

heterogeneity in meta-analysis

stats methodologists meeting

3 September 2015

basic framework

- k experiments ($i=1\dots k$)
- each gives an estimate x_i of θ , with variance s_i^2
- 'best' overall estimate is weighted (by $w_i=1/s_i^2$) mean of the x_i ...
- ...*if* the x_i are really all estimating the same θ

'The Combination of Estimates from Different Experiments' by William G. Cochran

Biometrics, Vol. 10, No. 1 (1954), pp. 101-129,

URL: <http://www.jstor.org/stable/3001666>

test of interactions on the *weighted* sum of squares of deviations

$$Q = \sum_{i=1}^k w_i (x_i - \bar{x}_w)^2,$$

where

$$w_i = 1/s_i^2$$

$$\bar{x}_w = \sum w_i x_i / \sum w_i$$

If the degrees of freedom n_i are large, Q follows the χ^2 distribution

‘Quantifying heterogeneity in a meta-analysis’

by J. P. T. Higgins & S. G. Thompson

Statist. Med. 2002; **21**:1539–1558

A test of homogeneity of the θ_i is provided by referring the statistic

$$\hat{\mu}_F = \frac{\sum w_i y_i}{\sum w_i} \quad Q = \sum w_i (y_i - \hat{\mu}_F)^2$$

to a χ^2 distribution with $k - 1$ degrees of freedom.

‘Quantifying heterogeneity in a meta-analysis’

by J. P. T. Higgins & S. G. Thompson

Statist. Med. 2002; **21**:1539–1558

A test of homogeneity of the θ_i is provided by referring the statistic

$$\hat{\mu}_F = \frac{\sum w_i y_i}{\sum w_i} \quad Q = \sum w_i (y_i - \hat{\mu}_F)^2$$

to a χ^2 distribution with $k - 1$ degrees of freedom.

$$H^2 = \frac{Q}{k - 1}$$

as a possible measure of the amount of heterogeneity

Assessing uncertainty in physical constants

Max Henrion and Baruch Fischhoff

Am. J. Phys. 54, 791 (1986);

pose each experiment i , for $i = 1, 2, \dots, N$, reports an estimate x_i with standard error σ_i , and \bar{x} is the group mean, weighted inversely by the variances. The sum of the squares of the normalized residuals, $h_i = (x_i - \bar{x})/\sigma_i$, should be distributed as the chi-squared statistic with $(N - 1)$ degrees of freedom, assuming the errors are independent and normally distributed with the reported standard deviations. This statistic can then be used to test the appropriateness of these assumptions.

Assessing uncertainty in physical constants

Max Henrion and Baruch Fischhoff

Am. J. Phys. 54, 791 (1986);

A related measure, the Birge ratio,¹⁸ R_B , assesses the compatibility of a set of measurements by comparing the variability among experiments to the reported uncertainties. It may be defined as the standard deviation of the normalized residuals:

$$R_B^2 = \sum_i h_i^2 / (N - 1). \quad (1)$$

Assessing uncertainty in physical constants

Max Henrion and Baruch Fischhoff

Am. J. Phys. 54, 791 (1986);

A related measure, the Birge ratio,¹⁸ R_B , assesses the compatibility of a set of measurements by comparing the variability among experiments to the reported uncertainties. It may be defined as the standard deviation of the normalized residuals:

$$R_B^2 = \sum_i h_i^2 / (N - 1). \quad (1)$$

¹⁸R. T. Birge, Phys. Rev. **40**, 207 (1932)

if there *is* heterogeneity (interaction, overdispersion...)

- move from fixed effects model:

$$\hat{\mu}_F = \sum w_i y_i / \sum w_i$$

- to 'random effects'

$$\hat{\mu}_R = \sum w_i^* y_i / \sum w_i^*$$

$$w_i^* = (w_i^{-1} + \hat{\tau}^2)^{-1}$$

$$\hat{\tau}_{DL}^2 = \frac{Q - (k - 1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}$$

what is the problem with an additional additive component of variance in random effects M-A?

$$\bar{w}_i^* = (\bar{w}_i^{-1} + \hat{\tau}^2)^{-1}$$

- no matter how big a trial is, its (statistical) weight is limited by τ ...
- ...which is particularly a problem if you are planning/reporting the next one

'Trial sequential methods for meta analysis'

