



AGENCY FOR HEALTHCARE RESEARCH AND QUALITY



Quantitative Synthesis

Chapter 3. Choice of Statistical Model for Combining Studies

Prepared for:

The Agency for Healthcare Research and Quality (AHRQ)

Training Modules for Systematic Reviews Methods Guide

www.ahrq.gov

Learning objectives

1. Describe a scenario for which a fixed effects versus random effects model may be appropriate.
2. List different types of estimators for random effects models.
3. Describe the strengths and weaknesses for each of the estimators for different types of data.

Statistical models for meta-analysis

- Meta-analysis can be performed with either a **fixed effects model** or a **random effects model**.
 - ▶ A *fixed effects model* assumes there is 1 single treatment effect across studies.
 - Therefore differences between studies' treatment effects are due to random variability.
 - ▶ A *random effects model* assumes that the true treatment effect varies from study to study, following a random distribution.
 - Differences between studies owe not just to sampling error, the true treatment effect also varies (following a normal distribution).
- Validity of these assumptions are difficult to empirically verify.

Choosing a model

- Factors affecting model choice include:
 - ▶ Number and size of included studies
 - ▶ Type of outcome
 - ▶ Potential bias
- Do not choose a model based on the significance level of a heterogeneity test.
 - ▶ For example, recommend against selecting fixed effects model because $P > 0.1$
 - ▶ Such approaches do not factor in all relevant information that should inform model choice.

Choosing a model

- Heterogeneity is inevitable to some degree.
- Therefore, random effects models are generally recommended, with some special considerations (e.g., rare binary outcomes, discussed later).
- When a systematic review includes small and large studies, and results of small studies are different from large studies:
 - ▶ Suggests publication bias.
 - ▶ Assumption of a random distribution of effect sizes is likely violated.
 - ▶ In this case neither fixed nor random effects models yield valid results, and studies should not be combined.

Choosing a random effects estimator

- The *DerSimonian and Laird (DL) estimator* is the most common random effect model.¹
 - ▶ DL estimator does not accurately reflect error association with parameter estimation.²
 - ▶ Bias is most pronounced with few studies and/or high between-study heterogeneity.
- Refined estimators have been proposed:
 - ▶ The Hartung and Knapp (HK) estimator^{3,4}
 - ▶ The Sidik and Jonkman (SJ) estimator⁵
 - ▶ Jointly referred to as the **HKSJ method**
 - ▶ Both use the t distribution and adjust the confidence interval

1. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177-88.

[https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)

2. Brockhaus AC, Bender R, Skipka G. The Peto odds ratio viewed as a new effect measure. Stat Med. 2014;33(28):4861-74. <http://dx.doi.org/10.1002/sim.6301>

3. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. Stat Med 2001;20(12):1771-82. PMID: 11406840. <http://dx.doi.org/10.1002/sim.791>

4. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. Stat Med 2001;20(24):3875-89. PMID: 11782040

5. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. Stat Med. 2002;21(21):3153-9.

Profile likelihood (PL) methods

- There is a move to use alternative random effects estimators instead of DL.
- Simulation studies suggest that the ***Profile Likelihood (PL) methods*** generally performs best.⁶
 - ▶ They better account for uncertainty in estimating between-study variance.
 - ▶ PL methods have best performance across more scenarios than other methods.
 - ▶ However, PL methods may overestimate confidence intervals in small studies with low heterogeneity.
 - ▶ The PL method also does not always converge.
- The DL method is also appropriate when between-study heterogeneity is low.

Rare binary outcomes



- For rare binary events (e.g., adverse events), few or zero events may occur in one or both trial arms.^{7,8}
 - ▶ In this case the binomial distribution is not well-approximated by the normal distribution; model choice is complex.
 - ▶ The DL method performs poorly in this scenario.

