WELCOME

Meta Analysis: A Brief Tutorial

Mark S. Kaiser
Department of Statistics
Iowa State University

Meta

- Meta can be an adjective, a prefix, or a noun.
- Meaning is very different in those different contexts:
  - In chemistry, meta refers to two positions in the benzene ring that are separated by one carbon atom.
  - In words adopted from Greek (Greek loanwords), meta is a prefix meaning after or beyond (metacarpus).
  - In philosophy, meta is a prefix added to a subject to denote another subject that analyzes the original at a more abstract level (metalinguistics).

Cambridge English Dictionary

Meta: (adjective): (of something that is written or performed) referring to itself or to something of its own type

- Meta Data: data about what is contained in data files
- Meta Analysis: an analysis of analyses

Broader Context

- Goal: Provide a systematic summarization of evidence about an issue that has been presented in the literature.
- Often, the issue of concern is efficacy of a medical treatment:
  Zhao *et al.* (2017), Association between calcium or vitamin D supplementation and fracture incidence in community-dwelling older adults: A systematic review and meta-analysis. *Journal of the American Medical Association* **318(24):** 2466-2482.
- Increasingly showing up in other scientific areas:
  Jia, L. (2020), Quantifying effects of contour tillage in controlling water erosion in China: A meta-analysis. *Catena* **195**, Article 104829. https://doi.org/10.1016/j.catena.2020.104829
- There are now even meta analyses of meta analyses:
  Sokouti *et al.* (2023), A meta-analysis of systematic reviews and meta-analyses to evaluate the psychological consequences of COVID-19. *BMC Psychology* **11**, Article 279. https://doi.org/10.1186/s40359-023-01313-0

Statistical Context

- Goal: Determine a statistical procedure to quantitatively combine the results of multiple scientific studies.
- Secondary Goal: Quantify heterogeneity among studies:
  - Settings (populations, geographic location, demographics)
  - Protocol (design, observational/measurement methods)
  - Endpoints (measured responses)

  Easiest of these is endpoints in cases for which all studies have estimated the same effect.

Underlying Concept

- There exists a common scientific mechanism behind what is observed in similar scientific studies.
- The effect of that mechanism may be observed with error or may be variable in manifestation in a set of studies.
- Observation of a constant with error versus variability in manifestation will help determine the statistical framework to be used.

Desired Benefits and Potential Pitfalls

Benefits

1. Ability to increase scope of inference (e.g., conclusions apply to a large population than those used in individual studies).
2. Increased "power" in determination of an effect or treatment difference.

Pitfalls

1. How to conduct an "unbiased" search for studies on a particular topic? Typically, studies are identified by searching the scientific literature. What about "grey literature"?
2. Not all studies are of equal quality in design and execution. Which should be included in a meta analysis?
3. Potential publication bias – do studies that fail to find a significant effect get published as readily as those that do find an effect?

Not All of Meta Analysis is Statistical

Following the Cochrane Handbook for Systematic Reviews of Interventions, a meta analysis generally involves the following steps:

1. Formulate Question – type of population, type of intervention, comparisons, definition of outcomes.
2. Search for relevant studies.
3. Select studies to include – acceptable protocols, information provided, correct population match, correct outcomes.
4. Assess sources of bias – selection bias, detection bias, reporting bias, attrition bias. Requires some judgement.
5. Examine issue of publication bias.
6. Assess heterogeneity among studies.
7. Determine statistical model and produce meta estimate of effect.

Our focus is steps 5 to 7, but usefulness of statistical results greatly impacted by steps 1 to 4 as well.

Statistical Idealization

- A collection of studies all conducted to a minimum level of required protocol (randomized, double blind, etc.)
- Each study reports the same summary measures of effect size or association, or data that allows computation of those measures:
  - Treatment group means or incidence rates.
  - Odds ratios or relative risk.
  - Differences in means (possibly standardized).
  - Correlations.
- Each study reports the same measures of uncertainty, or quantities that allow computation of those measures:
  - Sample sizes.
  - Standard errors or variances.
  - Confidence interval levels and endpoints.

This idealization creates potential conflict between inclusion of as many studies as possible and filtering studies for similar quality and reporting of results.

Publication Bias

- The phenomenon that studies finding significant results are more likely to appear in literature than those that don't.
- Interesting twist: even when published, studies with no significant results appear 2 to 3 years later than those with significant results.
- Causes may involve authors failing to submit work for publication if no positive results have been found.
- Clinical trials registries. https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries

Hopewell *et al.* (2007), Publication bias in clinical trials. *Cochrane Database of Systematic Reviews*, 2007 Issue 2, Article MR000006.

Stern, J.M. and Simes, R.J. (1977), Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal* **315**: 640-645.

Easterbrook *et al.* (1991), Publication bias in clinical research. *Lancet* **337**: 867-872.

Identifying Publication and Related Biases

Assessing Selective Outcome Reporting

- Not strictly publication bias but a similar phenomenon.
- For a set of N studies, construct a matrix of outcomes reported in which studies.
- If the studies mostly report the same outcomes but one or more fail to report a particular outcome it may be because a significant difference between treatment groups was not found for that outcome.
- Relative to a meta analysis on that outcome, having no useful information on it from a study in which it was (or should have been) measured or observed has the same effect as publication bias.
- Difficult to deal with, but should be reported as potential cause of bias if detected.

Williamson *et al.* (2005), Outcome selection bias in meta-analysis. *Statitical Methods in Medical Research* **14(5):** 515-524.

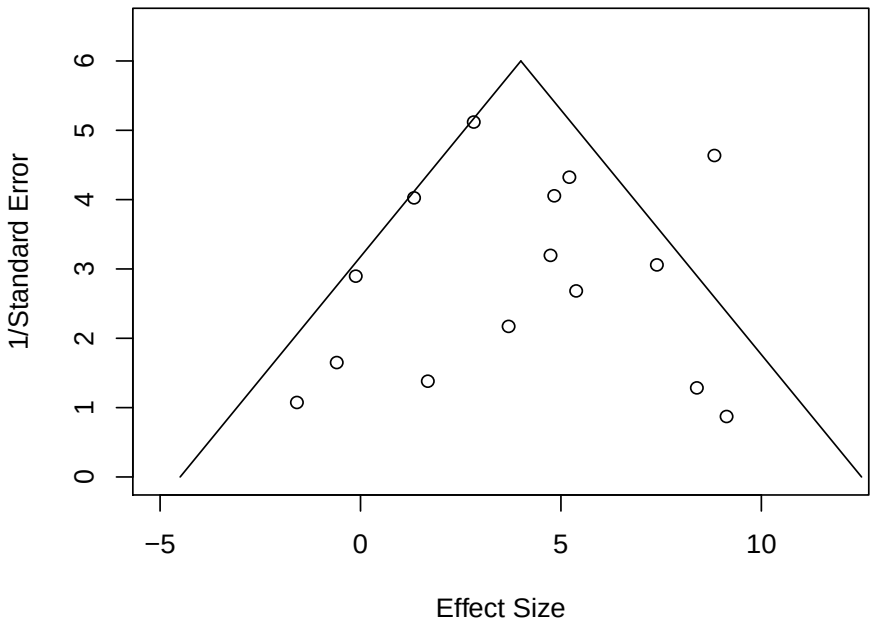
Identifying Publication and Related Biases

