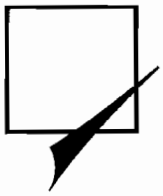
Evaluation

SAGE Publications (London,
Thousand Oaks and New Delhi)
DOI: 10.1177/1356389005053196
Vol 11(1): 95–109



Evidence-based Decision Making: Enhancing Systematic Reviews of Program Evaluation Results in Europe

WILLIAM R. SHADISH
University of California, USA

SALVADOR CHACÓN-MOSCOSO
University of Seville, Spain

JULIO SÁNCHEZ-MECA
University of Murcia, Spain

Over the last 25 years, meta-analysis has been widely used to study the effects of practical treatment interventions in fields ranging from psychology and education to medicine and public health. The present article first describes the impact of meta-analysis. The article then presents a description of some preliminary results of a study of the main design characteristics of published interventions in Europe. In order to foster both better experimental designs and more systematic reviews in the European context, and to promote collaboration between different countries and research groups, the authors then describe and discuss the Campbell Collaboration (C2), a new international organization aimed at fostering public policies and practices based on systematic reviews of high-quality evidence. Though there is some overlap between the work of the Campbell Collaboration and its sibling organization the Cochrane Collaboration, Campbell focuses more on systematic reviews in the social, behavioral and educational areas. Reviewers in those areas encounter difficulties over the use of non-randomized designs, difficulties in effect-size computation, and the need for new computer software to better serve the needs of meta-analytic reviews in the social sciences.

KEYWORDS: Campbell Collaboration; experiment; meta-analysis; program evaluation; quasi-experiment; systematic review

Introduction

It has been 26 years since Glass (1976) first coined the term meta-analysis in his Presidential Address to the American Educational Research Association. Glass distinguished between primary, secondary and meta-analysis, with these three phrases intended to distinguish between three different kinds of statistical analyses. According to Glass, primary analysis is 'the original analysis of data in a research study' (1976: 3). Secondary analysis is 'the re-analysis of data for the purposes of answering the original research question with better statistical techniques, or answering new questions with old data' (p. 3). Meta-analysis is 'the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings' (p. 3). In the next few years, Glass outlined a set of methods for doing meta-analysis, and provided a number of applications of those methods in both psychology and education.

A meta-analysis differs from a systematic review. A systematic review is 'a review that has been prepared using a systematic approach to minimizing biases and random errors which is documented in a materials and methods section' (Egger et al., 2001: 5). That is, what makes a review systematic is the methods it uses for identifying and selecting studies to be included in a review. Systematic reviews may or may not use the statistical methods of meta-analysis. For example, the review may determine that no studies exist that meet the selection criteria, or only one exists, making the use of meta-analytic methods unnecessary (Cooper, 1998; Hedges and Olkin, 1985).

The use of meta-analytic methods has spread from the social sciences such as psychology and education to medicine and public health, and even to fields such as entymology (Shadish et al., 2002). This article begins by making some observations about the use and impact of meta-analysis in both the United States and in Europe. Although many meta-analyses and systematic reviews have been done in both contexts, most of the primary experiments that are being reviewed in these efforts were conducted in the US. To shed light on this situation, some new data are presented about the kinds of research designs that have been used in the European context. The article ends with a discussion of the Campbell Collaboration (C2), a new international organization aimed at fostering public policies and practices based on systematic reviews of high-quality evidence. The international focus of the Campbell Collaboration may help to promote collaborative systematic reviews in both Europe and the US, so that both contexts may benefit from the work of each other.

The Impact of Meta-Analysis

Meta-analysis has had an enormous impact on how scientists review research literature. More researchers are doing meta-analyses and even devoting their careers to it, and meta-analytic methods and computer programs are advancing in sophistication and specialization. Reviews of the literature that used to be done narratively 25 years ago are often done meta-analytically today, especially reviews of treatment effectiveness. Presumably, the body of scientific knowledge

that has resulted from all this work is now more accurate, more detailed, and much better able to keep track of results from the ever-growing numbers of primary studies that are themselves appearing with increasing rapidity (Cooper and Hedges, 1994; Lipsey and Wilson, 2001; Shadish, 2002; Whitehead, 2002).

