Final Project

Benjamin Jakubowski

Abstract

In 2002, Higgins and Spiegelhalter published a report comparing different meta-analyses of clinical trials testing the benefits of intravenous magnesium (Mg) following myocardial infarction (MI). This report essentially replicates their work using open-source statistical tools (R and RSTAN). First, we briefly introduce frequentist and Bayesian methods for meta-analysis and frame the substantive problem (meta-analytic inference on the log odds ratio for mortality with and without intravenous Mg following MI). Then we introduce and conduct the two frequentist analyses (using the Peto method and DerSimonian and Laird (D-L) method). Next, we run two Bayesian random effects models under two different priors. Finally, we compare the results from these different methods, and conclude by addressing the substantive question of the efficacy of Mg.

Contents

1. Introduction

Introducing meta-analysis

Meta-analysis is an important method for statistical inference motivated primarily by the need to combine the results of replication studies. For example, early clinical development requires investigating the efficacy of a candidate therapy on small numbers of patients. If early results are promising, the clinical trial may be replicated, and if there is continued evidence of clinically beneficial effect a large study may eventually be conducted.

Since a single candidate therapy (or, more generally, a single statistical relationship) may be tested many times over many studies, there is a need for statistical methods to combine results across disparate studies. For example, in a clinical context there is a need to estimate the overall effect size of a new therapy. A naive approach to estimating the overall effect size would be to take a simple mean of effect sizes from each study; however, this would clearly ignore information about the studies (for example, their sample sizes). The statistical field of meta-analysis involves (among other lines of inquiry) developing more principled methods for estimating the overall effect size from multiple studies.

When conducting a meta-analysis, it is important to consider many factors, including

- **Number of trials included**: It is important to consider the number of trials included in the meta-analysis be-

cause we may want to estimate the between-study effect size distribution (using a random-effects model). With an insufficient number of studies, our estimate will be poor [1].

- **Trial inclusion criteria (including trial quality)**: It is important to consider which trials to include in a meta-analysis. The goal of meta-analysis is to pool effect size estimates from comparable studies; as such, it is important the included studies are in fact comparable [1]. Additionally, it is important to consider their quality. Including poorly run trials (with potentially biased results) will bias the final meta-analytic result. To risk statistical tautology, meta-analysis does not produce meaningful results from meaningless studies.

- **The sample sizes of the trials**: Studies with larger sample sizes are expected to have smaller variance in effect size estimates, and should be more heavily weighted in the final meta-analytic estimate.

- **Publication bias**: Unfortunately, our current system of scientific publication (typically) suppresses results that don't pass an $\alpha = 0.05$ test. This is expected to bias the sample of trials available for meta-analysis towards trials with larger observed effect sizes (since they are more likely to be published)[1]. While there are methods to assess publication bias in a sample of studies (for example, funnel plots comparing estimated effect sizes vs. sample size or standard error) the suspicion of bias plagues meta-analysis [2].

Bayesian vs. Frequentist Analysis

There are two fundamental views of statistics. The first is frequentist statistics. In a frequentist view, parameters are fixed and unknown. In the context of our analysis, under the frequentist view there is a true log odds ratio (θ) for mortality with and without intravenous Mg following MI for our given

population. We can sample from the population and estimate this parameter as $\hat{\theta}$, but we can never access the true value θ. Note our estimate $\hat{\theta}$ is drawn from the sampling distribution of this statistic, and as such is random (due to sampling error), but the underlying parameter itself is non-random.

In contrast, in the Bayesian view the parameter is itself a random variable- there is no fixed "true" value θ. Instead, θ is a random variable that has some distribution we can estimate. Our initial estimate (the 'prior' distribution) for this distribution is $P(\theta)$. After seeing data $\mathscr{D}$, we place an updated 'posterior' distribution on θ, namely $P(\theta|\mathscr{D})$ [1].