7. Bradburn MJ, Deeks JJ, Berlin JA, et al. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007;26(1):53- 77. PMID: 16596572. <http://dx.doi.org/10.1002/sim.2528>

8. Shuster JJ, Walker MA. Low-event-rate meta-analyses of clinical trials: implementing good practices. *Stat Med*. 2016. <http://dx.doi.org/10.1002/sim.6844>

Rare binary outcomes

- The ***Peto method*** performs best when event prevalence is $<1\%$ (least bias, highest power, best confidence interval coverage).^{9,10}
 - ▶ Peto method performs poorly if studies are imbalanced or have large ORs (i.e., outside range of 0.2 – 5.0).
 - ▶ With imbalanced treatment arms, large effect sizes, or more frequent outcomes (5-10%), the Mantel-Haenszel method (without correction factor) or fixed effects logistic regression are preferred.

9. Fleiss J. The statistical basis of meta-analysis. Stat Methods Med Res. 1993;2(2):121-45. PMID: 8261254.

http://dx.doi.org/10.1177/096228029300200_202_72.

10. Vandermeer B, Bialy L, Hooton N, et al. Meta-analyses of safety data: a comparison of exact versus asymptotic methods. Stat Methods Med Res. 2009;18(4):421-32. http://dx.doi.org/10.1177/096228020809255_9

Rare binary outcomes

- **Beta-binomial models** have attractive properties for meta-analyses of rare binary outcome data.¹¹
 - ▶ This model assumes that outcomes of each trial follow a binomial distribution, and these binomial probabilities follow a beta distribution.
- Given the existence of newer methods handling rare/zero events, avoid using older continuity correction methods.
 - ▶ Instead, use valid methods that include studies with zero events in one or both arms.
 - ▶ However, no method yields completely unbiased results with sparse data; this issue is intractable to some degree.

11. Kuss O. Statistical methods for metaanalyses including information from studies without any events-add nothing to nothing and succeed nevertheless. Stat Med. 2015;34(7):1097-116.

<http://dx.doi.org/10.1002/sim.6383>

Rare binary outcomes

- Recommendations for combining rare binary outcome data:
 - ▶ Avoid continuity corrections.
 - ▶ For studies with zero events in one arm, or sparse binary events generally, consider the Peto method, the Mantel-Haenszel method, or logistic regression without correction factor (when heterogeneity is low).
 - ▶ If heterogeneity is high and/or studies exist with zero events in both arms, consider recently-developed methods (e.g., beta-binomial model).
 - ▶ Conduct sensitivity analyses acknowledging data adequacy.

Bayesian methods



- Although the frequentist framework is most common practice, meta-analysis can also be implemented in a Bayesian framework:
 - ▶ Accommodates a variety of outcome types.
 - ▶ Using GLM with normal, binomial, Poisson, or multinomial likelihoods and various link functions.¹²
- The Bayesian posterior parameter distributions fully incorporate uncertainty about all parameters.
 - ▶ Thus, Bayesian credible intervals tend to be wider than confidence intervals produced by classical random-effects models.
- Use vague priors, if Bayesian methods are chosen.¹³

12. Dias S, Sutton AJ, Ades AE, et al. Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials. *Med Decis Making*. 2013;33(5):607-17.

<http://dx.doi.org/10.1177/0272989X12458724>

13. Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med*. 2005;24(15):2401-28. PMID: 16015676. <http://dx.doi.org/10.1002/sim.2112>

Summary of objective 1

Objective: Be able to describe a scenario for which a fixed effects versus random effects model may be appropriate.

Summary:

- **Fixed effects model** assumes there is a single treatment effect across studies, whereas a **random effects model** assumes that true treatment effect varies from study to study, and follows a random distribution.
- Factors affecting model choice include the number and size of included studies, and type of outcome, among others.
- Don't choose a model based on the significance level of a heterogeneity test.
- Heterogeneity is inevitable to some degree. Therefore, **random effects models are generally recommended**, with some special considerations.

Summary of objective 2

Objective: Be able to list different types of estimators for random effects models.

Summary:

- ***DerSimonian and Laird (DL) estimator*** is the most common random effect model.
- The Hartung and Knapp (HK) estimator and the Sidik and Jonkman (SJ) estimator, which are referred to as the ***HKSJ method***.
- As an alternative to the DL estimator, ***Profile Likelihood (PL) methods*** generally perform best.
- For rare binary outcomes, the ***Peto method*** or ***beta-binomial models***.

Summary of objective 3

Objective: Be able to describe the strengths and weaknesses for each of the estimators for different types of data.