In the US, meta-analysis has been used in public policy (Anderson, 1992) and has influenced the everyday work of practitioners in fields like education psychotherapy, and medicine (Lamas et al., 1992). In education, for example, meta-analytic reviews that suggested that smaller class sizes led to better outcomes generated public policies in many US states to reduce class size. Psychotherapy meta-analyses have been cited in public debates about whether psychotherapy should be reimbursed under public or private health insurance. In medicine, the Agency for Healthcare Research and Quality (AHRQ) of the US government is now funding a total of 15 Evidence-Based Practice Centers whose function is to do meta-analytic reviews of outcome research on medical and public health questions;[1,2] and the results of those reviews are then disseminated to physicians in an effort to influence their clinical decision making. The US General Accounting Office (GAO), a branch of the US Congress, has relied extensively on meta-analytic methods in making policy recommendations to the Congress (US General Accounting Office, 1992).

Since its origins in 1976, both the theory and practice of meta-analysis have expanded rapidly. By 1990, it was clear that thousands of meta-analyses had been published. For example, Cooper and Lemke (1991) found 420 meta-analyses in psychology by 1987; Lipsey and Wilson (1993) found 302 meta-analyses on the effects of practical treatments for applied problems in psychology and education; Dickersin et al. (1990) found 119 meta-analyses reporting results of medical clinical trials.

Today, over 25 years after the start of meta-analysis, a definitive count of the number of completed meta-analyses is probably impossible to obtain. In English-language psychotherapy alone, we have located over 250 meta-analyses;[3] and the more narrowly focused topic of marriage and family therapy had 20 meta-analyses by 2002 (Shadish and Baldwin, 2002). Such numbers are small compared to the quantity of meta-analyses appearing in medicine and public health.

Although meta-analytic work began and initially blossomed in the US, European researchers quickly began using and improving such methods. Of particular note is the Cochrane Collaboration (www.cochrane.org), an international collaboration with a Secretariat located in Oxford, England, founded in 1993 and named after the British epidemiologist Archie Cochrane. Its focus tends to be mostly on medical and public health interventions, and its electronic database, called the Cochrane Library (http://www.cochrane.org/reviews/clibintro.htm), lists many thousands of systematic reviews. Meta-analysis has been used less frequently to review studies in other areas in the European context, mainly in social and psychological areas (e.g. Lösel and Koferl, 1989; Redondo et al., 1997, 1999, 2002; Moreno et al., 2001; Sánchez-Meca et al., 1999).

However, our experience is that most of the primary research studies that have been included in European systematic reviews have been conducted in the US, not in Europe. The main exceptions to this are a few European reviews that

specifically limit their focus to a language-based literature, for example, German-language psychotherapy studies (e.g. Beelman and Schneider, 2003). The primary reason for this lacuna seems to be the comparative lack of high-quality experiments conducted in Europe in the areas reviewed. However, few good data exist on the prevalence of such experiments, so now some preliminary results of work the authors have done is presented to provide such data.

Description of Main Design Characteristics of Published Interventions in the European Context

In this section, some preliminary results are presented from the ongoing work to characterize various aspects of the research designs used in published articles about randomized and non-randomized interventions in Europe (Chacón-Moscoso et al., 2002, 2004). This information has been updated during the last three years (June 2001–June 2004) as part of a larger project to identify and characterize experimental work in Europe. The design dimensions examined are relevant to the quality of the evaluations and of the meta-analytic studies based on those evaluation results.

The following databases were used to obtain published studies of interventions in Europe: Psycinfo (1887–2004); ERIC (1966–2004); Current Contents (1999–2004) and EBSCO Online (1997–2004). The keywords used to select articles (alone and using all possible combinations) were the following: random; non-random; effect size; quasi-experimental; experimental; meta-analysis; intervention program; evaluation; social; education; assessment. After carrying out this search, the authors obtained a total population of 1414, of which 307 were studies of the effects of intervention programs in the European context. Because the review of these studies is still in progress, with only data from the abstracts being coded across all studies, the presented data are gleaned only from those abstracts. Although many of the design features that were coded are consistently reported in abstracts (e.g. whether random assignment was used), some are probably not well-reported in abstracts (e.g. treatment standardization). Once all studies have been fully coded, it will be possible to present a more complete and accurate version of what is currently only a pilot effort.

Studies were included if they reported sufficient data to code the dimensions of interest to us, and if they reported results of interventions on humans rather than on laboratory animal work. We treated different reports of the same study as one report if they all used the same participants and intervention. For data management, we have used procite-5 to manage data from online databases available at the University of Seville; and SPSS 11.0 was used to analyze data. In Table 1, a summary is presented of the main characteristics of these articles; and those results are summarized in narrative form here.