Construction of Funnel Plots

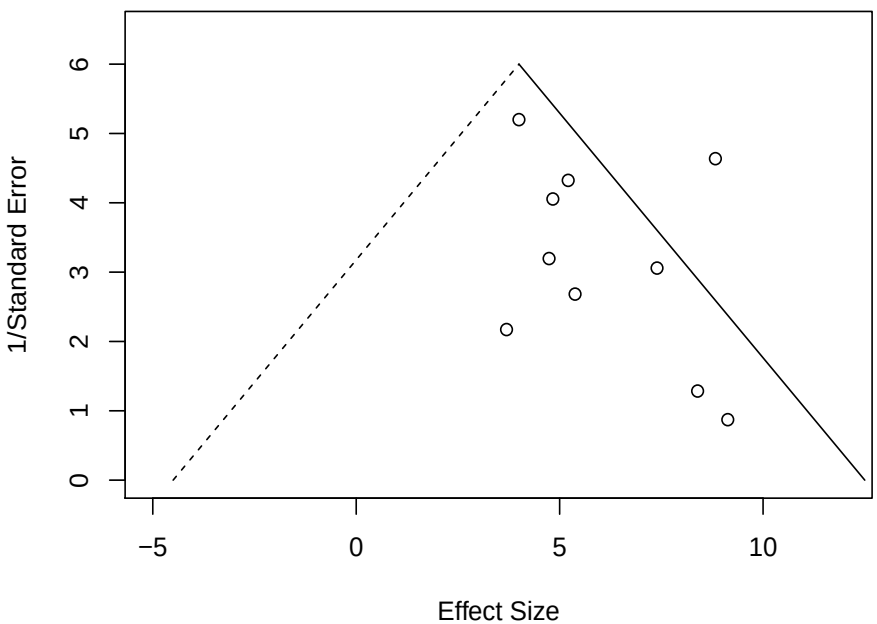
- Applicable when studies report the same summary measure of effect along with the same quantification of uncertaity (e.g., means, sample sizes and standard deviations or standard errors).
- Underlying concept is that larger sample sizes should lead to less uncertainty in estimated effect size.
- Plot reciprocal standard error on vertical axis and effect size on horizontal axis.
- A "funnel" that is symmetric, especially with respect to the base, provides no evidence of publication bias.
- A "funnel" that is asymmetric or is missing a part of the funnel suggests publication bias.

See also: Haidich, A.B. (2010), Meta-analysis in medical research. *Hipppokratia* **14**: 29-37.

Hyothetical Funnel Plot Examples



- No Evidence of Publication Bias



- Evidence of Publication Bias

Estimating Treatment Effects

Individual Studies

1. Mean differences (e.g., treatment and placebo):

$$T = \bar{Y}_1 - \bar{Y}_2.$$

2. Standardized mean differences (SMD).
   - Implemented by many software packages, such as R package metafor.
   - Computation to come.
   - Not without controversy:

     Gallardo-Gomez *et al.* (2024). Variability in meta-analysis estimates of continuous outcomes using different standardization and scale-specific re-expression models. *Journal of Clinical Epidemiology* **165**, 111213.

3. Odds ratios:

$$\text{OR} = \frac{Y_{1,1}\, Y_{2,2}}{Y_{2,1}\, Y_{1,2}}.$$

4. Estimates of binomial parameters:

$$\hat{p} = \frac{1}{m} Y$$

Estimating Standard Errors

Individual Studies

1. For mean differences, may be based on pooled sample variances (typical) or non-pooled:

$$\text{se}(\bar{Y}_1 - \bar{Y}_2) = \left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]$$

$$\text{se}(\bar{Y}_1 - \bar{Y}_2) = \left[\left(\frac{n_1 + n_2}{n_1 n_2}\right) S_p^2\right].$$

2. For standardized mean differences, same as (1), but more to come.
3. For log odds ratios:

$$\text{se}[\log(\hat{\text{OR}})] = \left(\sum_{i=1}^{2}\sum_{j=1}^{2} Y_{i,j}\right)^{1/2}.$$

4. For estimated binomial proportions:

$$\text{se}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}.$$

Standardized Mean Differences

- For continuous response data.
- When different studies use different scales of measurement.

  Example: Response of elderly patients to physical activity measured on two scales:
  - Short Physical Performance Battery: Evaluates lower extremity function. Possible scores range from 0 to 12.
  - Barthel Index: Measures functional disability in 10 daily living activities. Possible scores range from 0 to 99.

  Ortiz-Alonso *et al.* (2019). Effect of a simple exercise programme on hospitalization-associated disability in older patients: a randomized controlled trial. *Geriatic Medicine* **21**: 531-537.

- A mean difference and it's standard error are computed for each study as in previous slides, $(\bar{Y}_1 - \bar{Y}_2)$ and $\text{se}(\bar{Y}_1 - \bar{Y}_2)$.

Standardized Mean Differences

- Both the mean difference and it's standard error are standardized by dividing each by one of:
  1. A pooled estimate of standard deviation at baseline (before treatment).
  2. A pooled estimate of standard deviation at end of treatment.
  3. The average, across studies, of pooled standard deviations at baseline. This is called an internal reference.
  4. An externally determined standard error from some reference population or in the literature. This is called an external reference.
- The standardized mean differences and standardized standard errors are then used in the meta analysis estimation.
- Different standardizations can lead to different results (Gallardo-Gomez *et al.* already cited).

Again: there is some controversy about all of this, but the attempt is to allow a greater number of studies to be combined.

Estimation of an Overall Effect: Fixed Effect Model

- Context of Presentation: medical treatment effect.
- Objective: Estimate a single overall effect from a set of studies along with quantities needed for inference.
- Consider first a conceptualization of a common fixed effect in every study, observed with error.

Estimates from Individual Studies

- Estimated Treatment Effects: $\{T_s : s = 1, \ldots, S\}$.
- Estimated Standard Errors: $\{\text{se}(T_s) : s = 1, \ldots, S\}$.
- Estimated Variances (of treatment effects) $\{V_s : s = 1, \ldots, S\}$; $V_s = [\text{se}(T_s)]^2$.

Fixed Effect Model

$$T_s \sim \text{indep } \text{N}(\mu,\ V_s)$$

Notice: the estimated V_s are assumed to be known variances in the model.

