REFEREED PAPER

USING FOREST PLOTS TO INTRODUCE META-ANALYSIS, INCLUDING SIMPLE MODERATOR ANALYSIS, EARLY IN STATISTICS EDUCATION

CUMMING, Geoff
Statistical Cognition Laboratory, School of Psychological Science, La Trobe University, Australia
g.cumming@latrobe.edu.au

ABSTRACT

Meta-analysis is the quantitative integration of empirical studies that address the same or similar issues. It is usually the best way to draw research-based conclusions that can guide evidence-based practice by professionals, and evidence-based decision making by public policy makers. Meta-analysis is so important that students should learn about it very early in their statistics education. The close links between meta-analysis and practical conclusions drawn from bodies of research mean that meta-analysis is a vital element in outreach from statistics education. I describe software that uses forest plots to make the basic ideas of meta-analysis accessible, and my experience using it with beginning students. I use the software to illustrate two major models for meta-analysis, and introduce graphical extensions to forest plots that illustrate how the crucial topic of moderator analysis can be explained and, in simple cases, interpreted visually.

INTRODUCTION

Meta-analysis is a set of techniques for the quantitative integration of results from a number of empirical studies on the same or similar issues. First, a meta-analysis is likely to give a more precise estimate of population parameters than any of the contributing studies. Second, it may allow identification of moderating variables that influence major outcomes. Meta-analysis is becoming widely used in numerous disciplines as the main way to review bodies of literature and identify what conclusions are justified by the evidence. *Evidence-based practice* is increasingly the expectation in medicine, psychology, and many other professions, and meta-analysis is almost always the best way to integrate research findings and thus provide the evidence on which evidence-based practice must rely. Similarly, *evidence-based decision making* should wherever possible guide formation of corporate and public policy and, again, meta-analysis is usually the best way to integrate results and justify the conclusions drawn.

At ICOTS 7, I argued (Cumming, 2006) that meta-analysis is so important it should be included in the introductory statistics course. I outlined how software based on the forest plot has allowed me to do that with introductory psychology students. Here I describe my further 5 years experience teaching meta-analysis, and extend my argument in two main ways. First, I expand the justification for making meta-analysis an important topic in statistics education by discussing it in the context of the conference theme of *Statistics Education and Outreach*. Second, I describe extensions to my ESCI (Exploratory Software for Confidence Intervals) software that allow me to expand teaching about meta-analysis to include both fixed effect and random effects models, and also to include the essential topic of moderator analysis. To illustrate outreach, I'll next give three examples of meta-analysis in action.

THREE EXAMPLES OF META-ANALYSIS IN ACTION

Medicine is the discipline that is most advanced in its use of meta-analysis. The Cochrane Collaboration (www.cochrane.org) is a large co-operative worldwide effort to create and make publicly available thousands of reviews, based on meta-analysis, of medical research. The conclusions of Cochrane reviews are usually the best basis for evidence-based practice in health care. Meta-analysis thus makes an vital contribution to keeping medical practice up-to-date, and therefore contributes greatly to human wellbeing. This single example justifies identifying meta-analysis as a vital component of statistical outreach. It also means that all health care workers need at least a basic understanding of meta-analysis if they are to appreciate the basis for the guidance that Cochrane gives for their professional development and their everyday practice. Their statistics education must therefore include an introduction to meta-analysis.

My first example is medical, and heart-wrenching. Cot death, or Sudden Infant Death Syndrome (SIDS), is the death of healthy babies while sleeping. Its cause has long been a mystery. My wife and I, like many young parents in the late 1970s and early 1980s, followed the best advice at the time by putting our three babies to sleep on their front on a sheepskin. However, Gilbert, Salanti, Harden, and See (2005) used meta-analysis to integrate evidence on how sleeping position influences the risk of SIDS. They also reported a historical review of the advice given by authors of childcare books, including Doctor Spock, on sleeping position for babies. They found that by 1970 there was reasonably clear evidence that back sleeping was safer than front, although there was little good research until around 1986, by which time the advantage of back sleeping was very clear. However, they also found frequent recommendation of front sleeping, even as late as 1988. They estimated that, if meta-analysis had been available, and used to integrate research carried out before 1970, and if, consequently, back sleeping had been widely recommended from 1970, then as many as 50,000 infant deaths could have been prevented in Europe, the U.S., and Australasia. Statistical techniques can save large numbers of lives.