Type of publication Most of these studies have been published in scientific journals (95.4%), and only a small percentage were doctoral theses and book chapters.

Table 1. Main Characteristics of Published Intervention Programs in Europe

Variables	*Frequency*	*Percentage*
Type of publication		
Journal	293	95.4
Book	8	2.6
Doctoral thesis	2	0.7
Conference	2	0.7
Others	2	0.7
Theoretical orientation		
Specified	51	16.6
Inferred	218	71
Not codified	38	12.4
Context of implementation		
Clinical	31	10.1
Health	86	28
Education	70	22.8
Community	78	25.4
Organizational	10	3.3
Others	32	10.4
Age (mean = 23.09)		
Specified	71	23.1
Not specified	145	47.2
Not codified	91	29.6
Intervention context		
Rural	13	4.2
Urban	19	6.2
Mixed	222	72.3
Not codified	56	17.2
Random assignment of units		
Yes	40	13
Not, but specifying control techniques	218	71
Not and without control techniques	3	1
Not codified	46	15
Methodology		
Experimental	40	13
Quasi-experimental	114	37.1
Pre-experimental	105	34.2
Not codified	48	15.6
Sample size		
$n < 5$	5	1.6
$5 \leq n \leq 10$	20	6.5
$n > 10$	209	68.1
Not codified	73	23.8
Attrition		
<30%	194	63.2
≥30%	36	11.7
Without attrition	7	2.3
Not codified	70	22.8

Table 1. Continued

Variables	*Frequency*	*Percentage*
Follow-up period		
<6 months	174	56.7
6–12 months	34	11.1
>12 months	55	17.9
Not codified	44	14.3
Time of measurement		
Post-test only	158	51.5
Pre-test and post-test	109	35.5
Not codified	40	13
Type of standardized dependent variables		
Non-standardized	4	1.3
Standardized questionnaires	238	77.5
Objective measures	23	7.5
Not codified	42	13.7
Use of standardized dependent variables in pre-test and post-test		
None	3	1
Some	95	30.9
All	15	4.9
Not codified	194	63.2
Intervention homogeneity		
Not homogeneous	12	3.9
Homogeneous	251	81.8
Not codified	44	14.3
Effect size		
Significant	24	7.8
Not significant	7	2.3
Not specified	276	90

Theoretical orientation Most of these publications present little or no data about the theoretical frameworks that buttress the interventions (71%); and only 16.6 percent specified the theoretical orientation clearly.

Context of implementation The most frequent substantive contexts for these studies were health (28%), community action (25.4%), education (22.8%) and psychotherapy (10.1%). Studies of organizational interventions are relatively rare (3.3%).

Age The age of study participants is not frequently reported. However, it appears that studies about adolescents and older people are most frequent.

Intervention context Most studies were implemented in both rural and urban contexts at the same time (72.3%). Of the remainder 4.2 percent were only implemented in rural contexts and 6.2 percent were only implemented in urban contexts.

Random assignment of units Random assignment was used in only a very small percentage of studies (13%). Many more (71%) used another kind of control.

Methodology or design Most studies used a pre-experimental design (34.2%) or a quasi-experimental design (37.1%). Only a minority used a randomized experimental design (13%).

Sample size Most studies used a sample size of more than 10 participants, though the sample sizes were rarely very large.

Attrition A few studies experienced no attrition at all, but most (63.2%) had an attrition rate lower than 30 percent.

Follow-up period Most (56.7%) of the studies carried out a follow-up outcome assessment in the first six months. Long-term follow-ups occurred more rarely, either the second six months (11.1%) or over a year after the intervention (17.9%).

Time of measurement Relatively few (35.5%) of the studies used pre-intervention measurements.

Use of standardized dependent variables Most studies used post-test outcome variables measured by standardized instruments with suitable psychometric properties, for example, previously published questionnaires or objective measures (85%). Fewer (30%) used these measures at both pre-test and post-test, however.

Intervention homogeneity In most of the studies (81.8%), study participants received the same intervention in terms of duration, intensity, program components and contextual features.

Effect size Of the studies reviewed, 90 percent did not present the estimated effect size. When specified, only 7.8 percent were significant.