Estimation of an Overall Effect: Fixed Effect Model

- Density of T_s:

$$f_s(t|\mu) = \frac{1}{(2\pi V_s)^{1/2}} \exp\left[-\frac{1}{2V_s}(t-\mu)^2\right].$$

- Log likelihood and Derivative:

$$\ell(\mu) = \sum_{s=1}^{S}\left[-\frac{1}{2V_s}(t_s-\mu)^2 - \frac{1}{2}\log(2\pi V_s)\right]$$

$$\frac{\partial\,\ell(\mu)}{\partial\mu} = -\sum_{s=1}^{S}\left(\frac{t_s}{V_s}\right) + \mu\sum_{s=1}^{S}\frac{1}{V_s}.$$

- Maximum Likelihood Estimate of μ:

$$\hat{\mu} = \frac{\sum_{s=1}^{S} w_s t_s}{\sum_{s=1}^{S} w_s}$$

$$w_s = \frac{1}{V_s}.$$

Estimation of an Overall Effect: Fixed Effect Model

- $\hat{\mu}$ is an unbiased estimator of μ.
- $\hat{\mu}$ is a linear combination of independent T_s; $s = 1, \ldots, S$.

$$\hat{\mu} \sim \mathrm{N}[\mu, V(\hat{\mu}]$$

$$V(\hat{\mu}) = \left(\sum_{s=1}^{S} w_s \right)^{-2} \sum_{s=1}^{S} \left(w_s^2 V_s \right)$$

$$= \frac{1}{\sum_{s=1}^{S} w_s} = \frac{1}{\sum_{s=1}^{S} \frac{1}{V_s}}.$$

- A $(1-\alpha)100\%$ confidence interval for μ is then,

$$\left(\hat{\mu} - z_{1-\alpha/2} [V(\hat{\mu})]^{1/2},\ \hat{\mu} + z_{1-\alpha/2} [V(\hat{\mu})]^{1/2} \right).$$

- $z_{1-\alpha/2}$ rather than $t_{1-\alpha/2}$ because $V(\hat{\mu})$ is considered known, as a function of the known V_s.

Estimation of Overall Effect: Fixed Effect Model

- It is obviously incorrect to assume that the V_s are known, rather than estimated.
- The issue is not so much using $z_{1-\alpha/2}$ rather than $t_{1-\alpha/2}$ but rather that uncertainty in the V_s should affect the uncertainty in $\hat{\mu}$. Let $\boldsymbol{V} = \{V_s : \ s = 1, \ldots, S\}$. If we take into account that the V_s are estimated,

$$E(\hat{\mu}) = E[E(\hat{\mu}|\boldsymbol{V})] = E[\mu] = \mu,$$

so $\hat{\mu}$ remains unbiased. But,

$$\begin{aligned} \text{var}(\hat{\mu}) &= \text{var}[E(\hat{\mu}|\boldsymbol{V})] + E[\text{var}(\hat{\mu}|\boldsymbol{V})], \\ \text{var}[E(\hat{\mu}|\boldsymbol{V})] &= \text{var}(\mu) = 0, \\ E[\text{var}(\hat{\mu}|\boldsymbol{V})] &= E\left[\sum_{s=1}^{S} \frac{1}{V_s}\right]^{-1}. \end{aligned}$$

and the variance of $\hat{\mu}$ becomes complex.

Assessing Heterogeneity

- Several types of heterogeneity: Clinical Variation, Study Design Variation, Statistical Variation.

Statistical Heterogeneity

Recall from Statistical Conceptualization:

- Common scientific mechanism across studies.
- Constant effect observed with error OR
- Variable manifestation of effect.

Statistical heterogeneity often defined relative to the fixed effect model:

- Variability in observed effects of treatment or intervention beyond what would be expected by chance alone.
- May be a consequence of clinical or study design variability.
- May be a consequence of other factors that help determine exactly what effect the mechanism has in a given situation.

Quantifying Heterogeneity I

Cochran's Q

$$Q = \sum_{s=1}^{S} w_s (T_s - \hat{\mu})^2.$$

- Compare Q to a Chi-square distribution with $S - 1$ degrees of freedom.
- Problems: (i) if S small, low power; (2) if S too large, excessive power.
  For examples of each, see
  Higgins *et al.* (2003). Measuring inconsistency in meta-analysis. *British Medical Journal* **327**: 557-560.

Quantifying Heterogeneity II

Inconsistency Index

$$I^2 = 100\left(\frac{Q-(S-1)}{Q}\right).$$

- Proposed by
  Higgins, J.P.T. and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**: 1539-1558.
- Claim is that I^2 represents variability in effect estimates due to heterogeneity rather than sampling error.
- Intuition comes from the fact that, under perfect homogeneity ($E[T_s] = \mu$ for $s = 1, \ldots, S$),
$$E(Q) = S - 1.$$
- Various scales: 25 =low, 50 =moderate, 75 = high OR $0-40$ not important, $30-60$ moderate, $50-100$ substantial.

Exploring Heterogeneity

1. Sometimes, at least some of the potential factors causing heterogeneity may be known.
2. If information on such factors is available for each study, define a covariate vector,

$$\boldsymbol{x}_s = (x_{1,s}, \ldots, x_{p,s})^T$$

Examples: $x_{j,s}$ might be

- Proportion of sampling units (patients, subjects) less than a given age.
- Length of follow-up for effects such as disease incidence.
- Average disease severity for subjects in a study.

Meta Analysis Regression Model

$$T_s = \boldsymbol{x}^T \boldsymbol{\beta} + (V_s)^{1/2} \epsilon_s,$$

where $\epsilon_s \sim$ iid F location-scale, $E(\epsilon_s) = 0$ $\text{var}(\epsilon_s) = 1$, and V_s assumed known.
Estimation using weighted least squares.

Estimation of an Overall Effect: Random Effects Model

- Context of Presentation, Objective are same as for Fixed Effect Model.
- Conceptualization changes: effect manifests differently in different studies. Mechanism as a distribution.

Random Effects Model

- Same notation: Estimates of effect and variances (of that estimate) from individual studies, $\{T_s, V_s : s = 1, \ldots, S\}$.

$$T_s \sim \text{indep } \mathsf{N}(\theta_s, V_s)$$
$$\theta_s \sim \text{iid } \mathsf{N}(\mu, \tau^2).$$

May also be written as,

$$T_s = \theta_s + \epsilon_s; \qquad \theta_s \sim \text{iid } \mathsf{N}(\mu, \tau^2);\ \epsilon_s \sim \text{iid } \mathsf{N}(0, V_s)$$
$$T_s = \mu + \delta_s + \epsilon_s; \qquad \delta_s \sim \text{iid } \mathsf{N}(\mu, \tau^2);\ \epsilon_s \sim \text{iid } \mathsf{N}(0, V_s).$$

Estimation of an Overall Effect: Random Effects Model

For any of the equivalent forms of the random effects model, the marginal distributions of the T_s are:

$$T_s \sim \text{indep } \mathsf{N}(\mu, \tau^2 + V_s).$$

Interpretation:

- The expected effect is μ.
- The manifested effect in study s is θ_s (or $\mu + \delta_s$).

While the expected effect μ is certainly of interest, so is the distribution of effects among studies – which would also be assumed to apply to additional studies from the same "population" of studies from which our data have come.

Note that:

- As for the fixed effect model, the V_s are assumed known.
- Now two parameters to estimate, μ and τ^2.

Estimation of an Overall Effect: Random Effects Model

The Original Approach

- A combination of moment-based estimation of τ^2 followed by weighted least squares estimation of μ.
- For τ^2,

$$\hat{\tau}^2 = \begin{cases} \frac{Q-(S-1)}{C} & Q > (S-1) \\ 0 & \text{otherwise} \end{cases}$$

where

$$C = \sum_{s=1}^{S} w_s - \frac{\sum_{s=1}^{S} w_s^2}{\sum_{s=1}^{S} w_s},$$

with $w_s = 1/V_s$ as before.

Estimation of an Overall Effect: Random Effects Model

The Original Approach

- A combination of moment-based estimation of τ^2 followed by weighted least squares estimation of μ.
- For μ,

$$\hat{\mu} = \frac{\sum_{s=1}^{S} w_s^* T_s}{\sum_{s=1}^{S} w_s^*} \quad \text{where} \quad w_s^* = \frac{1}{\hat{\tau}^2 + V_s}.$$

- An asymptotic variance for $\hat{\mu}$ is estimated as

$$\hat{V}(\hat{\mu}) = \frac{1}{\sum_{s=1}^{S} w_s^*}.$$

- Note that uncertainty in estimation of τ^2 not quantified.

