

O'Rourke K (2006). A historical perspective on meta-analysis: dealing quantitatively with varying study results.



© Keith O'Rourke, Institute of Statistics and Decision Sciences, Duke University, 214 Old Chemistry Bldg. Durham, NC 27708, USA. Email: k_orourke@rogers.com

From gambling to astronomy

It was not until the 17th century, when the French mathematician [Blaise Pascal](#) developed mathematical ways of dealing with the games of chance used for gambling, that a science for dealing quantitatively with varying observations started to emerge (Franklin 2001). Whereas in games of chance these mathematical approaches allowed one to determine the value of possible gambles, it turned out they also allowed one to determine the best way to compare and combine observations made by different astronomers.

In the 1700s, there was not yet the strong and clear distinction made today between observations within a given study, and summarized results from different studies. These ideas were tackled in the 18th and 19th centuries by astronomers and mathematicians (Stigler 1986), such as [Gauss](#) and [Laplace](#) (Laplace 1820), and presented in a textbook ([Airy 1861](#)) published by [George Biddell Airy](#), the British Astronomer Royal. But it was only in the 20th century that statisticians addressed similar questions for the combination of clinical trial results. Summarizing results from different studies eventually became the formalized technique we refer to today as meta-analysis.

Karl Pearson and typhoid inoculation

The British statistician [Karl Pearson](#) was familiar with Airy's textbook and appears to have been the first to apply methods to combine observations from different clinical studies. He was asked to analyse data comparing infection and mortality among soldiers who had volunteered for inoculation against typhoid fever in various places across the British Empire with that of other soldiers who had not volunteered ([Pearson 1904](#)).

Pearson first re-grouped the study observations into larger groups, noting simply that he considered some groups too small. His reasoning here is not clear, though it might simply have been based on expediency, given the practical difficulty of carrying out many small analyses. This preliminary re-grouping of various studies into 'one study' would be considered an invalid technique today, although a re-analysis comparing the original studies with the collapsed studies used by Pearson shows that the collapsing had no practical consequence.

Pearson decided to look at the association of inoculation with infection separately from the association of inoculation with mortality. The observed study outcomes were presented in '2 by 2' tables in his Appendix B. He presented the results of his analyses in a table in which each study was assigned its own line showing its measure of effect, together with a measure of the within-study uncertainty. The last line gives a pooled estimate of the effect – his 'meta-analysis' – albeit without an estimate of the pooled uncertainty associated with this estimate.

There was less infection and death from typhoid in the inoculated groups than in the uninoculated groups, and, by the standards of the time (using two probable errors rather than two standard errors as the criterion), this difference was statistically significant in nine of the eleven comparisons. But Pearson was struck by the irregularity of the associations. Seeking some explanation for these varying effects, he considered the possibility that the soldiers who had volunteered for inoculation against typhoid might have been at lower initial risk of developing the disease. He notes that these uncertainties might be resolved by further scrutiny of the results in hand, but, significantly, proposes "an experimental inquiry":

Assuming that the inoculation is not more than a temporary inconvenience, it would seem to be possible to call for volunteers... [and] only to inoculate every second volunteer... with a view to ascertaining whether any inoculation is likely to prove useful... In other words, the "experiment" might demonstrate that this first step to a reasonably effective prevention was not a false one.

Karl Pearson appears to have been the first to analyse clinical trial results using meta-analysis. He was especially

Cite as:

O'Rourke K (2006). A historical perspective on meta-analysis: dealing quantitatively with varying study results. JLL Bulletin: Commentaries on the history of treatment evaluation (www.jameslindlibrary.org). [Brief history]

thorough about questioning the consistency of individual trial results and equally keen to discover clues from this for better future research.

The fertile field of agricultural statistics

Like Pearson, the British statistician [Ronald Fisher](#) had studied statistics from Airy's textbook, and was comfortable addressing the combination of different study results. During the 1920s and 1930s, Fisher worked at the Agricultural Research Station in Rothamstead. In his 1935 textbook, he gives an example of the appropriate analysis of multiple studies in agriculture, identifying the likely and real concern that fertilizer effects will vary by year and location (Fisher 1935). There were numerous references to and discussions of the analysis of multiple studies in the last book that Fisher wrote (Fisher 1956), in which he encouraged scientists to summarize their research in such a way to make the comparison and combination of estimates almost automatic, and the same as if all the data were available. Fisher's influence on meta-analysis is hard to exaggerate. For instance, one of the earliest publications warning about preferential publication of studies based on statistical significance acknowledged Fisher as the person responsible for stimulating the research ([Sterling 1959](#)).

