies. Figure 2 shows this plot along with 95% pseudo confidence interval for the effect. These lines give the region in which 95% of the trials should fall in the absence of heterogeneity. Note that there is no cause for concern given by this figure for the Bohlius *et al.* (2009) data.

Figure 2: Summer Meeting Womens' Attendance



We now proceed to develop a basic Bayesian meta-analysis on these same studies under the assumption of normal log-relative risk posterior distribution. If θ_k is the log-relative risk found in study k, then the distribution of studies is assumed to follow:

$$\theta_k|\mu,\sigma^2 \sim \mathcal{N}(\mu,\sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}(\theta_k-\mu)^2\right]$$

$$-\infty < \mu < \infty, 0 < \sigma^2 < \infty. \tag{2}$$

It is reasonable and convenient to specify inverse-gamma and normal conjugate priors for the unknown parameters according to::

$$p(\mu|m,\sigma^2/s_0) \propto (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2/s_0}(\mu-m)^2\right] \tag{3}$$

$$p(\sigma^2|\alpha,\beta) \propto (\sigma^2)^{-(\alpha+1)} \exp\left[-\beta/\sigma^2\right], \tag{4}$$

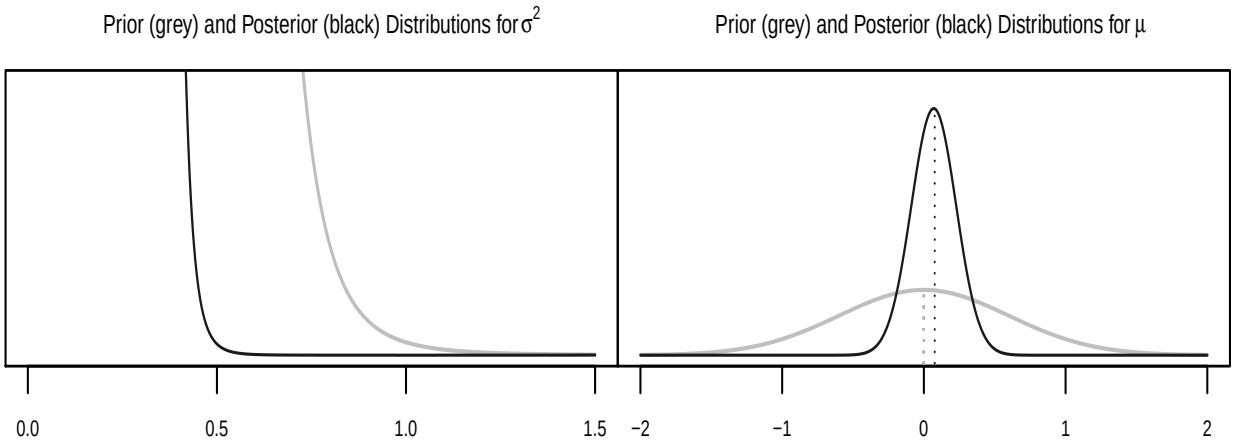
where the prior for μ is explicitly conditional on σ^2 in order to obtain conjugacy. This produces the following posterior distributions for the unknown parameters (Gill 2007, Chapter 3):

$$\mu|\sigma^2,\boldsymbol{\theta} \sim \mathcal{N}\left[\frac{K\bar{\theta}+ms_0}{K+s_0}, \frac{\sigma^2}{K+s_0}\right] \tag{5}$$

$$\sigma^2|\boldsymbol{\theta} \sim \mathcal{IG}\left(\alpha+\frac{K}{2}-\frac{1}{2}, \beta+\frac{1}{2}\sum_{k=1}^{K}(\theta_k\omega_k-\bar{\theta})^2/\sum_{k=1}^{K}\omega_k\right). \tag{6}$$

where $\bar{\theta}$ is weighted by the individual sample study sizes, ω_k. The parameter s_0 is a fixture that measures the researcher's strength of belief in the prior association of μ and σ, where it is common

Figure 3: Bayesian Treatment of Bohlius *et al.* Meta-Analysis



to specify it simply as one. Here we select $m = 0$, $s_0 = 1$, $\alpha = 15$, $\beta = 5$ to obtain a reasonably diffuse form for both prior distributions: $E(\sigma^2) = 0.35714$ (which roughly matches the variance of the log-relative risks), $\text{Var}(\sigma^2) = 0.01$, $E(\mu) = 0$, $\text{Var}(\mu) = E(\sigma^2) = 0.35714$. Setting $s_0 = 1$ also keeps the Bayesian analysis closer to a traditional approach by leaving out a subjective evaluation.