The problem of meta-analysis lends itself well to the Bayesian paradigm, since random-effects meta-analysis essentially matches the Bayesian perspective that there is in fact an underlying distribution on θ. In a random-effects model, we no longer view θ as a fixed parameter- instead there is an underlying distribution on θ (which can be attributed to heterogeneity between studies). For further context, there are essentially two steps in data generating process governing this model [1]:

1. **Draw θ^i from the distribution on θ**: Consider the ith study- choosing this study corresponds to drawing some effect size θ^i from the distribution on θ.

2. **Draw from the sampling distribution of $\hat{\theta}^i$**: Next, we estimate θ^i with our test statistic $\hat{\theta}^i$. This statistic is itself drawn from a sampling distribution, and as such is a random variable.

Under this model, it is very natural to take a Bayesian perspective, placing a prior distribution on θ and then updating our prior based on available data (trial results).

Meta-analysis in practice: treating MI with Mg
In this report, we will apply both frequentist and Bayesian meta-analysis methods to the problem of estimating the efficacy of intravenous magnesium in patients with acute myocardial infarction. This treatment was controversial in the 1990s due to the apparent contradiction between early meta-analyses and a later mega-trial. Higgins and Spiegelhalter [2] present the following timeline:

1. **First eight trials**: Fixed effect and random effects meta-analysis of eight early trials produced ORs of 0.65 (95% CI: 0.51, 0.82) and 0.55 (95% CI: 0.34, 0.89).

2. **First 14 trials**: Fixed effect and random effects meta-analysis of 14 early trials produced ORs of 0.57 (95% CI: 0.46, 0.71) and 0.47 (95% CI: 0.33, 0.68).

3. **Including ISIS-4 mega-trial**: Finally, including the results of the ISIS-4 mega-trial (published in 1995), meta-analysis produces fixed effect and random effects ORs of 1.01 (95% CI: 0.95, 1.07) and 0.53 (95% CI: 0.36, 0.77).

---

[1]For the sake of notation, I'm using P to represent either a density function or probability mass function

Thus, it is apparent the ISIS-4 mega-trial conflicted with the promising early meta-analyses. This conflict lead to both methodological and clinical controversy around the use meta-analysis and Mg for MI, respectively. We proceed by exploring this controversial dataset using both frequentist and Bayesian meta-analysis.

2. Introducing the models

Our analysis is divided into two sections. In the first analysis section (section 3), we will present two frequentist meta-analyses: a fixed-effects model using the Peto method, and a random-effects model using the DerSimonian and Laird (D-L) method. In the second analysis section (section 4), we will present two Bayesian random-effects models, corresponding to two priors: a reference prior and a skeptical prior. Before conducting these analyses, we first introduce these models.

Fixed- vs. random-effects models and sample sizes
As described above, there are two essential modeling frameworks for meta-analysis [1]:

1. **Fixed-effects models**: In a fixed-effects model, we assume there is a constant true effect size that is the same for all studies. Note this is a strong assumption, since it is asserts the homogeneity of studies potentially conducted (i) at different locations, (ii) on samples (in our context patient samples) with differing characteristics, (iii) using different protocols (ex: varying times of treatment administration). Under this assumption, variation in the estimated effect size is attributed to sampling error.

2. **Random-effects models**: In a random-effects model, we assume the true effect size varies between studies. Note this is a weaker assumption, since it is allows for heterogeneity between the studies' true effect sizes.

Importantly, the impact of a trial's sample size on its weight in the estimated meta-analytic treatment effect varies between fixed-effects and random-effects models. In a fixed-effects models, since we assume there is a constant true effect size θ, larger studies provide lower-variance estimates than smaller studies. As such, fixed-effects model estimates will be dominated by large studies (if the trial sample sizes are highly variable).

In contrast, in a random-effects model, each study represents a single draw θ^i from the underlying distribution on θ. As such, while the studies with larger sample sizes will have lower variance estimates $\hat{\theta}^i$ of θ^i, each study still provides useful information on the underlying distribution of θ. Hence, larger studies receive a (relatively) less-outsized weight in the meta-analytic treatment effect estimate, and smaller studies recieve a (relative) greater weight (compared to fixed-effect models).