One of Fisher's colleagues, [William Cochran](#) extended Fisher's approach and provided a formal random effects framework for it more in line with the earlier approach by Airy (Cochran 1937). Cochran, together with [Frank Yates](#) (another colleague of Fisher's), soon afterwards applied this in practice to agricultural data (Yates and Cochran 1938). Cochran continued to work on methods for the analysis of multiple studies throughout his career. Indeed, the last sentence in his last paper commented on the difficulties in dealing with study effects that vary over time and location (Cochran 1980).

Cochran also applied the method in medical research in an assessment of the effects of vagotomy (a surgical operation for duodenal ulcers), which was reported in an influential book entitled [Costs, Risks and Benefits of Surgery](#) ([Cochran et al. 1977](#)). Like Karl Pearson before him ([Pearson 1904](#)), Cochran commented on the need for data from controlled trials:

We could have come across a number of comparisons that were well done but not randomized – the type sometimes called observational studies. ... I would have been interested in including the observational studies so as to learn whether they agreed with the randomized studies and if not, why not? But the medical members of our team had been too well brought up by statisticians, and refused to look at anything but randomized experiments.

Meta-analysis and fair tests of social, educational and medical interventions

By the middle of the 20th century, the sheer volume of research reports forced researchers to consider how to develop and apply methods to synthesize the results produced. In 1940, for example, quantitative synthesis was used in an analysis of the results of 60 years' research by psychologists on extrasensory perception ([Pratt et al. 1940](#)). Finding themselves swamped with studies and in need of methods to make sense of the barrage of findings (Chalmers et al. 2002), other American social scientists and statisticians began to develop and apply methods for quantitative synthesis of the results of separate but similar studies (Light and Smith 1971; [Smith and Glass 1977](#)). In 1976, one of them, Gene Glass, coined the term 'meta-analysis' to refer to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass 1976, p 3). Articles and textbooks about meta-analysis followed soon after (Rosenthal 1978; Cooper and Rosenthal 1980; [Glass et al. 1981](#); [Hunter et al. 1982](#); [Light and Pillemer 1984](#); [Hedges and Olkin 1985](#)).

Application of meta-analysis by medical researchers began a few years later ([Stjernswärd 1974](#); Chalmers et al. 1977; Cochran et al. 1977; [Chalmers 1979](#)). Particularly influential was the first randomized trial conducted by Peter Elwood, Archie Cochrane and their colleagues to assess whether aspirin reduced recurrences of heart attack ([Elwood et al. 1974](#)). The results were suggestive of a beneficial effect, but not statistically convincing, so, as additional trials were reported, Elwood and Cochrane assembled and synthesized their results using meta-analysis ([Elwood 2004](#)). This left little doubt that aspirin could reduce the risk of recurrence, and the results were published in 1980 in an anonymous *Lancet* editorial ([Lancet 1980](#)). The editorial had actually been written by the British medical statistician Richard Peto. Based on earlier work (Peto et al. 1976; 1977), Peto and his colleagues went on to provide a detailed example (using randomized trials of beta-blockade following heart attack) to encourage clinicians to review randomized trials systematically, and to combine estimates of the effects of treatments considered to be the same, based on informed clinical judgment (Yusuf et al. 1985). When treatment effects varied among studies, Peto argued for testing and estimating the (fixed) weighted average of the varying treatment effects (Peto 1987). He and his colleagues therefore rejected the Airy/Cochran tradition of considering the variation of treatment effect as being like a random variable. The latter approach was promoted to medical researchers by DerSimonian and Laird (1986), who also provided simple approximate formulas for Cochran's formal random effects model.

As had happened in the social sciences a few years earlier, these developments in clinical research led to expository papers (L'Abbé et al. 1987; [Sacks et al. 1987](#); [Jenicek 1989](#); O'Rourke and Detsky 1989), special journal issues

(*Statistics in Medicine* 1987;6) and books ([Jenicek 1987](#); [Pettiti 1994](#); [Chalmers and Altman 1995](#)) directed at clinical researchers and clinicians. These publications tended to emphasize the importance of assessing the quality of the studies being considered for meta-analysis to a greater extent than the early work in social sciences had done (see, for example, [Jenicek 1987](#)). They also emphasized the importance of the overall scientific process (or epidemiology) involved ([Jenicek 1989](#); O'Rourke and Detsky 1989). Some of the more challenging statistical aspects of modelling this process were outlined for clinical researchers by (Eddy 1989).