These results do not support the Bohlius *et al.* (2009) conclusion of an increased mortality for the erythropoiesis-stimulating agent treatment. The posterior mean and standard error of the log-relative risk are 0.074 and 0.161 respectively, giving a 95% credible interval of $[-0.241 : 0.389]$. The prior and posterior forms for σ^2 and μ are given in Figure 3. In the second panel the prior and posterior means are shown with vertical dotted lines. It is clear that marginally acceptable results as seen in the Bohlius *et al.* meta-analysis can be mitigated by the inclusion of only slightly vague prior information.

4 Bayesian Model Averaging

Anything not known for certain in the Bayesian framework is treated probabilistically, and this includes possibly model selection. Bayesian model averaging proceeds from: assigning priors to *models*, and producing posterior distributions that are averaged across model space. Putting priors, either probabilities or distributions, on models allows the researcher to include extra information outside of the reported values.

We create a likelihood based on mixtures by requiring K distinct parametric forms that might be used to explain the data generation process, weighting these $\omega_1, \omega_2, \ldots, \omega_K$, where $\sum \omega_k = 1$. We will further require that each of the parametric forms is from the same family, but differing in parametric assignment. So $f_k(x|\theta_k)$ is the kth specification for x with values given in the parameter vector θ_k. Now the likelihood for this mixture is given by:

$$L_k(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \omega_k f_k(x_i|\theta_k). \tag{7}$$

Suppose now that we are choosing between: $M_1, M_2, \ldots, M_K$ and for each k determine a model prior: $p(M_k)$. The *integrated likelihood* for this setup is:

$$p(\mathbf{x}|M_k) = \int \underbrace{p(\theta_k|M_k, \mathbf{x})p(\theta_k|M_k)}_{\text{likelihood}\times\text{prior}} d\theta. \tag{8}$$

$$\pi(M_k|\mathbf{x}) = \frac{p(\mathbf{x}|M_k)p(M_k)}{\sum_{j=1}^{K} p(\mathbf{x}|M_j)p(M_j)}. \tag{9}$$

Consider now the posterior mean and variance for the j^{th} coefficient of the model indexed by k, $\boldsymbol{\theta}_j(k)$ and $Var_{\boldsymbol{\theta}_j}(k)$. Now:

$$p(\boldsymbol{\theta}_j \neq 0|\mathbf{x}) = \sum_{\boldsymbol{\theta}_j \in M_k} \pi(M_k|\mathbf{x}) \tag{10}$$

is just the "posterior probability that $\boldsymbol{\theta}_j$ is in the model." Also:

$$E[\boldsymbol{\theta}_j|\mathbf{x}] \approx \sum_{k=1}^{K} \boldsymbol{\theta}_j(k)\pi(M_k|\mathbf{x}) \tag{11}$$

and:

$$\text{Var}[\boldsymbol{\theta}_j|\mathbf{x}] \approx \sum_{k=1}^{K} \left[(Var_{\boldsymbol{\theta}_j}(k) + \boldsymbol{\theta}_j(k)^2)\pi(M_k|\mathbf{x})) - E[\boldsymbol{\theta}_j|\mathbf{x}]^2\right]. \tag{12}$$

The summation in both the expected value and the variance calculations are averages where the prior model weights sum to one. One point of controversy is that the researcher makes decisions about what variables to include in the various specifications and can therefore increase or decrease expected values or variances by strategical selection.