My second example is from education. *Teaching to Read, Historically Considered* is an intriguing book by Mathews (1966), who described how fashions for teaching reading have changed since Greek slaves taught rich Romans to read. The pendulum of popularity has swung back and forth between largely phonic methods and largely whole-word methods. Phonic methods pay careful attention to the ways letters represent sounds, whereas whole-word methods emphasize complete words in context. Mathews reviewed the research literature and concluded that phonic methods are more effective than whole-word. However, not everyone agreed, and fads and outspoken individuals, rather than evidence, seemed to guide educational policy and teachers' practices. The U.S. National Reading Panel (2000) reported a large meta-analysis on the effects of different teaching methods. Diane McGuinness, in her book *Early Reading Instruction: What Science Really Tells us About How to Teach Reading* (McGuinness, 2004), extended and discussed the Panel's findings. She concluded that we now know how to teach reading successfully: It's essential to focus early on phonics. Mathews was right, back in 1966. This is an example of meta-analysis not merely summarizing evidence, but determining in decisive fashion a long-standing and heated controversy, and giving clear practical guidance on an issue of crucial importance to the whole English-speaking world. The pendulum should swing no more.

My third example illustrates how meta-analysis can guide public policy making. It comes from Hunter and Schmidt (2004), who described how the U.S. Congress, in advance of considering funding for producing binary chemical weapons, commissioned a meta-analysis of studies that had evaluated such weapons. Hunter and Schmidt reported that the meta-analysis did not support the production of the weapons. They stated that: "This was not what the Department of Defense (DOD) wanted to hear, and the DOD disputed the methodology and the results. The methodology held up under close scrutiny, however, and in the end Congress eliminated funds for these weapons" (p. 30).

STATISTICS EDUCATION AND OUTREACH: META-ANALYSIS

I conclude from those examples, and many others, that meta-analysis is highly relevant to the conference theme in at least the following ways:

Understanding of evidence-based practice. Students who may go on to train in any profession that expects evidence-based practice need a critical understanding of how research evidence is integrated, and communicated in an accessible way to busy practitioners. They therefore need to understand meta-analysis. Citizens need a basic understanding of what is meant by evidence-based practice, and therefore they, too, also need at least a basic appreciation of what meta-analysis is and can provide.

Understanding of evidence-based decision making. Similar comments apply here, for any students who may train for any management or decision-making role, and for citizens to understand how public decision making should be based.

Understanding media reports. The media increasingly refer to research reviews based on meta-analysis, whether or not the term is mentioned, as, for example "A recent review of 26 studies found no evidence that mobile phones cause cancer…" Informed citizens thus need to appreciate the basics of how research evidence from individual experiments can be integrated.

Planning research. Researchers routinely consider previous research when planning their experiments, and routinely summarise previous research in the introduction of a journal article. Since 2005 *The Lancet*, a leading medical journal, has included in its instructions to authors the requirement that "The relation between existing and new evidence should be shown by direct reference to an existing systematic review and meta-analysis; if neither exists, authors are encouraged to do their own" (tinyurl.com/lancetma). Introducing that requirement was a pioneering step, but now journal editors across a number of disciplines are increasingly asking authors of empirical articles to include in the introduction a reference to an existing meta-analysis, or to do their own. That's an excellent development: Without good understanding of past research, how can we know whether this new study is necessary, or appreciate what contribution it makes? Meta-analysis is thus becoming an essential part of research planning.

Meta-analytic thinking. More broadly, researchers should always be thinking of their own study in the context of past studies in the same area, and possible future studies. They should take particular care to report their results in sufficient detail and in a way that assists their inclusion in future meta-analyses. Cumming and Finch (2001) described that attitude to research as *meta-analytic thinking*. Students should learn about and be encouraged to adopt meta-analytic thinking.

Hunt (1997) is an excellent book about meta-analysis, aimed at the general reader, which explains the basics and includes numerous examples from a wide range of fields. Hunt makes the important claim that social science can thank meta-analysis for the continuation of public funding for social science research. He argued that even very large studies on complex social issues rarely give clear-cut findings or clear advice. The application of meta-analysis, however, to a number of studies can often give very useful guidance for policy and practice, as well as for future research, even in messy applied areas. Decision makers can thus see the value of funding such research.

In summary, meta-analysis is a vitally important element in the statistical literacy required to be an educated citizen, as well as an essential component of the statistical education needed by anyone training to be a researcher, professional practitioner, or public or corporate decision maker. Modern statistics education must include meta-analysis.

