

An historical perspective on meta-analysis: dealing quantitatively with varying study results

Keith O'Rourke

J R Soc Med 2007;100:579–582

Additional material for this article is available from the James Lind Library website [www.jameslindlibrary.org], where it was previously published.

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. 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 century by astronomers and mathematicians such as Gauss and Laplace¹ and presented in a textbook published by George Biddell Airy,² 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.³

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 'two by two' 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.

By the standards of the time (using two probable errors rather than two standard errors as the criterion) all but two studies analysed by Pearson showed statistically significant associations of inoculation with infection and death from typhoid; but he 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 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 probable and real concern that fertilizer effects will vary by year and location.⁴ There were numerous references to and discussions of the analysis of multiple studies in the last book that Fisher wrote,⁵ 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 statistical significance acknowledged Fisher as the person responsible for stimulating the research.⁶

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.⁷ Cochran, together with Frank Yates (another colleague of Fisher's), soon afterwards applied this in practice to agricultural data.⁸ 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.⁹

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*.¹⁰ Like Karl Pearson before him,³ 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.¹¹ Finding themselves swamped with studies and in need of methods to make sense of the barrage of findings,¹² other American social scientists and statisticians began to develop and apply methods for quantitative synthesis of the results of separate but similar studies.^{13,14} 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.'¹⁵ Articles and textbooks about meta-analysis followed soon after.^{16–21}

Application of meta-analysis by medical researchers began a few years later.^{10,22–24} 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.²⁵ The results were suggestive of a beneficial effect but were not statistically convincing; therefore, as additional trials were reported, Elwood and Cochrane assembled and synthesized their results using meta-analysis.²⁶ This left little doubt that aspirin could reduce the risk of recurrence, and the results were published in 1980 in an anonymous *Lancet* editorial,²⁷ which had actually been written by the British medical statistician Richard Peto. Based on earlier work,^{28,29} 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.³⁰ When treatment effects varied among studies, Peto argued for testing and estimating the (fixed) weighted average of the varying treatment effects.³¹ 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,³² 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,^{33–36} special journal issues³⁷ and books^{38–40} 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.³⁸ They also emphasized the importance of the overall scientific process (or epidemiology) involved.^{35,36}

The importance of using systematic approaches to reducing bias in reviews of a body of evidence began to be distinguished as an issue separate from meta-analysis.^{41,42} 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,^{43,44} as well as international collaboration to prepare meta-analyses of all the randomized trials in some medical fields.⁴⁵

By the early 1990s, terminology was becoming confusing, and Chalmers and Altman⁴⁰ 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*,⁴⁶ and the 4th edition of Last's *Dictionary of Epidemiology*⁴⁷ 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.

Competing interests None declared.