Estimation of an Overall Effect: Random Effects Model

The Original Approach

- An approximate $(1-\alpha)100\%$ confidence interval is,

$$\left(\hat{\mu} - z_{1-\alpha/2}[\hat{V}(\hat{\mu})]^{1/2},\ \ \hat{\mu} + z_{1-\alpha/2}[\hat{V}(\hat{\mu})]^{1/2}\right)$$

- Although the formulas look analogous to those for the fixed effect model with V_s replaced by $\hat{\tau}^2 + V_s$, the justification is different.
- It is an asymptotic result, not a small-sample result as for the fixed effect model.
- This Original Approach Due To:
  DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**: 177-188.
- A number of other possible estimators of τ^2 have been proposed. See DerSimonian, R. and Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials* **28**: 105-114.

Estimation of an Overall Effect: Random Effects Model

Likelihood Estimation

- The log likelihood for the random effects model may be written as,

$$\ell(\mu, \tau^2) = -\frac{1}{2}\sum_{s=1}^{S} \log\left[2\pi(\tau^2 + V_s)\right] - \frac{1}{2}\sum_{s=1}^{S}\left[\frac{(t_s - \mu)^2}{\tau^2 + V_s}\right].$$

- Taking derivatives and setting equal to zero gives,

$$\hat{\mu}(\tau^2) = \frac{\sum_{s=1}^{S}\left(\frac{t_s}{\tau^2 + V_s}\right)}{\sum_{s=1}^{S}\left(\frac{1}{\tau^2 + V_s}\right)}. \tag{1}$$

$$\sum_{s=1}^{S}\frac{1}{\tau^2 + V_s} = \sum_{s=1}^{S}\left(\frac{t_s - \mu}{\tau^2 + V_s}\right)^2. \tag{2}$$

Estimation of an Overall Effect: Random Effects Model

Likelihood Estimation

To find maximum likelihood estimates:

1. Substitute (1) into (2), maximize in τ via one-dimensional search, such as equal interval search, to find $\hat{\tau}^2$.
2. Substitue $\hat{\tau}^2$ back into (1) and compute $\hat{\mu}(\hat{\tau}^2)$.

Taking second derivatives and negative expected values gives the information matrix,

$$I(\mu, \tau^2) = \begin{pmatrix} \sum w_s^* & 0 \\ 0 & \frac{1}{2}\sum (w_s^*)^2 \end{pmatrix}. \tag{3}$$

Inverting gives,

$$\hat{V}(\hat{\mu}) = \frac{1}{\sum_{s=1}^{S} w_s^*} \quad \text{and} \quad \hat{V}(\hat{\tau}^2) = \frac{2}{\sum_{s=1}^{S} (w_s^*)^2}.$$

$$w_s^* = \frac{1}{\hat{\tau}^2 + V_s}$$

Estimation of an Overall Effect: Random Effects Model

Likelihood Inference

- Based on asymptotic normality of maximum likelihood estimators, approximate $(1-\alpha/2)100\%$ confidence intervals are then,

$$\left(\hat{\mu} - z_{1-\alpha/2}[\hat{V}(\hat{\mu})]^{1/2}, \hat{\mu} + z_{1-\alpha/2}[\hat{V}(\hat{\mu})]^{1/2}\right),$$
$$\left(\hat{\tau}^2 - z_{1-\alpha/2}[\hat{V}(\hat{\tau}^2)]^{1/2}, \hat{\tau}^2 + z_{1-\alpha/2}[\hat{V}(\hat{\tau}^2)]^{1/2}\right).$$

- The random effects model is often used when Cochran's Q or the Inconsistency Index indicate considerable heterogeneity. The approximate confidence interval for τ^2 gives us another way to examine the question of heterogeneity, directly as part of model fitting.
- Estimates and intervals for μ and τ^2 should not exhaust our inference from the random effects model, although few researchers using meta analysis go beyond them (and do not make inference about τ^2).

Estmation of an Overall Effect: Random Effects Model

Likelihood Inference About Additional Studies

- Consider an additional study drawn from the same population as those observed. According to the first form of the random effects model given, the mechanism will manifest itself in such a study with a value θ_0, which is a value taken from,

$$\theta_0 \sim \text{iid } \mathrm{N}(\mu, \tau^2).$$

- Thus, we can consider probabilities such as $P(\theta_0 \in \mathcal{A})$ for $\mathcal{A} \subset \mathbb{R}$,

$$P(\theta_0 \in \mathcal{A}) = \frac{1}{(2\pi\tau^2)^{1/2}} \int_{t \in \mathcal{A}} \exp\left[-\frac{1}{2\tau^2}(t-\mu)^2\right] dt. \qquad (4)$$

- Without even worrying about formal decision-theoretic predication, we can make inference about such probabilities by substituting $\hat{\mu}$ and $\hat{\tau}^2$ into 4 and evaluating the integral.

Bayesian Estimation in Meta Analysis

Fixed Effect Model

- Data Model: $T_s \sim \mathrm{N}(\mu, V_s)$.
- Prior: $\pi(\mu)$. Normal Example: $\pi(\mu) = \mathrm{N}(M_0, V_0)$.
- Posterior: $p(\mu|\boldsymbol{t}) \propto \pi(\mu) f(\boldsymbol{t}|\mu)$
- For Normal Example: $p(\mu|\boldsymbol{t}) = \mathrm{N}(M_\mu, V_\mu)$.

$$M_\mu = \frac{\frac{M_0}{V_0} + \sum_{s=1}^{S} \frac{t_s}{V_s}}{\frac{1}{V_0} + \sum_{s=1}^{S} \frac{1}{V_s}},$$

$$V_\mu = \frac{1}{\frac{1}{V_0} + \sum_{s=1}^{S} \frac{1}{V_s}}.$$

Inference for Fixed Effect Model

- Inference consists of making probability statements on the basis of the posterior.
- A point estimate of μ might be posterior mean (or mode).
- A $(1-\alpha)100\%$ (central) credible interval has endpoints given by the $\alpha/2$ and $1-\alpha/2$ quantiles of the posterior. If these are z_1 and z_2 the interpretation is

$$P(z_1 < \mu < z_2) = 1 - \alpha,$$

  which is quite proper using epistemic probability.
- Algebra will show that, for $\hat{\mu}$ the mle, if $M_0 > \hat{\mu}$ then $M_\mu > \hat{mu}$ and if $M_0 < \hat{\mu}$ then $M_\mu < \hat{\mu}$ – this is shrinkage of the frequentist estimator toward the prior mean.