These data suggest some potential reasons why past meta-analyses have included relatively few experiments that were conducted in Europe. To begin with, relatively few such experiments exist from Europe compared to the many thousands that have been conducted in the US. In addition, without the data to calculate effect sizes, a study cannot be included in a meta-analysis. And perhaps most important, the fact that most studies used non-experimental or quasi-experimental methods would often result in the elimination of such studies from systematic reviews using strict inclusion criteria. A real paucity of randomized experiments exists in the European context.

For example, consider two meta-analyses about the same substantive area, the effectiveness of delinquency rehabilitation programs, one of them carried out in Europe (Redondo et al., 2002) and the other one in the US (Lipsey, 1992). In the meta-analysis carried out by Lipsey (1992), half of the studies were

randomized; in the European meta-analysis, only 8.7 percent were randomized A similar pattern has been observed in a paper presented in the last Campbell Collaboration Colloquium, where we compared the designs used in published work about interventions implemented in the US and other countries (Chacón-Moscoso et al., 2004).[4]

A possible explanation for differences detected between USA and Europe design feature characteristics is that in Europe there is a higher degree of centralization and locally implemented intervention programs. This can add to difficulties in carrying out meta-analysis if it results in a restriction of range in the sheer number of studies done, or if it results in the relative exclusion of some kinds of designs compared to others.

The Campbell Collaboration (C2)

One effort that promises to help remedy this situation is the recent formation of the Campbell Collaboration. The Campbell Collaboration is an international organization that aims to prepare, maintain and disseminate high-quality, systematic reviews of studies of effectiveness of social and educational policies and practices (Petrosino et al., 2001). By supporting the production of these reviews and by disseminating results in an accessible fashion, the Campbell Collaboration intends to contribute to decisions in practice, policy and to public understanding (www.campbellcollaboration.org).[5]

The decision to establish the Campbell Collaboration was taken by 80 people from four countries at an exploratory meeting at University College London in July 1999. The Collaboration was formally established at a meeting at the University of Pennsylvania on 24–5 February 2000. C2 was modeled after the Cochrane Collaboration, with its Secretariat initially at the University of Pennsylvania but with an international membership. Its annual meetings currently alternate between the US and outside the US (e.g. Stockholm, Lisbon), and it includes participants from around the globe. It was named after the US scientist Donald T. Campbell because of his long interest in using experiments to find effective solutions to social problems. Hence it has also emphasized the use of high-quality experimental designs in the primary research that goes into systematic reviews.

The systematic reviews of research evidence prepared and maintained by contributors to the Campbell Collaboration's Review Groups are designed to meet the needs of those with a strong interest in high-quality evidence on 'what works'. These include members of the public who want to keep abreast of the best evidence on the effects of social and educational policies and practices, service providers, policy makers, educators and their students, and professional researchers. Like Cochrane reviews, Campbell systematic reviews are published electronically so that they can be updated promptly as relevant additional evidence emerges, and amended in the light of criticisms and advances in methodology.

The Campbell Collaboration collaborates closely with its sibling organization, the Cochrane Collaboration. Cochrane has traditionally prepared and maintained systematic reviews of the effects of interventions in healthcare. Campbell

is meant to complement Cochrane by reviewing and synthesizing evidence on social and behavioral interventions and public policy, including education, criminal justice and social welfare, among other areas. However, boundaries are blurred between the purviews of the two organizations, with Cochrane having done many reviews in areas such as psychotherapy that could easily fall under the auspices of both organizations.

To do a systematic review under Campbell Collaboration auspices, researchers begin with a Review Protocol, that is, a document that sets out the reviewers' intentions with regard to (a) the topic and (b) the methods to be used in carrying out a proposed review for inclusion in the Campbell Database of Systematic Reviews of Interventions and Policy Evaluation (C2-RIPE), a component of the Campbell Library. Editors of Collaborative Review Groups and other referees deemed appropriate by the Editors appraise and give feedback on protocols before actual reviews are conducted. The protocols are published in the Campbell Database of Systematic Reviews and are subject to comments and criticisms from users of that Database. The main advantages of using protocols are to not duplicate efforts with respect to the same areas of review and to have the possibility of a continuous feedback throughout the review process.

People considering taking on the preparation and maintenance of a Campbell Review first need to discuss and register the title with one of the Campbell Collaborative Review Groups. This is done to help to avoid duplication of effort, especially the unhappy situation that two separate teams begin work on two separate reviews addressing the same (or similar) questions, only to find out about this duplication of effort late on in the process.