5 A New Approach To Bayesian Meta-Regression

We now propose a principled Bayesian approach to meta-analyzing similar medical studies that combines Bayesian model averaging with Bayesian meta-regression. The key idea is to simultaneously average across covariates and estimates of treatment effects for the included studies. This incorporates standard Bayesian meta-analysis in the sense that the efficacy of the treatment variable is measured in the usual way, but also in the presence of model-averaged covariates. We also avoid the controversial practice of stipulating our own models for Bayesian model averaging since the literature studied provides the individual mix of covariates included. Furthermore, the fully Bayesian treatment means that we include priors on the model choice itself, which can be parameterized if desired with individual model sample sizes, citation counts, or other substantive criteria. The end-product is a substantially richer approach to meta-analysis and meta-regression than currently practiced alternatives.

Consider a set of K studies (and therefore models) $M_1, \ldots, M_K$, each with a set of estimated coefficients, $\boldsymbol{\gamma}_k$, of observed length J_k. The latter notation is generally necessary since the models will have a different mix of covariates considered. It is more useful to consider a meta-set of coefficients that is the union of those considered across the K models, $\boldsymbol{\Gamma}_K = \cup_{k=1}^{K} \boldsymbol{\gamma}_k$, with cardinality

J. Now index every set of individual model coefficients according to $j = 1, \ldots, J$ where $\theta_{jk} = 0$ and $\sigma^2_{\theta_{jk}} = 0$ if model k omits explanatory variable k from the specification.

Since we do not possess every study's dataset there is no $\mathbf{X}_k$ to directly consider here. The notation from here forward suppresses conditionality on $\mathbf{X}_k$ in each model statement, and it is implied that all model statements are conditional on data. Instead our "data" is simply the set of $\mathbf{\Gamma}_K$ pulled from the individual regression tables. Due to standard reporting conventions we *do* have access to the point estimates (usually the vector mode of a likelihood function) and the standard errors (usually from the diagonal of the negative expected inverse of the Hessian matrix, square root). We take the overtly Bayesian approach of assuming that these are indicators of normally distributed posterior means and variances for the regression effects, an assumption that is supported asymptotically. So the posterior distribution for each considered model, $k = 1, \ldots, K$ is a conventional combination of likelihood and prior information, starting with just the model (M_k) and the treatment effect (θ_k):

$$\pi(M_k|\theta_k) \propto \underbrace{p(\theta_k|M_k)}_{\theta_k \text{ likelihood}} \quad \underbrace{p(M_k).}_{\text{model prior}} \tag{13}$$

The likelihood function in (13) for θ_k comes from integrating the parameter posterior distribution over these parameters:

$$p(\theta_k|M_k) = \int_{\boldsymbol{\gamma}_k} \underbrace{p(\boldsymbol{\gamma}_k|M_k, \theta_k)}_{\boldsymbol{\gamma}_k \text{ likelihood}} \underbrace{p(\boldsymbol{\gamma}_k|M_k)}_{\boldsymbol{\gamma}_k \text{ prior}} d\boldsymbol{\gamma}_k. \tag{14}$$

This term has several names: the *normalizing constant*, the *normalizing factor*, the *marginal likelihood*, and the *prior predictive distribution*, but it is simply integrating the denominator in the full calculation of the posterior distribution for the estimated coefficients: $\pi(\boldsymbol{\gamma}_k|\theta_k) = L(\boldsymbol{\gamma}_k|\theta_k)p(\boldsymbol{\gamma}_k)/p(\theta_k)$. Sometimes this is a difficult calculation, but in an MCMC context (as done here) it can be produced relatively easily with methods due to Chib (1995) for Gibbs samplers and Chib and Jeliazkov (2001) for Metropolis-Hastings algorithms (see Gill 2007, Chapter 12 for an overview). The $P(M_k)$ can be values or distributions here, according to:

$$p(M_k) = \frac{n_k}{\sum_{k=1}^{K} n_k} \xi(M_k), \quad \sum_{k=1}^{K} p(M_k) = 1, \tag{15}$$

where n_k is the sample size from the kth study, and ξ_k is a substantively produced prior distribution based on existing knowledge, or simply set to 1 for all cases when prior information is low. It is also possible to insert qualitative knowledge from similar studies that do not qualify for inclusion in the K selected through this mechanism.