THE FOREST PLOT TO MAKE META-ANALYSIS VISUAL AND ACCESSIBLE

In Cumming (2006) I described a module of ESCI ("ESS-key", *Exploratory Software for Confidence Intervals*, http://www.latrobe.edu.au/psy/esci/) that pictures a simple meta-analysis as a forest plot. Figure 1 is an example. For some seven years I have used such pictures to introduce large classes of beginning psychology undergraduates to the basic ideas of meta-analysis. They see the software in lectures, then use it for their own exploration in tutorials and at home. I then revisit meta-analysis with the same students two or three years later. Often it is meta-analysis and forest plots that students spontaneously mention as the topic they remember clearly from first year. "It just made so much sense" is a typical comment, music to any teacher's ear. My experience since 2006 has reinforced my conviction that it is possible, even easy, to introduce beginning students to the basics of meta-analysis, and that forest plots are the key to success.

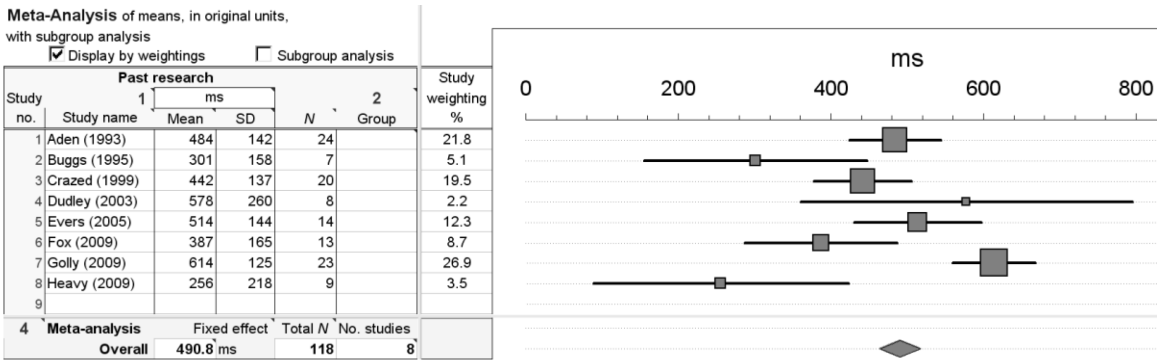

Figure 1. A fixed-effect meta-analysis. The figure shows an image from ESCI (*Exploratory Software for Confidence Intervals*, which runs under Microsoft Excel) of a meta-

analysis of fictitious data. At right is a forest plot, in which the main result of each study is represented as a square marking the mean, and a 95% confidence interval. Eight studies reported mean response times for a decision making task. The mean, SD, and *N* are entered for each study. The weightings used to combine the studies in the meta-analysis are reported as percentages and indicated by the relative sizes of the squares marking the means. The diamond represents the result of a fixed effect meta-analysis and its 95% confidence interval.

Forest plots are a wonderful way to assemble and present the major findings of several studies, or even a whole research literature. I use forest plots in ESCI to discuss estimation, the near irrelevance of statistical significance testing in meta-analysis, cumulative meta-analysis, the weighting of studies, publication bias and the file drawer effect, and the need for careful, critical thought when conducting a meta-analysis.

FIXED EFFECT AND RANDOM EFFECTS MODELS FOR META-ANALYSIS

The next steps are to consider the two major models for meta-analysis, and moderator analysis. Developments of ESCI allow forest plots to serve both these educational goals. The *fixed effect* model assumes that every study estimates a single fixed population mean, μ. This is the simplest model, but the assumption is very strong, and in many cases unrealistic. Figure 1 shows a fixed effect meta-analysis of the results of 8 fictitious studies that are quite disparate, as the lack of overlap among the confidence intervals attests. The very short diamond hardly seems a justifiable conclusion from the set of 8 studies. It gives a very precise estimate of what is true in the world, despite major disagreement among the individual studies.

Meta-Analysis of means, in original units, with subgroup analysis

☑ Display by weightings ☐ Subgroup analysis

Study no.	Study name	Mean (ms)	SD	*N*	Group (2)	Study weighting %
1	Aden (1993)	484	142	24		14.4
2	Buggs (1995)	301	158	7		11.5
3	Crazed (1999)	442	137	20		14.3
4	Dudley (2003)	578	260	8		8.5
5	Evers (2005)	514	144	14		13.6
6	Fox (2009)	387	165	13		12.9
7	Golly (2009)	614	125	23		14.6
8	Heavy (2009)	256	218	9		10.2
9						
4	**Meta-analysis**	Random effects		Total *N*	No. studies	
	Overall	452.13 ms		118	8	