For both Cochrane and Campbell, however, the move toward doing systematic reviews of social interventions brought new challenges that were less frequently encountered in reviews of the effects of health interventions. Both organizations have formed committees to work on these issues, some with cross-membership and some working independently. The key issues they have faced include the following.

Experimental Design

In medical care, the randomized clinical trial is the most commonly used methodology for investigating treatment effects, and the Cochrane Collaboration tends to limit its reviews to studies using that method. Nonetheless, the Cochrane Collaboration has not adhered to this limitation strictly. For example, the Cochrane Library has registered reviews that consist in whole or in part of case control designs (Thompson and Rivara, 2001; Thompson et al., 2000), observational studies (Ansink and van der Velden, 2002), non-randomized control group designs (Osborn et al., 2002a, 2002b) and interrupted time series quasi-experiments (Briggs et al., 2002; Grilli et al., 2000). In this context, the Campbell Collaboration has also allowed the inclusion of randomized and non-randomized experiments. Both organizations adopted such policies for many compelling reasons.

First, randomized experiments are not as frequent in many social science fields. For some questions in social science there may never be enough randomized

studies to warrant a research synthesis. Thus, confining reviews of social science interventions only to randomized experiments might significantly curtail the topics about which systematic reviews could be done. Yet policy makers and practitioners will make decisions anyway, so providing them with a thoughtful, critical review of the findings in the field can be useful.

Second, non-randomized studies may offer some kinds of information less widely available in randomized experiments. For example, non-randomized designs may examine intervention features, settings or kinds of service recipients that are less amenable to random assignment, and that therefore would be under-represented in a database limited only to randomized experiments.

Third, non-randomized experiments themselves vary enormously in quality. For example, time-series and regression-discontinuity designs tend to yield generally strong causal inferences while post-test-only studies with no control group rarely do. To exclude all non-randomized designs discourages researchers from conducting high-quality non-randomized experiments.

Fourth, randomized experiments also vary in quality. Sometimes, they suffer from extremely high differential attrition rates, low sample sizes, and poor treatment implementation. These flaws can make their results far less helpful than better-conducted randomized experiments or even than some high-quality non-randomized experiments.

Fifth, while the synthesis of results across studies will not cancel out consistent and pervasive design flaws, to the extent that the strengths of some studies, randomized or not, compensate for the weaknesses of others (and vice versa), clear patterns that emerge across flawed bodies of evidence can be informative.

Neither Campbell nor Cochrane have developed rigid solutions for dealing with these issues. Campbell, for example, developed a Research Design Policy Brief that gave strong preference to randomized experiments, but included non-randomized experiments as long as results for the two kinds of experiments were reported separately. Cochrane has several working groups on experimental design issues, and review groups have considerable latitude in making such decisions depending on the nature of the literature they are reviewing. Considerably more development on these fronts can be expected over time.

Meta-analytic Methodology

Effect-size calculation Given that effect sizes are a defining feature of meta-analysis, it is crucial that they be computed accurately. This is quite a bit more difficult for the social sciences than for many health reviews, for the following reasons. In evaluating the effects of healthcare interventions, results are often reported in the form of a dichotomous outcome, for example, the patient survived or died, the patient improved or deteriorated, or the patient developed the disease or stayed disease free. Such dichotomous outcomes are routinely reported in a very small number of ways in primary research reports, such as the percentage of successful cases in treatment or control groups, or a fourfold table listing the number of successes and failures in each group. From these data, effect-size indices such as odds ratios are easily computed and aggregated.

The situation in social, behavioral and educational research is quite different. Results are sometimes reported as dichotomies, but most often they are reported on continuous scales for which the appropriate statistic is usually the standardized mean difference statistic, often called Cohen's *d*. Unlike the case with dichotomous outcomes, however, results on continuous scales are reported in a very wide variety of ways. Some of those ways lend themselves to simple effect-size computation, but others of them require extensive, detailed computations that are not widely known and that are subject to error from a variety of sources. Additional problems arise when continuous outcomes have been dichotomized by the researcher prior to reporting results (Sánchez-Meca et al., 2003). One computer program that is devoted to calculating the *d* statistic, for example, lists over 40 methods for doing so, methods that include various forms of multifactoral experiments, change score results, analysis of covariance results, and multivariate methods.[6] As a result, researchers in both Cochrane and Campbell have had to pay significant attention to locating and evaluating pertinent effect-size computation methods for experiments in the social, behavioral and educational areas.