The joint posterior of the unknown quantities for model k is produced from (14) omitting the integration, and including the model priors:

$$\pi(M_k, \boldsymbol{\gamma}_k) \propto p(\boldsymbol{\gamma}_k|M_k, \theta_k)p(\boldsymbol{\gamma}_k|M_k)p(M_k), \tag{16}$$

which is the appropriate combination of data and two sources of prior information: a prior on the model and a prior on the coefficients given the model specification. From this we get the model-averaged posterior distribution for the jth coefficient from summing over K-sized model-space,

conditional on the treatment effect:

$$\pi(\gamma_j|\theta_k) \propto \sum_{k=1}^{K} p(\gamma_j|M_k, \theta_k)p(\gamma_j|M_k)p(M_k), \tag{17}$$

where the sum provides a zero contribution if $\theta_j \notin M_k$, from the definition of $\mathbf{\Gamma}_K$. Thus $\pi(\gamma_j|\theta_k)$ is the posterior distribution of some regressor effect of interest conditional on the treatment effect, over the space spanned by the models included in the meta-analysis. However, these γ_j coefficients have differing statistical reliability across the K model reflected by their posterior variance. Similar to the standard treatment for random effects variance specifications in the meta-analysis literature (eg. Harville 1977, Laird and Ware 1982, Dersimonian and Laird 1986) is to weight the coefficient variance contributions using both the within- and between-variance components:

$$\omega_{jk} = \frac{\sigma^2_{\gamma_{jk}} + \sigma^2_{\gamma_{jK}}}{\sum_{k=1}^{K}(\sigma^2_{\gamma_{kj}} + \sigma^2_{\gamma_{jK}})} \tag{18}$$

where $\sigma^2_{\gamma_{jK}}$ is the "between" model variance for the jth coefficients across the K models. It is important to include both within-trial variances of treatment effects and the remaining between-trial heterogeneity (Thompson and Higgins 2002). This is a simple empirical calculation in MCMC estimation processes since it can be calculated numerically from the saved chain values, or it can be calculated directly from the individual model tables according to: $\sigma^2_{\gamma_{jK}} = \frac{1}{K-1}\sum_{k=1}^{K}(\gamma_{kj} - \bar{\gamma}_j)^2$, where $\bar{\gamma}_j$ is just the mean of the jth coefficients. Therefore (17) with weighting by ω_{jk} becomes:

$$\pi(\gamma_j|\theta_k) \propto \sum_{k=1}^{K} p(\gamma_j|M_k, \theta_k)p(\gamma_k|M_k)p(M_k)\omega_{jk}. \tag{19}$$

Notationally subsumed in this definition is the specification of the $\xi(M_k)$, the observed set of $\sigma^2_{\gamma_{jk}}$ from the studies, as well as each n_k. Also built-in to this form are prior statements about models, regression coefficients, and variance components. In the next sections we will put more specific distributional assumptions on these models statements to determine linear and generalized linear model forms.

5.1 Linear Model Specification

First we consider a standard linear model specification for the continuously-measured mean outcome variable of the kth model (study) is:

$$y_k = \mathbf{X}_k\mathbf{\Omega}_k^{-1}\boldsymbol{\gamma}_k + \theta_k t + \eta_{k[t]} + \epsilon_k, \tag{20}$$

where:

$\mathbf{X}_k\mathbf{\Omega}_k^{-1}\boldsymbol{\gamma}_k$:	a standard linear additive effect of mean covariates, with a diagonal $J \times J$ weighting matrix $\mathbf{\Omega}_k$ containing study specific coefficient weights ω_{jk}
t :	$t = 0$ for control, and $t = 1$ for treatment
θ_k :	estimated treatment effect in study k weighted by the inverse of the sample size
$\eta_{k[t]}$:	random effect in study k for treatment status, assumed: $\sim N(0, \sigma^2_{k[t]})$
ϵ_{ik} :	residual value for ith case, assumed: $\sim N(0, \sigma^2_k)$.