Bayesian Estimation in Meta Analysis

Random Effects Model

- Data Model: $T_s \sim$ indep $\mathsf{N}(\theta_s, V_s)$.
- Mixing Dist'n or Prior 1: $\theta_s \sim$ iid $\mathsf{N}(\mu, \tau^2)$.
- Prior or Prior 2: $\pi(\mu, \tau^2)$.
  Simple Example:

$$\begin{aligned} \pi(\mu, \tau^2) &= \pi_1(\mu)\, \pi_2(\tau^2), \\ \pi_1(\mu) &= \mathsf{N}(M_0, V_0). \\ \pi_2(\tau^2) &= \mathsf{IG}(\alpha, \beta). \end{aligned}$$

- Posterior: $p(\mu, \boldsymbol{\theta}, \tau^2) \propto f(\boldsymbol{t}|\boldsymbol{\theta})\, g(\boldsymbol{\theta}|\mu, \tau^2)\, \pi(\mu, \tau^2)$.
- Posterior approximated by Markov Chain Monte Carlo.

Bayesian Estimation in Meta Analysis

Random Effects Model

Posterior Approximation:

- Could try to integrate out $\boldsymbol{\theta}$,

$$m(\boldsymbol{t}|\mu, \tau^2) = \int f(\boldsymbol{t}|\boldsymbol{\theta})\, g(\boldsymbol{\theta}|\mu, \tau^2)\, d\boldsymbol{\theta}.$$

  Then simulate from posterior

$$p(\mu, \tau^2) \propto m(\boldsymbol{t}|\mu, \tau^2)\, \pi(\mu, \tau^2),$$

  most likely using a Metropolis-Hastings type algorithim.

- Could include $\boldsymbol{\theta}$ (dimension S) in full posterior

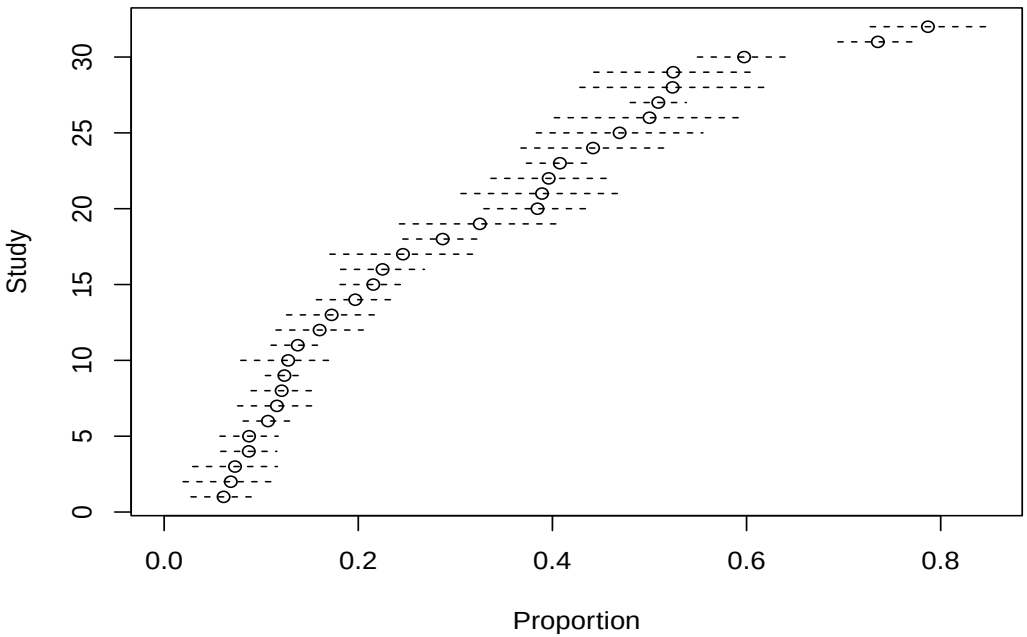
$$p(\mu, \boldsymbol{\theta}, \tau^2) \propto f(\boldsymbol{t}|\boldsymbol{\theta})\, g(\boldsymbol{\theta}|\mu, \tau^2)\, \pi(\mu, \tau^2),$$

  and then mostly likely use an overall Gibbs Sampling algorithm.

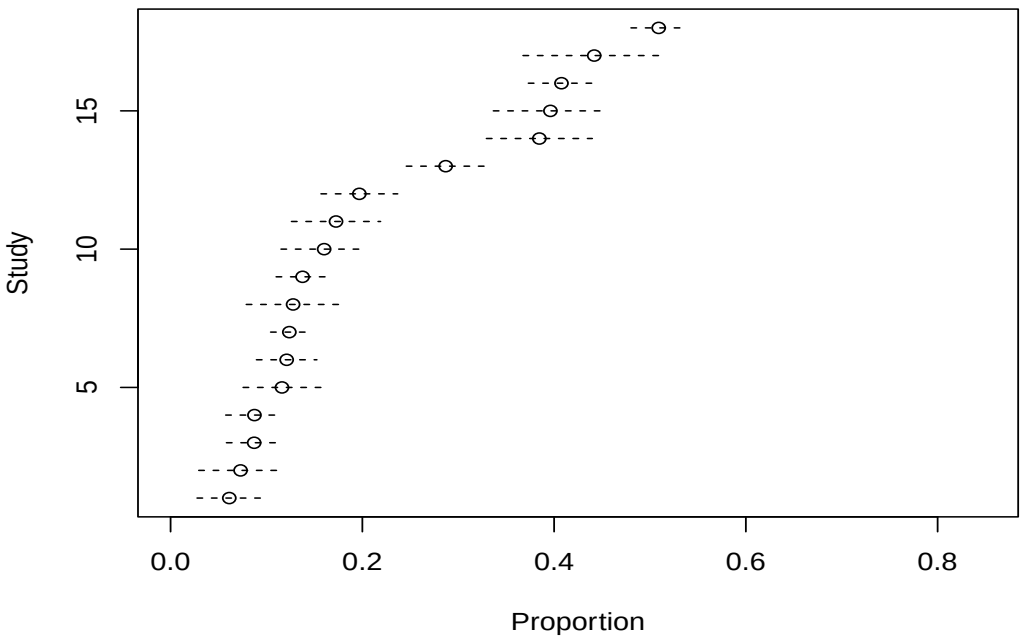
Meta Analysis Example

1. Meta analysis concerning drug-resistant epilepsy (DRE): Kalilani *et al.* (2018). The epidemiology of drug-resistant epilepsy: A systemic review and meta-analysis. *Epilepsia* **59 (12)**: 2179-2193.
2. A Number of Issues Examined:
   - Definition of DRE.
   - Region (US, Europe, other)
   - Association with Risk Factors.
3. Potential articles from search: $1,711$ unique.
4. Articles meeting inclusion criteria: 35 with estimates for 36 groups.
5. Some studies reported incidence, some prevalence (4 at pop level).
6. Reporting in this article is sloppy – numbers in tables do not add up to what is claimed in the text, etc.
7. For illustration, used 32 studies that reported prevalence or incidence of DRE among patients with epilepsy. Compared to a subset of 18 studies with values known to be incidence.

DRE: Forest Plots



- Full Data: 32 Studies



- Reduced Data: 18 Studies

Results: Fixed Effect Model

Estimation for Full Data Set

- $\hat{\mu} = 0.225$
- $\text{var}(\hat{\mu}) = 1.3 \times 10^{-5}$
- 95% confidence interval: (0.218, 0.232)

Analysis of Heterogeneity

- Cochran's Q: $2,594.57$
  Associated $p-$value: 0
- Inconsistency Index: 98.80
  Conclusion: substantial heterogeneity.

Several studies (references in Kalilani *et al.*) give prevalence of DRE among epilepsy patients from $30 - 40\%$.
This overall estimate considerably lower.