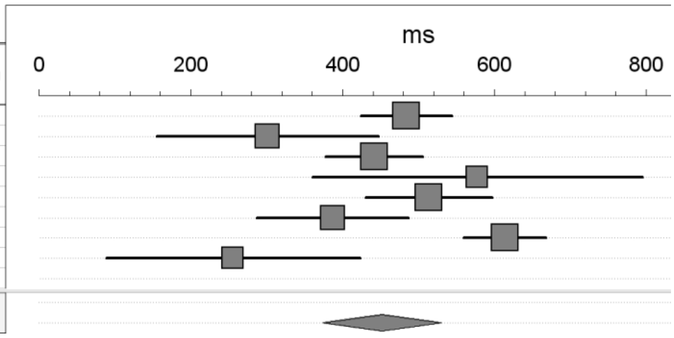

Figure 2. The same as Figure 1, except that the meta-analysis uses a random effects model. The study weightings vary much less from study to study and the confidence interval on the result, shown by the length of the diamond, is much wider.

The *random effects* model for meta-analysis assumes that each study estimates a different population mean, μ_i, sampled from a super-population with overall mean, μ. This model is more complex, and usually assumes the super-population is normally distributed and the μ_i are sampled randomly from that super-population. These are very strong assumptions, but even so a random effects meta-analysis is usually in practice a better choice than the fixed effect model. Figure 2 shows the random effects meta-analysis of the data shown in Figure 1. The much longer diamond seems a more justifiable conclusion to draw from the disparate studies.

A single mouse click switches between the models, so students can easily compare the two, and see, for example, how the two models give different patterns of study weightings as well as different overall results. In addition, it is easy to explore any other aspect of a set of studies, such as the effects of varying the extent to which the studies give similar or disparate results.

ESCI reports three statistics that measure the extent of *heterogeneity* between studies. Heterogeneity is study-to-study variability that is greater than can readily be accounted for by sampling variability. For the 8 studies in Figures 1 and 2, one of those statistics is I^2 = 86.5%, which is an estimate that 86.5% of the total study-to-study variability reflects the variability of the

μ_i in the super-population. In other words, there is very large heterogeneity, consistent with the lack of overlap of some of the confidence intervals in the forest plot.

ACCOUNTING FOR HETEROGENEITY: THE SEARCH FOR MODERATORS

The first large contribution of meta-analysis is to give an overall mean that is a more precise estimate than that given by any of the individual studies. That's the diamond at the bottom of a forest plot. The second major contribution can be an analysis of what variable, or variables, can account for heterogeneity. Such variables are known as *moderating variables*, or *moderators*. The simplest moderator is a dichotomous variable, which is a variable, such as gender, that can take just one of two categorical values. ESCI illustrates moderator analysis for the simplest case of a single dichotomous moderator.