Here $\mathbf{X}_k$ is a p_k length vector containing the mean of the explanatory variables with a leading vector of ones for the constant, and $\mathbf{y}_k$ is a scalar. In practice, Bayesian models are inherently multilevel reflecting both the nature of prior specifications and the reality of medical data that naturally aggregates cases into care or study units. Accordingly, the random effects term, $\eta_{k[t]}$ can be parameterized for treatment-specific effects:

$$E[\eta_{k[t]}] = \zeta_{0k[t]} + \zeta_{1k[t]} Z_{1k} + \ldots + \zeta_{qk[t]} Z_{qk},$$

for q group-level explanatory variables defined in the treatment hierarchy with the common distributional assumption:

$$\eta_{k[t]} \sim N(E[\eta_{k[t]}], \tau^2_{\eta k}).$$

In the case where $E[\eta_{k[t]}] = \zeta_{0k[t]}$, this is reduced to a standard random effects specification.

If we label π^C_k as the true mean effect in the control group and π^T_k as the true mean effect in the treatment group, then the model in (20) can be expressed with the notation:

$$\begin{aligned} \mathbf{y}_k &= \pi^C_k + \eta_{k0} + \epsilon_k, && \text{for } t = 0 \\ \mathbf{y}_k &= \pi^T_k + \eta_{k1} + \epsilon_k, && \text{for } t = 1. \end{aligned} \tag{21}$$

This notation also helps with nonlinear forms.

Make the following notational simplification:

$$\Delta_k = (y_k - \mathbf{X}_k \mathbf{\Omega}_k^{-1} \boldsymbol{\gamma}_k - \eta_{k[t]}),$$

which is complete expression for the expected residuals without including $-\theta_k t$. Then the contribution from the kth model to the coefficient likelihood function is:

$$\begin{aligned} L_k(\boldsymbol{\gamma}_k, \sigma^2_k, \theta_k | \mathbf{X}_k, y_k, \mathbf{\Omega}_k, M_k) &= (2\pi\sigma^2_k)^{-\frac{n}{2}} |\mathbf{\Omega}_k|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2_k} (\Delta_k - \theta_k t)'(\Delta_k - \theta_k t) \right] \\ &= (2\pi\sigma^2_k)^{-\frac{n}{2}} |\mathbf{\Omega}_k|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2_k} (\Delta^2_k - 2\Delta_k \theta_k t + \theta^2_k t) \right]. \end{aligned} \tag{22}$$

Taking expectations over the unobserved regression equation gives:

$$L_k(\boldsymbol{\gamma}_k, \sigma^2_k, \theta_k | \mathbf{\Omega}_k, M_k) = (2\pi\sigma^2_k)^{-\frac{n}{2}} |\mathbf{\Omega}_k|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2_k} (\sigma^2_k + \theta_k + \theta^2_k t) \right], \tag{23}$$

since we know by standard linear models assumptions that $E[\Delta_k] = 0$, and $E[\Delta^2_k] = \sigma^2_k + E[\theta_k t]$. The important feature of (23) is that every term is either contained in the regression summary of the article that contains model k or stipulated by the meta-regression setup here. This gives the full likelihood function across the K models for inference:

$$L(\boldsymbol{\gamma}, \sigma^2, \theta | \mathbf{\Omega}, \mathbf{M}) = \prod_{k=1}^{K} (2\pi\sigma^2_k)^{-\frac{n}{2}} |\mathbf{\Omega}_k|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2_k} (\sigma^2_k + \theta_k + \theta^2_k t) \right]. \tag{24}$$

The left-hand side terms are non-subscripted here because these are the meta-estimates.