Computer software The Cochrane Collaboration has made a major contribution to the conduct of meta-analysis with its Cochrane Collaboration Reviewers' Handbook (www.cochrane.org/cochrane/hbook.htm), and the associated RevMan software. However, initial experience with that protocol showed that it needed considerable adaptation for optimal use in reviews of social, behavioral and educational interventions. Meta-analysts in medical care tend to code relatively few variables to use as covariates of effects sizes, and their meta-analytic data sets usually have a fairly simple structure. In contrast, meta-analysts in the social sciences tend to code many substantive and methodological variables, and their data sets often have a very complex structure. For example, they often code variables about the study, about each treatment contrast within the study (and social science studies often have up to four or five treatment conditions, leading to n[n-1] comparisons), about each measure within the study, and about each effect size calculated for each measure-comparison combination. Consequently, researchers doing reviews of social, behavioral and educational interventions have devoted significant effort to identifying or developing software that might be more appropriate to their work.

Campbell Collaboration Accomplishments

These methodological quandaries notwithstanding, the Campbell Collaboration has already made significant progress towards producing systematic, meta-analytic reviews that may help to improve both policy and practice. First, it has established the Campbell Collaboration's Social, Psychological, Educational, and Criminological Trials Register (C2-SPECTR), a database of randomized and possibly randomized trials currently containing over 11,700 entries, searchable over the Internet at no cost.[7] Second, it has established the C2 Library, a register of Campbell systematic reviews of studies of interventions in the social, behavioral and education arenas (C2-RIPE).[8] At last count, 75 systematic reviews have

been registered in the library across education (13), social work (25), criminal justice (36) and methods (1), the vast majority of which are still in progress. The Campbell Collaboration, along with the American Institutes for Research, was awarded an $18.5 million dollar grant from the US Department of Education to establish a national *What Works Clearinghouse*, which will summarize evidence on the effectiveness of different programs, products and strategies intended to enhance academic achievement and other important educational outcomes. The clearinghouse will help provide education decision makers with the information they need to make choices guided by the best available scientific research.[9] Specifically, it will develop:

- an educational interventions registry that identifies potentially replicable programs, products and practices that are claimed to enhance important student outcomes, and synthesizes the scientific evidence related to their effectiveness;
- an evaluation studies registry, which is linked electronically to the educational interventions registry, and contains information about the studies constituting the evidence of the effectiveness of the program, products and practices reported;
- an approaches and policies registry that contains evidence-based research reviews of broader educational approaches and policies;
- a test instruments registry that contains scientifically rigorous reviews of test instruments used for assessing educational effectiveness;
- an evaluator registry that identifies evaluators and evaluation entities that have indicated their willingness and ability to conduct quality evaluations of education interventions.

Initial results from this clearinghouse are just beginning to appear as this article goes to print.

Although the focus of this section has been on the Campbell Collaboration, efforts similar to both Cochrane and Campbell are proliferating, including both the ESRC Evidence Network and the Evidence for Policy and Practice Information and Co-ordinating Centre in the United Kingdom, and the Meta-analysis Unit in Spain.[10,11,12] All of them are contributing in important ways to the development of systematic review methodology for social, behavioral and educational interventions.

Conclusion

In conclusion, then, from its start over 25 years ago (Glass, 1976), meta-analysis has grown into a multi-million-dollar enterprise producing thousands of systematic reviews that have made important contributions to both policy making and practice. More importantly, both meta-analysis and systematic reviews are still in their infancy. After another 25 years have passed, the authors believe these endeavors will have developed even more than they have to date, contributing to science, to policy making, and to practice in ways that are just beginning to be imagined today.

Notes

1. See www.ahrq.gov/
2. See www.ahrq.gov/clinic/epc/#centers
3. See http://faculty.ucmerced.edu/wshadish/index.htm
4. A copy of this paper is available at http://innoevalua.us.es/sajopang/congresos/washington/qual.pdf
5. In Europe, the Nordic Campbell Center has already been founded (www.sfi.dk/sw1270.asp).
6. See http://faculty.ucmerced.edu/wshadish/index.htm
7. See http://geb9101.gse.upenn.edu/
8. See www.campbellcollaboration.org/Fralibrary.html
9. See www.w-w-c.org/
10. See www.evidencenetwork.org
11. See http://epi.ioe.ac.uk/EPPIWeb/home.aspx
12. See www.um.es/UnidadMA1.html

References

Anderson, C. (1992) 'Congress Looks for Methods to Assess Clinical Research', *Nature* 357(6373): 5.