The importance of using systematic approaches to reducing biases in reviews of a body of evidence began to be distinguished as an issue separate from meta-analysis ([Mulrow 1987](#); [Oxman and Guyatt 1988](#)). This emphasis was manifested most explicitly in the late 1980s by the creation of global trialists' groups to conduct collaborative 'overviews' - meta-analyses based on individual patient data from their respective studies (Early Breast Cancer Trialists' Collaborative Group 1988; [Antiplatelet Trialists' Collaboration 1988](#)), as well as international collaboration to prepare meta-analyses of all the randomized trials in some medical fields ([Chalmers et al. 1989](#)).

By the early 1990s, terminology was becoming confusing, and Chalmers and Altman ([1995](#)) suggested that the term 'meta-analysis' should be restricted to the process of statistical synthesis considered in this commentary. This convention has now been adopted in some quarters. For example, the second edition of the *BMJ* publication *Systematic Reviews* is subtitled 'meta-analysis in context' ([Egger et al. 2001](#)), and the 4th edition of Last's *Dictionary of Epidemiology* (Last 2001) gives definitions as follows:

***Systematic Review:** The application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic. Meta-analysis may be, but is not necessarily, used as part of this process.*

***Meta-Analysis:** The statistical synthesis of the data from separate but similar, i.e. comparable studies, leading to a quantitative summary of the pooled results.*

Just as debates seem likely to continue about the statistical methods used for meta-analysis, so also will debates continue about terminology. What is certain, however, is that we will continue to have to deal quantitatively with varying study results.

This *James Lind Library* commentary has been republished in the *Journal of the Royal Society of Medicine* 2007;100:579-582.

Further reading

Chalmers I, Hedges LV, Cooper H (2002). A brief history of research synthesis. *Evaluation and the Health Professions* 25: 12-37.

Hunt M (1997). *How Science takes stock: Story of meta-analysis*. New York: Russell Sage Foundation.

Olkin I (1990). History and Goals. In: Wachter, Straf, eds. *The future of meta-analysis*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

O'Rourke K (2002). Meta-analytical themes in the history of statistics: 1700 to 1938. *Pakistan Journal of Statistics* 18:285-299.

References

Airy GB (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. London: Macmillan and Company.

Antiplatelet Trialists' Collaboration (1988). Secondary prevention of vascular disease by prolonged anti-platelet treatment. *BMJ* 296:320-331.

Chalmers I (1979). Randomized controlled trials of fetal monitoring 1973-1977. In: Thalhammer O, Baumgarten K, Pollak A, eds. *Perinatal Medicine*. Stuttgart: Georg Thieme: 260-265.

Chalmers I, Altman DG (1995). *Systematic Reviews*. London: BMJ Publications.

Chalmers I, Enkin M, Keirse MJNC (1989). *Effective care in pregnancy and childbirth*. Oxford: Oxford University Press.

Chalmers I, Hedges L, Cooper H (2002). A brief history of research synthesis. *Evaluation and the health professions*.

Chalmers TC, Matta RJ, Smith H, Kunzler A-M (1977). Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *New England Journal of Medicine* 297:1091-96.

Cochran WG (1937). Problems arising in the analysis of a series of similar experiments. *Journal of Royal Statistical Society Supplement* 4(1):102-118.

Cochran WG (1980). Summarizing the results of a series of experiments. 80-2, 21-33. Durham, NC, Proceedings of the 25th Conference on the Design of Experiments in Army Research Development and Testing, U.S. Army Research Office.

Cochran WG, Diaconis P, Donner AP, Hoaglin DC, O'Connor NE, Peterson OL, Rosenoer VM (1977). Experiments in surgical treatments of duodenal ulcer. In: Bunker JP, Barnes BA, Mosteller F, eds. *Costs, risks and benefits of surgery*. Oxford: Oxford University Press, p 176-197.

Cooper HM, Rosenthal R (1980). A comparison of statistical and traditional procedures for summarizing research. *Psychological Bulletin* 87:442-449.