It is reasonable and convenient to specify the standard conditional conjugate prior distributions, conditional normal for the regression coefficients and inverse-gamma for the variances components, since these can be made as diffuse as necessary and guarantee a proper posterior form (see Gill 2007, Chapter 4). Specifically:

$$p(\boldsymbol{\gamma}|\sigma^2) = (2\pi)^{-\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{\gamma}-\mathbb{G})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\gamma}-\mathbb{G})\right],$$
$$p(\sigma^2) \propto \sigma_k^{-(a-p)}\exp\left[-\frac{b}{\sigma_k^2}\right],$$

and:

$$p(\theta) = (2\pi\sigma_\theta^2)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma_\theta^2}(\theta-\mathbb{T})'(\theta-\mathbb{T})\right], \tag{25}$$

with $\boldsymbol{\Sigma} = \sigma_k^2\mathbf{I}$, and user chosen prior-parameters $\mathbb{G}$, a, b, and $\mathbb{T}$. Note that $\mathbb{G}$ and $\mathbb{T}$ are not subscripted since we specify a set of J prior distributions for the full set of coefficient values across all of the included models. The resulting contribution to the joint posterior distribution by the kth model is given by:

$$\begin{aligned}\pi(\boldsymbol{\gamma},\sigma^2,\theta,\mathbf{M}|\boldsymbol{\Omega}_k) &= L_k(\boldsymbol{\gamma}_k,\sigma_k^2|\mathbf{X}_k,y_k,\boldsymbol{\Omega}_k)p(\boldsymbol{\gamma}_k|\sigma_k^2,M_k)p(\theta)p(\sigma_k^2|M_k)p(M_k)\\
&\propto \sigma_k^{-n}|\boldsymbol{\Omega}_k|^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma_k^2}(\sigma_k^2+\theta_k+\theta_k^2 t)\right]\frac{n_k}{\sum_{k=1}^K n_k}\\
&\quad\times|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{\gamma}-\mathbb{G})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\gamma}-\mathbb{G})\right]\sigma_\theta^{-1}\exp\left[-\frac{1}{2\sigma_\theta^2}(\theta-\mathbb{T})'(\theta-\mathbb{T})\right]\\
&\quad\times\sigma_k^{-n-a+p}|\boldsymbol{\Omega}_k|^{-\frac{1}{2}}\exp\left[-\frac{b}{\sigma_k^2}\right]\xi(M_k),\end{aligned} \tag{26}$$

where $\xi(M_k)$ is a substantively informed prior for model k if desired. Therefore the joint posterior distribution of the coefficients and variance term alone is:

$$\begin{aligned}\pi(\boldsymbol{\gamma}_k,\sigma_k^2,\theta|\boldsymbol{\Omega}) \propto \prod_{k=1}^K\Bigg\{&\sigma_k^{-2n-a+p}|\boldsymbol{\Omega}_k|^{-\frac{1}{2}}\exp\left[-\frac{b}{\sigma_k^2}\right]\sigma_\theta^{-1}\exp\left[-\frac{1}{2}\sigma_\theta^{-2}(\theta-\mathbb{T})'(\theta-\mathbb{T})\right]\\
&\times\exp\left[-\frac{1}{2}\sigma_k^{-2}\left(\sigma_k^2+\theta_k+\theta_k^2 t+\boldsymbol{\gamma}_k'\boldsymbol{\gamma}_k-2\mathbb{G}'\boldsymbol{\gamma}_k+\mathbb{G}'\mathbb{G}\right)\right]\frac{n_k}{\sum_{k=1}^K n_k}\xi(M_k)\Bigg\},\end{aligned}$$

from (17). This is somewhat awkward looking but can be estimated directly with MCMC tools.

5.1.1 Replication of A Linear Study

5.2 Nonlinear Logit Model Specification

Consider now a dichotomous outcome variable of interest, such as whether the patient has recovered, died, worsened, moved up or down disease stages, and so on. The logit link function for these models is preferred for its mathematical qualities, although we can choose to describe a dichotomous treatment effect several different ways. We continue to label π_k^C as the true mean effect in the control group and π_k^T as the true mean effect in the treatment group, but now also specify:

r_k^C : the number of events in the control group for study k

r_k^T : the number of events in the treatment group for study k

n_k^C : the size of the control group for study k

n_k^T : the size of the treatment group for study k.