Ansink, A. and J. van der Velden (2002) 'Surgical Interventions for Early Squamous Cell Carcinoma of the Vulva (Cochrane Review)', *The Cochrane Library, 3*. Oxford: Update Software.

Beelman, A. and N. Schneider (2003) 'The Effects of Psychotherapy with Children and Adolescents: a Review and Meta-analysis of German-language Research', *Zeitschrift für Klinische Psychologie und Psychotherapie* 32(2): 129–43.

Briggs, C. J., P. Capdegelle and P. Garner (2002) 'Strategies for Integrating Primary Health Services in Middle- and Low-income Countries: Effects on Performance, Costs and Patient Outcomes (Cochrane Review)', *The Cochrane Library, 3*. Oxford: Update Software.

Chacón-Moscoso, S., J. Sánchez-Meca, D. Alarcón-Rubio and F. Marín-Martínez (2002) 'Main Characteristics of Intervention Designs of Published Papers in the European Context', Paper presented at the European Evaluation Society conference, Seville, Spain, 12–14 October.

Chacón-Moscoso, S., J. Sánchez-Meca, D. Alarcón-Rubio, F. Marín-Martinez, S. Sanduvete-Chaves and T. B. Huedo-Medina (2004) 'Quality of Program Interventions: a Comparative Study between USA and Europe', Paper presented at the 4th Campbell Collaboration Colloquium, Washington, DC, USA: February.

Cooper, H. (1998) *Synthesizing Research*, 3rd edn. Thousand Oaks, CA: Sage.

Cooper, H. and L. V. Hedges, eds (1994) *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Cooper, H. M. and K. M. Lemke (1991) 'On the Role of Meta-analysis in Personality and Social Psychology', *Personality and Social Psychology Bulletin* 17(3): 245–51.

Dickersin, K., K. Higgins and C. L. Meinert (1990) 'Identification of Meta-analyses: the Need for Standard Terminology', *Controlled Clinical Trials* 11(1): 52–66.

Egger, M., G. Davey Smith and K. O'Rourke (2001) 'Rationale, Potentials, and Promise of Systematic Reviews', in M. Egger, G. Davey Smith and D. G. Altman (eds) *Systematic Reviews in Health Care: Meta-analysis in Context*, pp. 3–19. London: BMJ Books.

Glass, G. V. (1976) 'Primary, Secondary, and Meta-analysis', *Educational Researcher* 5(10): 3–8.

Grilli, R., N. Freemantle, S. Minozzi, G. Domenighetti and D. Finer (2000) 'Mass Media Interventions: Effects on Health Services Utilization (Cochrane Review)', *The Cochrane Library, 3.* Oxford: Update Software.

Hedges, L. V. and I. Olkin (1985) *Statistical Methods for Meta-analysis.* Orlando, FL: Academic Press.

Lamas, G. A., M. A. Pfeffer, P. Hamm, J. Wertheimer, J.-L. Rouleau and H. Braunwald (1992) 'Do the Results of Randomized Clinical Trials of Cardiovascular Drugs Influence Medical Practice?', *New England Journal of Medicine* 327(4): 241–7.

Lipsey, M. W. (1992) 'Juvenile Delinquency Treatment: a Meta-analytic Inquiry into the Variability of Effects', in T. D. Cook, H. Cooper, D. S. Cordray et al. (eds) *Meta-analysis for Explanation: a Casebook*, pp. 83–127. New York: Russell Sage Foundation.

Lipsey, M. W. and D. B. Wilson (1993) 'The Efficacy of Psychological, Educational, and Behavioral Treatment: Confirmation from Meta-analysis', *American Psychologist* 48(12): 1181–209.

Lipsey, M. W. and D. B. Wilson (2001) *Practical Meta-analysis.* Thousand Oaks, CA: Sage.

Lösel, F. and P. Koferl (1989) 'Evaluation Research on Correctional Treatment in West Germany: A Meta-analysis', in H. Wegener, F. Lösel and J. Haisch (eds) *Criminal Behaviour and the Justice System: Psychological Perspectives*, pp. 334–55. New York: Springer-Verlag.

Moreno, P. J., F. X. Méndez and J. Sánchez-Meca (2001) 'Effectiveness of Cognitive-behavioural Treatment in Social Phobia: a Meta-analytic Review', *Psychology in Spain* 5(1): 17–25.