DerSimonian R, Laird N (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* 7:177-188.

Early Breast Cancer Trialists' Collaborative Group (1988). Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. *New England Journal of Medicine* 319:1681-92.

Eddy DM (1989). The confidence profile method: a Bayesian method for assessing health technologies. *Operations Research* 37:210-228.

Egger M, Davey Smith G, Altman DG (2001). *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books.

Elwood P (2004). The first randomised trial of aspirin for heart attack and the advent of systematic overviews of trials. The James Lind Library (<http://www.jameslindlibrary.org/>).

Elwood PC, Cochrane AL, Burr ML, Sweetnam PM, Williams G, Welsby E, Hughes SJ, Renton R (1974). A randomised controlled trial of acetyl salicylic acid in the secondary prevention of mortality from myocardial infarction. *BMJ* 1,436-440.

Fisher RA (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Fisher RA (1956). *Statistical methods and scientific inference*. Edinburgh: Oliver and Boyd.

Franklin J (2001). *The Science of conjecture: Evidence and probability before Pascal*. Baltimore and London: The Johns Hopkins University Press.

Glass GV (1976). Primary, secondary and meta-analysis of research. *Educational Researcher* 10:3-8.

Glass GV, McGaw B, Smith ML (1981). *Meta-analysis in social research*. Newbury Park: Sage Publications.

Hedges LV, Olkin I (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

Hunter JE, Schmidt FL, Jackson GB (1982). *Meta-analysis: cumulating research findings across studies*. Beverly Hills, Ca: Sage Publications.

Jenicek M (1987). *Méta-analyse en médecine. Évaluation et synthèse de l'information clinique et épidémiologique*. St. Hyacinthe and Paris: EDISEM and Maloine Éditeurs.

Jenicek M (1989). Meta-analysis in medicine: where we are and where we want to go. *Journal of Clinical Epidemiology* 42:35-44.

L'Abbé KA, Detsky AS, O'Rourke K (1987). Meta-analysis in clinical research. *Annals of Internal Medicine* 107:224-232.

Lancet (1980). Aspirin after myocardial infarction. *Lancet* 1:1172-3.

Laplace P-S (1820). *Théorie analytique des probabilités*. Oeuvres complètes 7 (3rd edition). Paris: Courcier, p lxxvii.

Last JM (2001). *A dictionary of epidemiology*. 4th edition. Oxford: Oxford University Press.

Light RJ, Pillemer DB (1984). *Summing up*. Cambridge: Harvard University Press.

Light RJ, Smith PV (1971). Accumulating evidence: Procedures for resolving contradictions among research studies. *Harvard Educational Review* 41:429-471.

Mulrow CD (1987). The medical review article: state of the science. *Annals of Internal Medicine* 10:485-88.

O'Rourke K, Detsky AS (1989) Meta-analysis in Medical Research: strong encouragement for higher quality in individual research efforts. *Journal of Clinical Epidemiology* 42:1021-1024.

Oxman AD, Guyatt GH (1988). Guidelines for reading literature reviews. *Canadian Medical Association Journal* 138:697-703.

Pearson K (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal* 3:1243-1246.

Peto R (1987). Discussion. *Statistics in Medicine* 6:242.

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer* 34:585-612.

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *British Journal of Cancer* 35:1-39.

Pettiti DB (1994). *Meta-analysis, decision analysis, and cost-effectiveness analysis: Methods for quantitative synthesis in medicine*. New York: Oxford University Press.

Pratt JG, Rhine JB, Smith BM, Stuart CE, Greenwood JA (1940). Extra-sensory perception after sixty years: a critical appraisal of the research in extra-sensory perception. New York: Henry Holt.

Rosenthal R (1978). Combining results of independent studies. *Psychological Bulletin* 85:185-193.

Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC (1987). Meta-analyses of randomized controlled trials. *New England Journal of Medicine* 316:450-455.

Smith ML, Glass GV (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32:752-760.

Sterling TD (1959). Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *Journal of the American Statistical Association* 54:30-34.

Stigler SM (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Stjernswärd J (1974). Decreased survival related to irradiation postoperatively in early breast cancer. *Lancet* 304:1285-1286.

Yates F, Cochran WG (1938). The analysis of groups of experiments. *Journal of Agricultural Science* 28:556-580.

Yusuf S, Peto R, Lewis J, Collins R, Sleight P (1985). Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Disease* 27:335-371.