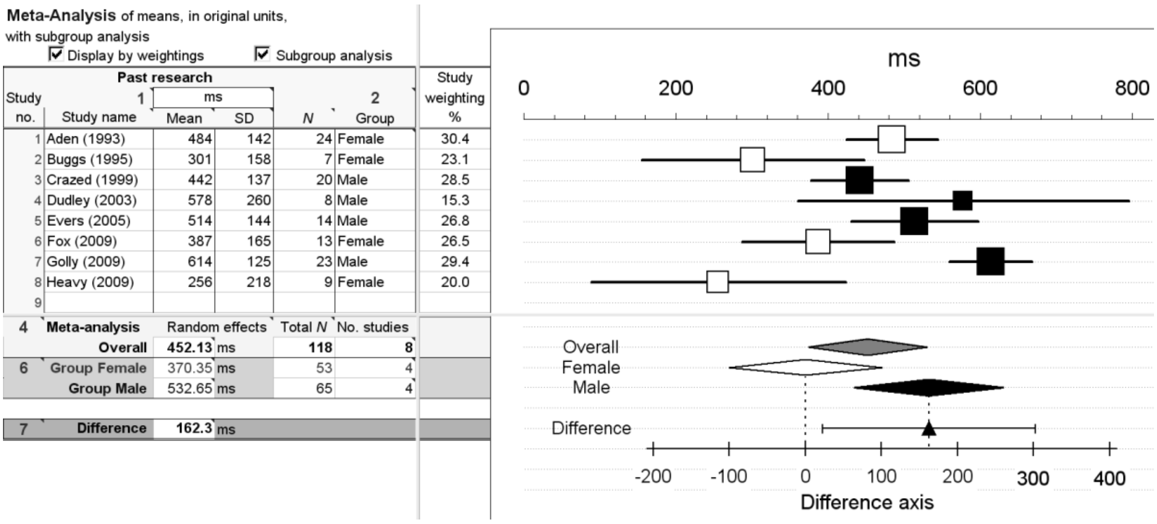

Figure 3. The same studies with the same data as Figures 1 and 2, but now each study is identified as employing either all male or all female participants. Random effects meta-analysis is applied first to all 8 studies, and the result shown as the grey diamond. Then random effects meta-analysis is applied separately to the two subgroups of studies, split by gender. The white squares and diamond refer to females, and the black to males. A difference axis is displayed at the bottom, its zero lined up with the overall female mean. The small black triangle marks on that difference axis the difference between the overall male and female means, with the 95% confidence interval on that difference.

I supposed that the 8 studies in Figures 1 and 2 employed either all female or all male participants. In Figure 3 the 'Group' column marks for each study the gender of the participants. We can therefore investigate gender as a possible moderator. Figure 3 shows an overall random effects meta-analysis (the grey diamond), then the same model applied to just the female subset of studies (white diamond) and the male subset (black diamond). The male minus female difference between the overall means is marked by the small black triangle on the difference axis at the bottom. The difference is 162 ms, 95% CI [22, 302], and the 95% confidence interval on that difference is displayed as the error bars on the small black triangle.

The configuration of diamonds, difference axis, and triangle with error bars is a novel way to display the result of a moderator analysis, but makes the important aspects of such an analysis visually clear, and provides a good basis for interpretation. The male and female diamonds overlap only a little, and the confidence interval on the difference, although wide, does not include zero. We can conclude that gender is a moderator of response time to the decision making task under study: Males are distinctly slower, on average, than females.

One of the great strengths of meta-analysis is that it can permit identification of moderators. It may even be able to assess the effect of a moderator when none of the individual studies is able to do that. In my set of 8 fictitious studies, for example, no single study compared males and females, but the moderator analysis shown in Figure 3 could do so. Moderator analysis can allow meta-analysis to yield theoretical and practical conclusions from a set of studies that go well beyond just weighted averages of study outcomes.

CONCLUSIONS

Researchers need to have meta-analysis in mind as they plan, analyse, and report research. Meta-analysis is the main tool we have to draw justified conclusions from a body of research and to formulate advice to support evidence-based practice by professionals, and evidence-based decision making by policy makers. Meta-analysis is a vital part of the outreach from statistics to society.

The American Psychological Association (APA) recently released the sixth edition of its *Publication Manual* (APA, 2010), which is highly influential because it is used by millions of students around the world, and more than 1,000 journals in a wide range of disciplines. This edition includes greatly increased coverage of meta-analysis and gives detailed advice on how to report meta-analyses. This is a further indication that meta-analysis is now mainstream.

Meta-analysis needs to be included in statistics education, and its close link with practical examples and practical research-based advice means that it is a vital part also of the outreach from statistics education to society. Forest plots provide a practical way to make meta-analysis readily understandable to students even at beginning levels. They can also help explain different models. I hope my development of the forest plot will prove useful in making moderator analysis also understandable to students. My book *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis* (Cumming, 2012) explains more about meta-analysis, including moderator analysis and how to make best use of ESCI, and provides examples.

ACKNOWLEDGEMENTS

I thank the Australian Research Council for support, and Fiona Fidler and Jerry Lai for comments on a draft.

REFERENCES

American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Cumming, G. (2012). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Cumming, G. (2006). *Meta-analysis: Pictures that explain how experimental findings can be integrated*. In Proceedings of ICOTS7, The 7th International Conference on Teaching Statistics. Salvador, Brazil, July. Retrieved from www.stat.auckland.ac.nz/~iase/publications/17/C105.pdf Also available from tinyurl.com/teachma

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 530–572.

Gilbert, R., Salanti, G., Harden, M., & See, S. (2005). Infant sleeping position and the sudden infant death syndrome: Systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology, 34*, 874-887.

Hunt, M. (1997). *How science takes stock. The story of meta-analysis*. New York: Sage

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Mathews, M. M. (1966). *Teaching to read, historically considered.* Chicago: University of Chicago Press.

McGuinness, D. (2004). *Early reading instruction: What science really tells us about how to teach reading*. Cambridge, MA: MIT Press.

National Reading Panel (2000). *Report*. Washington, DC: National Institute of Child Health and Human Development.