Osborn, D. A., M. J. Cole and H. E. Jeffery (2002a) 'Opiate Treatment for Opiate Withdrawal in Newborn Infants (Cochrane Review)', *The Cochrane Library, 3.* Oxford: Update Software.

Osborn, D. A., H. E. Jeffery and M. J. Cole (2002b) 'Sedatives for Opiate Withdrawal in Newborn Infants (Cochrane Review)', *The Cochrane Library, 3.* Oxford: Update Software.

Petrosino, A., R. F. Boruch, H. Soydan, L. Duggan and J. Sánchez-Meca (2001) 'Meeting the Challenges of Evidence-based Policy: the Campbell Collaboration', *Annals of the American Academy of Political and Social Science* 578(6: Nov): 14–34.

Redondo, S., V. Garrido and J. Sánchez-Meca (1997) 'What Works in Correctional Rehabilitation in Europe: A Meta-analytical Review', in S. Redondo, V. Garrido, J. Pérez and R. Barberet (eds) *Advances in Psychology and Law: International Contributions*, pp. 499–523. Berlin: De Gruyter.

Redondo, S., J. Sánchez-Meca and V. Garrido (1999) 'The Influence of Treatment Programmes on the Recidivism of Juvenile and Adult Offenders: an European Meta-analytic Review', *Psychology, Crime and Law* 5: 251–78.

Redondo, S., J. Sánchez-Meca and V. Garrido (2002) 'Crime Treatment in Europe: a Review of Outcome Studies', in J. McGuire (ed.) *Offender Rehabilitation and Treatment: Effective Programmes and Policies to Reduce Re-offending*, pp. 113–41. Chichester: Wiley.

Sánchez-Meca, J., F. Marín-Martínez and S. Chacón-Moscoso (2003) 'Effect Size Indices for Dichotomized Outcomes in Meta-analysis', *Psychological Methods* 8(4): 448–67.

Sánchez-Meca, J., J. Olivares and A. I. Rosa (1999) 'The Problem of Tobacco Addiction: Meta-analysis of Behavioural Treatments in Spain', *Psychology in Spain* 3(1): 36–45.

Shadish, W. R. (2002) 'Meta-analysis in the United States: Some Observations', Paper presented at the European Evaluation Society conference, Seville, Spain: 12–14 October.

Shadish, W. R. and S. A. Baldwin (2002) 'Meta-analysis of MFT Interventions', in D. Sprenkle (ed.) *Effectiveness Research in Marriage and Family Therapy*, pp. 339–70. Washington, DC: American Association of Marriage and Family Therapy.

Shadish, W. R., T. D. Cook and T. D. Campbell (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Thompson, D. C. and F. P. Rivara (2001) 'Pool Fencing for Preventing Drowning in Children (Cochrane Review)', *The Cochrane Library, 4*. Oxford: Update Software.

Thompson, D. C., F. P. Rivara and R. Thompson (2000) 'Helmets for Preventing Head and Facial Injuries in Bicyclists (Cochrane Review)', *The Cochrane Library, 3*. Oxford: Update Software.

US General Accounting Office (1992) *Cross Design Synthesis: a New Strategy for Medical Effectiveness Research* (GAO/PEMD-92–18). Washington, DC: US Government Printing Office.

Whitehead, A. (2002) *Meta-analysis of Controlled Clinical Trials*. Chichester: Wiley.

WILLIAM R. SHADISH is Professor at the University of California, Merced. He is author of over 100 articles and chapters. Please address correspondence to: School of Social Sciences, Humanities, and Arts, University of California, Merced, PO Box 2039, Merced, CA 95344, USA.
[email: wshadish@ucmerced.edu]

SALVADOR CHACÓN-MOSCOSO is Associate Professor, University of Seville. Please address correspondence to: Departamento de Metodología de Psicología Experimental, Facultad de Psicología, Universidad de Sevilla, C/ Camilo José Cela, 41018 Sevilla, Spain. [email: schacon@us.es]

JULIO SÁNCHEZ-MECA is Professor, University of Murcia. His main research interests are related to meta-analysis and statistical modelling. Please address correspondence to: Departamento Psicología Básica y Metodología, Facultad de Psicología, Campus de Espinardo, Apdo 4021, Universidad de Murcia, 30100-Murcia, Spain. [email: jsmeca@um.es]