REFERENCES

- Laplace P-S. *Théorie Analytique des Probabilités. Oeuvres Complètes* 7 (3rd edition). Paris: Courcier, 1820: lxxvii
- Airy GB. *On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations*. London: Macmillan and Company, 1861
- Pearson K. Report on certain enteric fever inoculation statistics. *BMJ* 1904;**3**:1243–6
- Fisher RA. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935
- Fisher RA. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd, 1956
- Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *J Am Stat Assoc* 1959;**54**:30–4
- Cochran WG. Problems arising in the analysis of a series of similar experiments. *J Roy Stat Soc* 1937;**4**(Suppl):102–18
- Yates F, Cochran WG. The analysis of groups of experiments. *J Agric Sci* 1938;**28**:556–80
- Cochran WG. *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, 1980
- Cochran WG, Diakonius P, Donner AP, et al. 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, 1977:176–97
- Pratt JG, Rhine JB, Smith BM, Stuart CE, Greenwood JA. *Extra-Sensory Perception after Sixty Years: A Critical Appraisal of the Research in Extra-Sensory Perception*. New York: Henry Holt, 1940
- Chalmers I, Hedges L, Cooper H. A brief history of research synthesis. *Evaluation and the Health Professions*. 2002
- Light RJ, Smith PV. Accumulating evidence: Procedures for resolving contradictions among research studies. *Harv Educ Rev* 1971;**41**:429–71
- Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977;**32**:752–60
- Glass GV. Primary, secondary and meta-analysis of research. *Educ Researcher* 1976;**10**:3–8
- Rosenthal R. Combining results of independent studies. *Psychol Bull* 1978;**85**:185–93
- Cooper HM, Rosenthal R. A comparison of statistical and traditional procedures for summarizing research. *Psychol Bull* 1980;**87**:442–9
- Glass GV, McGaw B, Smith ML. *Meta-Analysis in Social Research*. Newbury Park: Sage Publications, 1981
- Hunter JE, Schmidt FL, Jackson GB. *Meta-Analysis: Cumulating Research Findings Across Studies*. Beverly Hills, CA: Sage Publications, 1982
- Light RJ, Pillemer DB. *Summing Up*. Cambridge: Harvard University Press, 1984
- Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press, 1985
- Stjernswärd J. Decreased survival related to irradiation postoperatively in early breast cancer. *Lancet* 1974;**304**:1285–6
- Chalmers TC, Matta RJ, Smith H, Kunzler A-M. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *NEJM* 1977;**297**:1091–6
- Chalmers I. Randomized controlled trials of fetal monitoring 1973–1977. In: Thalhammer O, Baumgarten K, Pollak A, eds. *Perinatal Medicine*. Stuttgart: Georg Thieme, 1979:260–5
- Elwood PC, Cochrane AL, Burr ML, et al. A randomised controlled trial of acetyl salicylic acid in the secondary prevention of mortality from myocardial infarction. *BMJ* 1974;**1**:436–40
- Elwood P. The first randomised trial of aspirin for heart attack and the advent of systematic overviews of trials. *The James Lind Library* 2004: <http://www.jameslindlibrary.org/>
- Aspirin after myocardial infarction. *Lancet* 1980;**1**:1172–3
- Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;**34**:585–612
- Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br J Cancer* 1977;**35**:1–39
- Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985;**27**:335–71
- Peto R. Discussion. *Stat Med* 1987;**6**:242
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trial* 1986;**7**:177–88
- L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987;**107**:224–32
- Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *NEJM* 1987;**316**:450–5
- Jenicek M. Meta-analysis in medicine: where we are and where we want to go. *J Clin Epidemiol* 1989;**42**:35–44
- O'Rourke K, Detsky AS. Meta-analysis in Medical Research: strong encouragement for higher quality in individual research efforts. *J Clin Epidemiol* 1989;**42**:1021–4
- Special issue. *Stat Med* 1987;**6**:881–944

- 38 Jenicek M. *Méta-Analyse en Médecine. Évaluation et Synthèse de L'information Clinique et Épidémiologique*. St. Hyacinthe and Paris: EDISEM and Maloine Éditeurs, 1987
- 39 Pettiti DB. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*. New York: Oxford University Press, 1994
- 40 Chalmers I, Altman DG. *Systematic Reviews*. London: BMJ Publications, 1995
- 41 Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;**10**:485–8
- 42 Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;**138**:697–703
- 43 Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. *NEJM* 1988;**319**:1681–92
- 44 Antiplatelet Trialists' Collaboration. Secondary prevention of vascular disease by prolonged anti-platelet treatment. *BMJ* 1988;**296**:320–31
- 45 Chalmers I, Enkin M, Keirse MJNC. *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press, 1989
- 46 Egger M, Davey Smith G, Altman DG. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001
- 47 Last JM. *A Dictionary of Epidemiology*. 4th edition. Oxford: Oxford University Press, 2001

Further reading

- a Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Evaluation and the Health Professions* 2002;**25**:12–37
- b Franklin J. *The Science of Conjecture: Evidence and Probability before Pascal*. Baltimore and London: The Johns Hopkins University Press, 2001
- c Hunt M. *How Science Takes Stock: Story of Meta-Analysis*. New York: Russell Sage Foundation, 1997
- d Olkin I. History and Goals. In: Wachter KW, Straf ML, eds. *The Future of Meta-Analysis*. Cambridge, MA: The Belknap Press of Harvard University Press, 1990
- e O'Rourke K. Meta-analytical themes in the history of statistics: 1700 to 1938. *Pakistan J Stat* 2002;**18**:285–99
- f Stigler SM. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1986

RSM Membership – New Benefits



- Reciprocal Clubs worldwide – now including access to over 100 USA university clubs and over 20 others worldwide
- Wifi in 1 Wimpole Street and Chandos House so you can access the internet
- Remote access to 5 medical databases
- Remote access to a package of full-text e-journals from early 2008
- Webcasts of RSM Meetings and more regional meetings



To keep updated on the latest news email membership@rsm.ac.uk to receive our monthly ebulletin or visit www.rsm.ac.uk