We proceed to specify the frequentist fixed-effect and random-effects models used in our analysis. Note the subsequent sections draw heavily from Schwarzer *et al.* [3]

Table 1. Data produced by study i

	Deaths	No Death	Group size
Magnesium group	r_i^M	q_i^M	$n_i^M = r_i^M + q_i^M$
Control group	r_i^C	q_i^C	$n_i^C = r_i^C + q_i^C$
Marginal counts	$r_i^M + r_i^C$	$q_i^M + q_i^C$	$n_i^M + n_i^C$

Frequentist fixed-effect models: Peto method

The Peto method is a method for fixed-effect modeling of binary data. The ith trial in our meta-analysis produces the data described in Table 1. The number of deaths are assumed to follow a binomial distributions: $r_i^M \sim \text{Binomial}(n_i^M, p_i^M)$ and $r_i^C \sim \text{Binomial}(n_i^C, p_i^C)$. Note under a fixed-effect model, $p_i^M = p_j^M$ and $p_i^C = p_j^C$ for all i, j.

Each trial produces estimates $\hat{p}_i^M = r_i^M / n_i^M$ and $\hat{p}_i^C = r_i^C / n_i^C$. In addition, we can estimate the odds ratio $\psi = \frac{p^M(1-p^M)}{p^C(1-p^C)}$ using the ith study:

$$\hat{\psi}_i = \frac{\hat{p}_i^M (1 - \hat{p}_i^M)}{\hat{p}_i^C (1 - \hat{p}_i^C)}$$

Unfortunately, this estimator fails when we have 0 counts for either deaths in the treatment or control group (as we do in our meta-analysis for the Bertschat trial). One approach to dealing with this issue is to add a continuity correction. However, we will instead use the Peto method, which does not required any such correction. The Peto estimate of the odds ratio is

$$\hat{\psi}_i^* = \exp\left(\frac{r_i^M - \text{E}(r_i^K | \cdots; \psi = 1)}{\text{Var}(r_i^K | \cdots; \psi = 1)} \right)$$

where "$\cdots$" refers to the group sizes and marginal counts from Table 1, and the expectation and variance are over a hypergeometric distribution (found under the assumption the true odds ratio $\psi = 1$).

Per usual, we take the log transform of this odds ratio $\log(\hat{\psi}_i^*)$. The variance of this transformed statistic is

$$\hat{\text{Var}}(\log \hat{\psi}_i^*)) = \frac{1}{\text{Var}(r_i^K | \cdots; \psi = 1)}$$

With the Peto estimate of the odds ratio in hand, we now turn to the Peto method for meta-analysis. The Pet method is essentially an inverse variance method, where the log of the final meta-analytic Peto odds ratio. $\log(\hat{\psi}_{Peto})$ is a weighted average (convex combination) of the log-scaled Peto-estimated odds ratios:

$$\log(\hat{\psi}_{Peto}) = \frac{1}{\sum_{i=1}^K w_k} \sum_{i=1}^K w_k \log \hat{\psi}_k^*$$

where $w_i = \frac{1}{\hat{\text{Var}}(\log \hat{\psi}_i^*)}$.

The variance of the Peto-estimated log odds ratio is

$$\hat{\text{Var}}(\log \hat{\psi}_{Peto}) = \frac{1}{1 / \sum_{i=1}^K \hat{\text{Var}}(\log \hat{\psi}_k^*)}$$

From this point estimate and estimated variance, we can easily derive confidence interval and conduct statistic tests.

Frequentist random-effects models: D-L method

3. Frequentist analysis

4. Bayesian analysis

5. Conclusion

References

[1] Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd, 2009.

[2] J.P. Higgins and D.J. Spiegelhalter. Being sceptical about meta-analyses: a bayesian perspective on magnesium trials in myocardial infarction. *Int J Epidemiol.*, 31(1):96–104, 2002.

[3] Guido Schwarzer, James R. Carpetner, and Gerta Rücker. *Meta-Analysis with R*. Springer International Publishing, 2015.