This setup naturally leads to a binomial specification for the two groups in each study:

$$r_k^C \sim BN(n_k^C, \pi_k^C) \qquad\qquad r_k^T \sim BN(n_k^T, \pi_k^T), \tag{27}$$

where interest is obviously on estimating π_k^C and π_k^T. Further define μ_k as the baseline risk for this study, which is modeled by $\mathbf{X}_{ik}\mathbf{\Omega}_k^{-1}\boldsymbol{\gamma}_k$ as done before in the linear case. We assume that δ_k (the estimated treatment effect in study k) is normally distributed with true mean δ and variance τ^2. There are three popular strategies for modeling the treatment effect dating back to DerSimonian and Laird (1986).

Table 1: Approaches to Modeling Treatment Effects for Binary Outcomes

	Specification	Estimated Effect
Log-Odds Ratio:	$\text{logit}(\mu_k^C) = \mu_k$ $\text{logit}(\mu_k^T) = \mu_k + \delta_k$ $\delta_k = \text{logit}(\pi_k^T) - \text{logit}(\pi_k^C)$	$\frac{r_k^T}{n_k^T} - \frac{r_k^C}{n_k^C}$
Absolute Relative Risk:	$\mu_k^C = \mu_k$ $\mu_k^T = \mu_k + \delta_k$ $\delta_k = \pi_k^T - \pi_k^C \in [-\pi_k^C : 1 - \pi_k^C]$	$\log\left(\frac{r_k^T}{n_k^T}\right) - \log\left(\frac{r_k^C}{n_k^C}\right)$
Log Relative Risk:	$\log(\mu_k^C) = \mu_k$ $\log(\mu_k^T) = \mu_k + \delta_k$ $\delta_k = \log(\pi_k^T) - \log(\pi_k^C) < -\log(\pi_k^C)$	$\log\left(\frac{r_k^T/(n_k^T - r_k^T)}{r_k^C/(n_k^C - r_k^C)}\right)$

These are given in Table **??** as described by Warn, Thompson, and Spiegelhalter (2002). To separate-out the effects of the covariates and the treatment affect we specify the treatment model, using one of these three choices, outside of the standard generalized linear model inverse-link function:

$$y_k = g^{-1}(\theta_{ik}\mathbf{\Omega}_k^{-1}\boldsymbol{\gamma}_k + \eta_{k[t]}) + \theta_k t + \epsilon_{ik}, \tag{28}$$

where the individuals terms are defined as in (20). In this way, the common link function choice in the original models (logit, probit, log-linear, ordinal, etc.) does not constrain how we choose to evaluate the estimated treatment effect. Furthermore, since it is easy to transform between the three treatment approaches in the included published works, we can make a common choice without worrying about the original authors' choice.

The expression in (28) is not the only choice available. Haydon *et al.* (2005) choose to model δ_k with a set of covariates making it a composite intervention effect.

6 Measuring Heterogeneity

Heterogeneity is a function of the differences between study design, data, and modeling approach for a set of studies in some meta-analysis. Bayesian and non-Bayesian approaches both need to consider whether the estimated level of heterogeneity makes the studies suspect as a cohesive group. Since collection of the included studies is the most time-consuming and controversial component of any meta-analysis project, this is an important consideration. Presumed or suspected heterogeneity between studies is the primary reason to move from a fixed effects to a random effects model. Here we are concerned with whether the inevitable differences between included studies exceeds a subjective level. Usually this level is a function of variance, although there are different ways of measuring it (Petitti 2001). Often the level of heterogeneity is measured with the Q-statistic (DerSimonian and Laird 1986), which is the sum of squares of the treatment effect about the common mean weighted by the inverse of the estimated mean in the study: $Q = \omega_k(\delta_k - \bar{\delta})^2$, where $\bar{\delta} = \frac{1}{K}\sum \omega_k \delta_k$. Under the null hypothesis of homogeneity between studies this is distributed χ^2 with $K-1$ degrees of freedom. Note that this assumes a common sampling variance such that $\omega_k = 1/s_k^2$ is considered to be a draw from the same distribution, although there is some controversy about the efficacy of this assumption. The Q-statistic sampling variance from the kth study is defined for three effects of interest in Table 2.