E Kulinskaya & J Wood

RESEARCH SYNTHESIS METHODS 5(3):212-220 Sep 2014

- MRC requires a comprehensive review of existing evidence before funding trials
- guidelines of (e.g.) JAMA and the Lancet state that reports of new clinical trials must explain how they affect previous research with direct reference to existing meta-analysis
- which suggests performing power calculations for sample size for the 'next trial'...
- ...however, for large estimated or assumed values of τ^2 , several extra trials are needed, regardless of how large each may be

'EXPLAINING HETEROGENEITY IN META-ANALYSIS: A COMPARISON OF METHODS'

S.G. THOMPSON & S.J. SHARP

Statist. Med. 18, 2693-2708 (1999)

- incorporating residual heterogeneity into a meta-regression model:
 - allow a multiplicative factor:

$$y_i \sim N(\alpha + \beta x_i, \phi v_i)$$

- include an additive between-study variance component τ^2 :

$$y_i \sim N(\alpha + \beta x_i, v_i + \tau^2).$$

THOMPSON & SHARP (cont)

- The rationale for using a multiplicative factor for variance inflation is weak... has little intuitive appeal... leads to the same dominance of large studies over small studies that has been criticized in the context of “fixed-effect” meta-analysis... we do not recommended them.
- The use of an additive component of variance to represent heterogeneity between studies is more intuitively appealing, and of course is the usual way of representing heterogeneity in meta-analysis...

‘Confidence intervals for random effects meta-analysis...’ by M. Henmi & J.B. Copas

Stat Med. 2010 Dec 20;29(29):2969-83

- ...fixed effects estimates are less sensitive to (publication) biases than random effects estimates, since they put relatively more weight on the larger studies and relatively less weight on the smaller studies.
- Whereas the DerSimonian-Laird interval is centred on a random effects estimate, we centre our confidence interval on a fixed effects estimate, but allow for heterogeneity by including an assessment of the extra uncertainty induced by the random effects setting

'The Combination of Estimates from Different Experiments' by William G. Cochran

Biometrics, Vol. 10, No. 1 (1954), pp. 101-129,

URL: <http://www.jstor.org/stable/3001666>

- “This paper discusses methods for combining a number of estimates x_i of some quantity θ , made in different experiments...
- “It is important to find out whether the x_i agree with one another within the limits of their experimental errors...
- “...If they do not, the type of overall mean that will be useful for future action requires careful consideration.”

[supporting info.]

full quote from Cochran 1954

This paper discusses methods for combining a number of estimates x_i of some quantity θ , made in different experiments. For the i th estimate we have an unbiased estimate s^2 of its variance, based on n_i degrees of freedom. It is important to find out whether the x_i agree with one another within the limits of their experimental errors. If they do not, i.e. if interactions are present, the type of overall mean that will be useful for future action requires careful consideration. However, in most cases the problem of estimating the mean of the x_i , at least over some subgroup of the experiments, will remain.

full quote from THOMPSON & SHARP (1999)

The rationale for using a multiplicative factor for variance inflation is weak. The idea that the variance of the estimated effect within each study should be multiplied by some constant has little intuitive appeal, and leads to the same dominance of large studies over small studies that has been criticized in the context of “fixed-effect” meta-analysis. Thus, despite the fact that such analyses are easy to carry out, and might therefore be used as a quick and approximate way of assessing the impact of residual heterogeneity on the results, we do not recommend them in practice. The use of an additive component of variance to represent heterogeneity between studies is more intuitively appealing, and of course is the usual way of representing heterogeneity in meta-analysis without covariates as well as in many other situations

[full quote from Henmi & Copas \(2010\)](#)

[Stat Med.](#) 2010 Dec 20;29(29):2969-83. doi: 10.1002/sim.4029. Epub 2010 Oct 20.

Confidence intervals for random effects meta-analysis and robustness to publication bias.

Abstract

The DerSimonian-Laird confidence interval for the average treatment effect in meta-analysis is widely used in practice when there is heterogeneity between studies. However, it is well known that its coverage probability (the probability that the interval actually includes the true value) can be substantially below the target level of 95 per cent. It can also be very sensitive to publication bias. In this paper, we propose a new confidence interval that has better coverage than the DerSimonian-Laird method, and that is less sensitive to publication bias. The key idea is to note that fixed effects estimates are less sensitive to such biases than random effects estimates, since they put relatively more weight on the larger studies and relatively less weight on the smaller studies. Whereas the DerSimonian-Laird interval is centred on a random effects estimate, we centre our confidence interval on a fixed effects estimate, but allow for heterogeneity by including an assessment of the extra uncertainty induced by the random effects setting. Properties of the resulting confidence interval are studied by simulation and compared with other random effects confidence intervals that have been proposed in the literature. An example is briefly discussed.