Results: Random Effects Model

DerSimonian/Laird Estimation for Full Data Set

- $\hat{\mu} = 0.298$
- $\hat{\tau}^2 = 0.0346$
- $\text{var}(\hat{\mu}) = 0.0011$
- 95% confidence interval: (0.233, 0.363)

Maximum Likelihood for Full Data Set

- $\hat{\mu} = 0.298$
- $\text{var}(\hat{\mu}) = 0.0013$
- 95% confidence interval for μ: (0.229, 0.368)
- $\hat{\tau}^2 = 0.0396$
- $\text{var}(\hat{\tau}^2) = 0.0001$
- 95% confidence interval for τ^2: (0.0197, 0.0595)

Results: Random Effects Model

Maximum Likelihood for Reduced Data Set

- $\hat{\mu} = 0.215$
- $\text{var}(\hat{\mu}) = 0.0012$
- 95% confidence interval for μ: (0.150, 0.280)
- $\hat{\tau}^2 = 0.0196$
- $\text{var}(\hat{\tau}^2) = 4.6 \times 10^{-5}$
- 95% confidence interval for τ^2: (0.0063, 0.0329)

1. DerSimonian/Laird and MLE similar for μ from Full Data, but likelihood interval wider. Not inconsistent with $30 - 40\%$ literature estimate but toward lower end.
2. MLE also provides interval for τ^2: completely above zero.
3. MLE for μ from Reduced Data smaller than estimate from Full. Upper CI limit does not even reach 30%.
4. MLE for τ^2 from Reduced Data also smaller than from Full, but still greater than zero.

Bayesian Analysis for Reduced Data

Prior Specification

- Recall Model:

$$T_s \sim \text{indep } \mathrm{N}(\theta_s, V_s) \quad \theta_s \sim \text{iid } \mathrm{N}(\mu, \tau^2)$$
$$\pi(\mu) = N(M_0, V_0) \quad \pi(\tau^2) = IG(\alpha, \beta).$$

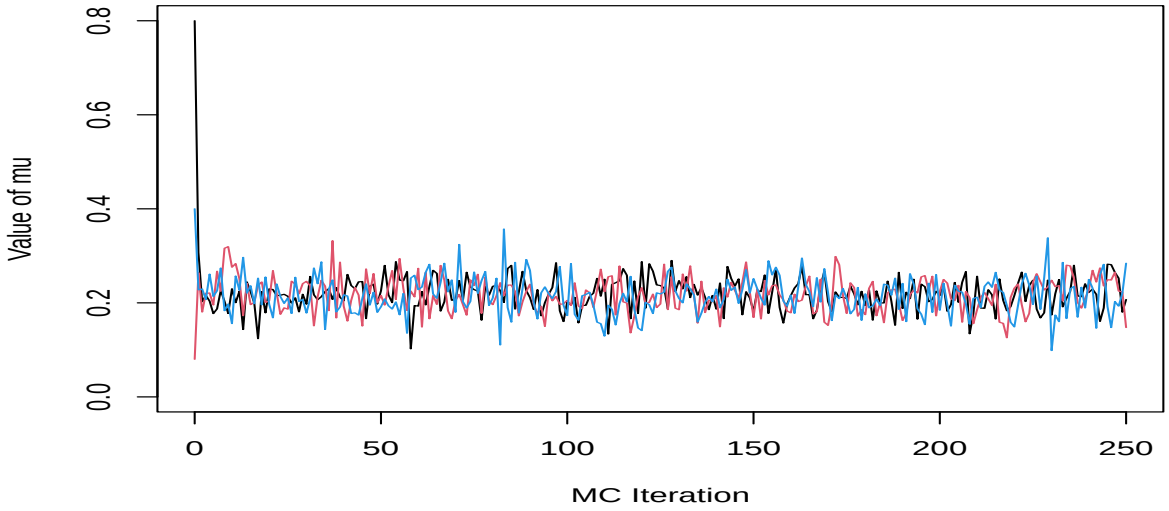
- Based on previous literature as cited in Kalilani *et al.* Set $M_0 = 0.35$.
- Want most prior probability between 0.1 and 0.8, and small probability less than 0. Set $V_0 = 0.1$ which gives,

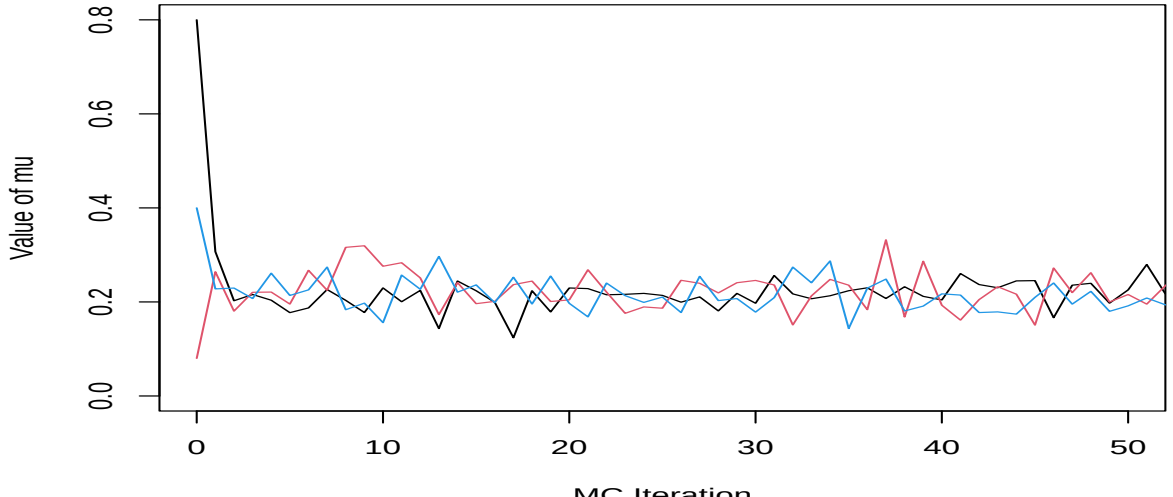
$$P(0.10 < \mu < 0.80) = 0.9938 \qquad P(\mu < 0) = 0.0002$$

- For reasonably diffuse, yet controlled, prior for τ^2 set $\alpha = 1$, $\beta = 0.01$. Gives, by Monte Carlo, a mean of about 0.10 and variance of about 10.