Table 2: Q-Statistic For Different Treatment Models

Log-Odds Ratio:	$s_k^2 = \frac{1}{r_k^C n_k^C} - \frac{1}{n_K^C} + \frac{1}{r_k^T n_k^T} - \frac{1}{n_K^T}$
Absolute Relative Risk:	$s_k^2 = \frac{r_k^C(1-r_k^C)}{n_k^C} + \frac{r_T^C(1-r_k^T)}{n_k^T}$
Log Relative Risk:	$s_k^2 = \frac{1}{n_k^C r_k^C(1-r_k^C)} + \frac{1}{n_k^T r_k^T(1-r_k^T)}$

It is well-known that the Q-statistic provides a low power test for small K, and is also overly sensitive to minor heterogeneity for large K. On alternative assuming a random effects model is τ^2, the between-study variance measured in units of the outcome variable. Since this is measured par-

ticularistically on the study outcome, often τ^2 values cannot be compared between meta-analyses. To make up for this deficiency, Higgins and Thompson (2002) created the I^2 statistic, which is the percent of total between variance of the effect sizes due to true heterogeneity $I^2 = 100\frac{Q-(K-1)}{Q}$ (for $Q > 0$, otherwise zero). This is interpreted as the percent of total variability that is due to between-study variance. Obviously τ^2 and I^2 are proportionally related.

7 Describing the Sampler For Estimation

8 Application To (Something Cancer Oriented)

9 Conclusions

10 References

Bohlius F, Schmidlin K, Brillant C, Schwarzer G, Trelle S, Seidenfeld J, Zwahlen M, Clarke M, Weingart O, Kluge S, Piper M, Rades D, Steensma D P, Djulbegovic B, Fey M F, Ray-Coquard I, Machtay M, Moebus V, Thomas G, Untch M, Schumacher M, Egger M, Engert A. Recombinant human erythropoiesis-stimulating agents and mortality in patients with cancer: a meta-analysis of randomised trials. *Lancet* 2009; **373**; 1532-1542.

Chib S. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 1995; **90**; 1313-1321.

Chib S, Jeliazkov, I. Marginal Likelihood from the Metropolis-Hastings Output. *Journal of the American Statistical Association* 2001; **96**; 270-281.

DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**; 17788.

Egger M, Smith G. Misleading meta-analysis. *British Medical Journal* 1995; **6982**; 752-754.

Gill, J. *Bayesian Methods For the Social and Behavioral Sciences.* New York: Chapman & Hall/CRC; 2007.

Harville D A. Maximum likelihood approaches to variance components estimation and to related problems. *Journal of the American Statistical Association* 1977; **72**: 320-340.

Hayden J, van Tulder M W, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. *Annals of internal medicine* 2005; **142**; 776-785.

Higgins J P T, Spiegelhalter D J. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology* 2002; **31**; 96-104.

Higgins J P T, Thompson S G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **11**; 1539-1558.

Higgins J P T, Thompson S G, Spiegelhalter D J. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society. Series A* 2009; **172**; 137-159.

Laird N M, Ware J H. Random-effects models for longitudinal data. *Biometrics* 1982; **38**; 963974.

Lopes H F, Müller P, Rosner G L. Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. *Biometrics* 2003; **59**; 66-75.

Petitti, D. Approaches to heterogeneity in meta-analysis. *Statistics in Medicine* 2001; **20**; 3625-3633.

Sterne J A C, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* 2001; **54**; 10461055.

Thompson S G, Higgins J P T. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 2002; **21**; 1559-1573.

Warn, Thompson, and Spiegelhalter. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales *Statistics in Medicine* 2002; **21**; 1601-1623.