Summary:

- DL estimator
 - ▶ Appropriate when between-study heterogeneity is low
 - ▶ Does not accurately reflect error association with parameter estimation
 - ▶ Bias most pronounced with few studies and/or high between-study heterogeneity
 - ▶ Performs poorly for rare binary events (e.g., adverse events)
- PL methods:
 - ▶ Better account for uncertainty in estimating between-study variance
 - ▶ Best performance across more scenarios
 - ▶ May overestimate confidence intervals in small studies with low heterogeneity
- Peto method
 - ▶ Best when event prevalence is $<1\%$ (least bias, highest power, best confidence interval coverage)
 - ▶ Performs poorly if studies are imbalanced or have large odds ratios (i.e., outside range of 0.2 – 5.0)

Recommendations for Chapter 3. Statistical models for meta-analysis

- PL method appears to generally perform best. The DL method is also appropriate when between-study heterogeneity is low.
- **For study-level aggregated binary data and count data**, use of a generalized linear mixed effects model assuming random treatment effects is also recommended.

Recommendations for Chapter 3. Statistical models for meta-analysis

- **For rare binary events:**
 - ▶ Avoid methods that use continuity corrections.
 - ▶ For studies with zero events in one arm, or studies with sparse binary data but no zero events, obtain an estimate using the Peto method, the Mantel-Haenszel method, or a logistic regression approach, without adding a correction factor, when the between-study heterogeneity is low.
 - ▶ When the between-study heterogeneity is high, and/or there are studies with zero events in both arms, more recently developed methods such as a beta-binomial model could be explored and used.
 - ▶ Conduct sensitivity analyses with acknowledgement of the inadequacy of data.
- If choosing Bayesian methods, use of vague priors is supported.

Author



- This presentation was prepared by Jonathan Snowden, Ph.D.
- The presentation is based on the chapter entitled “Choice of Statistical Model for Combining Studies” in the Methods Guide for Comparative Effectiveness Reviews (available at: <https://doi.org/10.23970/AHRQEPCMETHGUIDE3>)

References



1. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-88.
[https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
2. Brockhaus AC, Bender R, Skipka G. The Peto odds ratio viewed as a new effect measure. *Stat Med*. 2014;33(28):4861-74. <http://dx.doi.org/10.1002/sim.6301>
3. Hartung J, Knapp G. On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Stat Med* 2001;20(12):1771-82. PMID: 11406840. <http://dx.doi.org/10.1002/sim.79164>
4. Hartung J, Knapp G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat Med* 2001;20(24):3875-89. PMID: 11782040
5. Sidik K, Jonkman JN. A simple confidence interval for meta-analysis. *Stat Med*. 2002;21(21):3153-9.
6. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001;20(6):825-40. PMID: 11252006. <http://dx.doi.org/10.1002/sim.650>
7. Bradburn MJ, Deeks JJ, Berlin JA, et al. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007;26(1):53-77. PMID: 16596572. <http://dx.doi.org/10.1002/sim.2528>
8. Shuster JJ, Walker MA. Low-event-rate meta-analyses of clinical trials: implementing good practices. *Stat Med*. 2016. <http://dx.doi.org/10.1002/sim.6844>
9. Fleiss J. The statistical basis of meta-analysis. *Stat Methods Med Res*. 1993;2(2):121-45. PMID: 8261254. <http://dx.doi.org/10.1177/09622802930020020272>.
10. Vandermeer B, Bialy L, Hooton N, et al. Meta-analyses of safety data: a comparison of exact versus asymptotic methods. *Stat Methods Med Res*. 2009;18(4):421-32. <http://dx.doi.org/10.1177/0962280208092559>
11. Kuss O. Statistical methods for metaanalyses including information from studies without any events-add nothing to nothing and succeed nevertheless. *Stat Med*. 2015;34(7):1097-116. <http://dx.doi.org/10.1002/sim.6383>
12. Dias S, Sutton AJ, Ades AE, et al. Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials. *Med Decis Making*. 2013;33(5):607-17. <http://dx.doi.org/10.1177/0272989X12458724>
13. Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med*. 2005;24(15):2401-28. PMID: 16015676. <http://dx.doi.org/10.1002/sim.2112>