Markov Chain Details

- Rapidly mixing.
- Starting values (1) $\mu = 0.8$, $\tau^2 = 1$; (2) $\mu = 0.4$, $\tau^2 = 0.01$; (3) $\mu = 0.08$, $\tau^2 = 0.3$





Bayesian Analysis for Reduced Data

- Burn-In: 250
- MC Sample: $10,000$

Results for μ

Min	Q1	Q2	Mean	Q3	Max
0.076	0.192	0.216	0.217	0.242	0.383

95% credible interval: (0.144, 0.290)

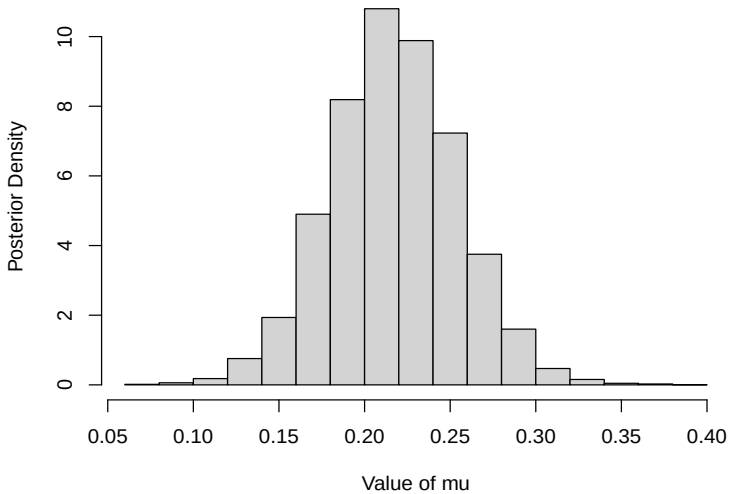
Results for τ^2

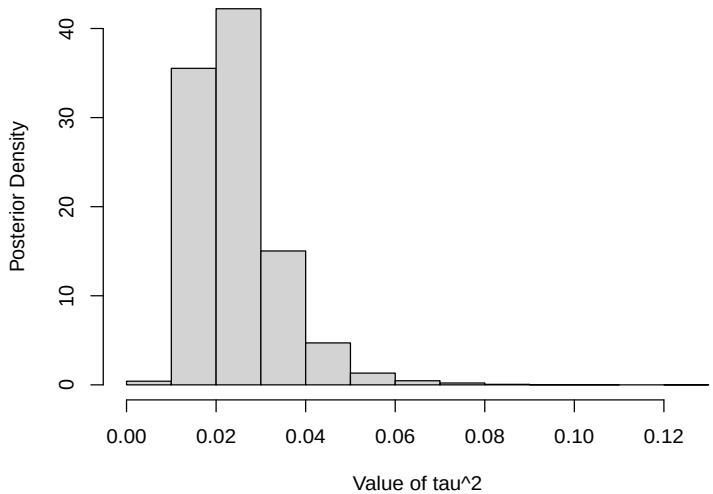
Min	Q1	Q2	Mean	Q3	Max
0.007	0.018	0.023	0.025	0.029	0.124

95% credible interval: (0.012, 0.049)

Bayesian Analysis for Reduced Data

- Results for μ similar to maximum likelihood, both point estimates (0.215 vs. 0.217) and intervals (0.150, 0.280) vs. (0.144, 0.290).
- MLE and posterior mean for τ^2 similar (0.020 vs. 0.023).
- Bayesian interval for τ^2 shifted to right relative to likelihood interval.





Reason for Difference in Intervals for τ^2:

- Likelihood interval based on asymptotic normality of maximum likelihood estimators.
- Posterior distribution of τ^2 definitively skew right.

Summary of Estimates – Random Effects Model

Point Estimates

Data	Method	Point Estimate μ	Point Estimate τ^2
Full	DerSimonian/Laird	0.298	0.035
	Maximum Likelihood	0.298	0.040
Reduced	Maximum Likelihood	0.215	0.020
	Bayes	0.217	0.025

Interval Estimates

Data	Method	95% Interval For μ	95% Interval For τ^2
Full	DerSimonian/Laird	$(0.233, 0.363)$	NA
	Maximum Likelihood	$(0.229, 0.368)$	$(0.018, 0.060)$
Reduced	Maximum Likelihood	$(0.150, 0.280)$	$(0.006, 0.033)$
	Bayes	$(0.144, 0.290)$	$(0.012, 0.049)$

Conclusions

About Drug-Resistant Epilepsy

1. Considerable heterogeneity among studies.
2. Results from Full Data toward lower end of $30 - 40\%$ range given in the literature.
3. Restricting studies to those that estimate only incidence gives smaller estimate than full data, and considerably smaller than $30 - 40\%$.

About Statistical Analysis

1. Likelihood estimation of random effects model provides more information than original DerSimonian/Laird approach.
2. Likelihood estimate of heterogeneity (τ^2) greater than for original method.
3. Bayesian analysis largely corroborates likelihood analysis, shifts interval for τ^2 to the right due to skew posterior. Superior for predictive analysis.

What About the Variances?

A potential major failing in all of the approaches to estimation presented is the assumption that variances of estimated effects from individual studies are known.

1. Typical modeling assumption is that If $T_s \sim$ indep $\mathrm{N}(\mu, V_s)$ or $T_s \sim$ indep $\mathrm{N}(\theta_s, V_s)$.
   - If responses are continuous, V_s may be the standard error of a mean or difference in means, perhaps with a pooled sample variance. Assuming T_s is an estimate of μ but the estimated standard error is V_s is quite questionable.
   - Even worse, when responses are binary as in the DRE example, the effect estimator (T_s) is $\hat{p}_s = y_s/n_s$ and $V_s = \hat{p}_s(1-\hat{p}_s)/n_s$. How can $\hat{p}_s$ be an estimator but V_s a fixed, known quantity?
2. One thought might be to model the V_s as well as the T_s. But The V_s are functions of the same fundamental responses (and sample sizes) as are the T_s. Let $\boldsymbol{y}_s$ be those responses in study s and let c_s denote any relevant constants (e.g., group sample sizes).
   - $T_s = T(\boldsymbol{y}_s, c_s)$.
   - $V_s = V(\boldsymbol{y}_s, c_s)$.

Incorporating Variances

Continuous Responses

- If T_s is a mean or a difference in group means, then

$$V_s = k_s W_s^2,$$

  where W_s^2 is a sample variance, either simple or pooled, and k_s is a known function of sample size(s), such as $k_s = 1/n$ or $k_s = (1/n_1 + 1/n_2)$.
- In this case, we could recover an value for W_s^2 from V_s as

$$W_s^2 = k_s^{-1} V_s.$$

- Make the additional (perhaps Herculean) assumption that individual responses are normally distributed, $Y_{s,i} \sim \text{indep } \mathrm{N}(\theta_s, \sigma_s^2)$.
- Then we have pairs (T_s, W_s^2); $s = 1, \ldots, S$ for which T_s and W_s are independent.

Incorporating Variances

Continuous Responses

Data Model:

$$T_s \sim \text{indep } \mathsf{N}(\theta_s, k_s\sigma_s^2),$$

$$W_s \sim \text{indep } \mathsf{Ga}\left(\frac{\nu}{2}, \frac{\nu}{2\sigma_s^2}\right),$$

where ν can be determined from k_s (e.g., $\nu = n - 1$ or $\nu = n_1 + n_2 - 2$). This comes from the sampling distribution of a sample variance S^2,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Mixing Distributions:

$$\theta_s \sim \text{iid } \mathsf{N}(\mu, \tau^2) \qquad \sigma_s^2 \sim \text{iid } \mathsf{IG}(\alpha, \beta).$$

Here, IG is conditionally conjugate for σ^2 in a normal one-sample setting.

Prior:

$$\pi(\mu, \tau^2, \alpha, \beta)$$

Incorporating Variances

Binary (Binomial) Responses

Here, the situation is simple because both mean (T_s) and squared standard error (V_s) are functions of a single parameter.

Data Model:

$$T_s \sim \text{Binom}(\theta_s, m_s); \quad m_s \text{ known for } s = 1, \ldots, S.$$

Mixing Distributions:

$$\theta_s \sim \text{iid Beta}(\alpha, \beta).$$

Prior:

$$\pi(\alpha, \beta).$$

The effect of binomial sample sizes m_s will be realized through the impact of the likelihood on the joint posterior:

$$p(\alpha, \beta, \{\theta_s : s = 1, \ldots, S\} | \boldsymbol{t}) \propto \left[\prod_{s=1}^{S} f(t_s | \theta_s, m_s)\, g(\theta_s | \alpha, \beta) \right] \pi(\alpha, \beta).$$

Some Additional Twists on Meta Analysis

Individual Meta Analysis

1. Combining results across a set of studies for which information is available for individual sampling units from each study.
   - Analysis is straightforward via mixed linear models.
   - Could alternatively use hierarchical models, linear or nonlinear.
2. Combining results across a set of studies in which individual level data are available for some studies but only aggregate data from others.
   - Modification of a mixed linear model in which responses are individual level for some studies and T_s for studies with only aggregate data.
   - Pure error variance in mixed model is taken as an unknown parameter for individual level responses and known V_s for aggregate responses.
   - The effects of this rather odd form for error variances has not been well examined.

Riley *et al.* (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* **27**: 1870-1893.

Some Additional Twists on Meta Analysis

Network Meta Analysis:
Comparing three or more treatments or interventions in a single analysis, say A, B, and C. Key is to combine what are called direct and indirect evidence.

Direct Comparisons

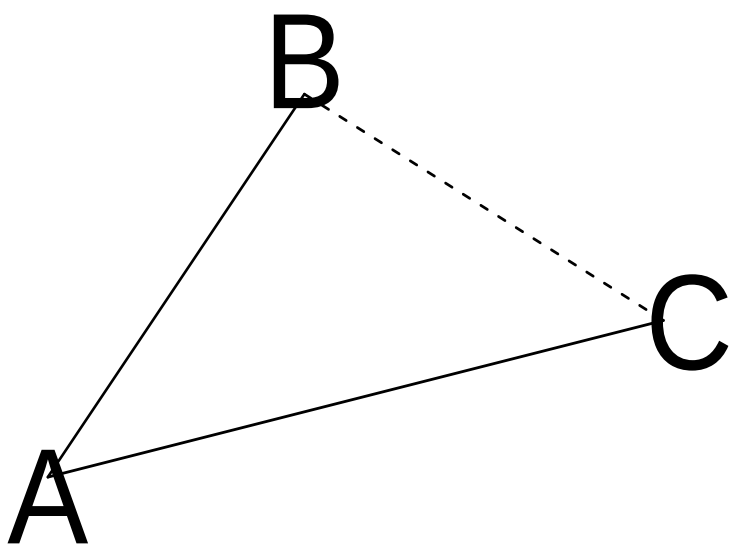
- Standard pairwise meta analysis.
- For our example, suppose that interventions A and B have been compared using difference in means, as have interventions A and C, each from sets of individual studies.

Indirect Comparisons

- Comparison between interventions which have not been used together in any individual study.
- For our example, suppose interventions B and C have never been included together in any individual studies.
- A comparison of B and C would be indirect.

Why Is it Called Network Meta Analysis?

- Consider three factors (interventions) to be compared.
- Depict interventions as nodes in a graph (or network).
- Edges join interventions that have been directly compared in one or more studies.



Studies With Direct Comparisons of Factors A–B and A–C
No Studies With Direct Comparison of Factors B–C

Assessing Indirect Evidence

1. Basic Idea: Let $MD(A, B)$ denote mean difference A minus B. Then

$$MD(B, C) = MD(A, C) - MD(A, B),$$
$$\text{var}[MD(B, C)] = \text{var}[MD(A, C)] + \text{var}[MD(A, B)].$$

   $(1 - \alpha)100\%$ Confidence Interval:

$$MD(B, C) \pm z_{1-\alpha/2} \left(\text{var}[MD(B, C)]\right)^{1/2}.$$

2. Supposes the realized effect of A is the same in trials containing A and B as in trials containing A and C.
   - May not be true if factors causing heterogeneity (sometimes called *modifiers*) differ between trials with A and C from those with A and B.
   - One approach compares distributions of modifiers among studies.

Chaimani, A. *et al.* (2019). Undertaking network meta-analyses. In Higgins, *et al.*, eds. *Cochrane Handbook for Systematic Reviews of Interventions*, version 6.5. Available from cochrane.org/handbook.

Combining Direct and Indirect Evidence

- Suppose now that there are some studies comparing B and C directly.
- Estimates can be obtained from an ANOVA-type model. Let T_s denote effect from study $s = 1, \ldots, S$ as before, for whatever comparison is made, $MD(A, C)$, $MD(A, B)$, or $MC(B, C)$.
- Let $\boldsymbol{x}_s = (x_{1,s}, x_{2,s})^T$ where

$$x_{1,s} = \begin{cases} 1 & \text{if } T_s = MD(A, C) \\ 0 & \text{if } T_s = MD(A, B) \\ 1 & \text{if } T_s = MD(B, C) \end{cases}$$

$$x_{2,s} = \begin{cases} 0 & \text{if } T_s = MD(A, C) \\ 1 & \text{if } T_s = MD(A, B) \\ -1 & \text{if } T_s = MD(B, C) \end{cases}$$

- Model(s):

$$T_s = \boldsymbol{x}_s^T \boldsymbol{\beta} + (V_s)^{1/2} \epsilon_s,$$

$$T_s = \boldsymbol{x}_s^T \boldsymbol{\beta} + \tau \delta_s + (V_s)^{1/2} \epsilon_s,$$

where both δ_s and ϵ_s are iid $N(0, 1)$.

Estimation for Network Regression Models

Fixed Effect Model

$$T_s = \boldsymbol{x}_s^T \boldsymbol{\beta} + (V_s)^{1/2} \epsilon_s; \quad \epsilon_s \sim \text{iid N}(0,1).$$

- Estimation using weighted least squares, $w_s = 1/V_s$.
- Inference from exact theory but assuming variances V_s known.

Random Effects Model

$$T_s = \boldsymbol{x}_s^T \boldsymbol{\beta} + \tau \delta_s + (V_s)^{1/2} \epsilon_s; \quad \epsilon_s, \delta_s \sim \text{iid N}(0,1).$$

- Estimation using iteratively re-weighted least squares as τ^2 unknown and $\text{var}(T_s) = \tau^2 + V_s$.
- Inference from asymptotic normality.
- Alternatively, Bayesian analysis.

Wrapping It Up

1. Meta analysis stems from medical applications and, to some extent, psychology, but is starting to show up in many other fields.
2. A principled approach to combining information from multiple sources or studies to produce a single estimate of some effect.
3. Does have some glaring deficiencies, the most obvious of which is the assumption that $V_s;\ s = 1, \ldots, S$ are known, not estimates.
   - It seems a "given" that this assumption is made in all aspects of meta analysis.
   - Fixed effect model, random effects model, model for combined individual and aggregate data, network-based models.
   - Original DerSimonian/Laird estimation, other least squares estimation (such as network models), likelihood estimation, Bayesian estimation.
4. Appears to be a lack of statistical effort devoted to the assumption of known V_s. To my knowledge:
   - No assessments of the impact on results.
   - No suggestions for methods to avoid the assumption, other than pure individual-level analysis.