

Methods Group Newsletter
 Issue 2
 September 2002

Greetings! Welcome to the second issue of the Campbell Collaboration's Methods Group Newsletter. The purpose of this newsletter is to keep interested individuals abreast of C2 developments that pertain to methodological issues. As such, we have a lot inside. For example ... a joint venture between the Campbell Collaboration and the American Institutes for Research has been awarded a contract from the U.S. Department of Education for the What Works Clearinghouse ... two methods policy briefs have been approved by the C2 Steering Committee ... a new training group has been approved ... and the first C2 Methods Conference is scheduled for September. Enjoy!

What Works Clearinghouse



A joint venture between the Campbell Collaboration and the American Institutes for Research has been awarded \$18.5 million over five years for the **What Works Clearinghouse** (WWC). The WWC will provide the following easily accessible and searchable databases:

- 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.

1st C2 Methods Group Conference

Posters & Presenters

- Meta-analysis for correlation matrices: Random effects refinements (Adam Hafdahl)
- Broadening the scope of available evidence (Philip Satherley)
- Secondary data analysis as research synthesis (Joel Garner)
- Empirically-based effect size estimates: Normative summaries in social psychology (Dan Richard & Charles Bond)
- Choosing appropriate criteria for assessing the quality of 'qualitative' research for inclusion in systematic reviews: a review of the published literature (Angela Harden)
- Evaluation results generalization, design quality, and meta-analysis (Salvador Chacon Moscoso)
- Methodologies for the systematic synthesis of non-experimental designs (David Gough, Ann Oakley & Diana Elbourne)



Reports of Methods Groups' Activities

C2 Training Group Established

By Betsy Becker and Therese Pigott

In Spring of 2002 a proposal was submitted, and accepted, to form a group focused on training under the umbrella of the Campbell



Collaboration (C2) Methods Group. Conveners Betsy Becker (Michigan State University) and Therese Pigott (Loyola University Chicago) worked with the Campbell Steering Committee to develop the concept for this new area of interest.

The C2 Training Group will serve as the formal body in the C2 Methods Group to address the goals of providing training and support for review groups within C2. As C2 review groups in different substantive areas (i.e., education, criminal justice, etc.) begin to conduct and possibly solicit C2 reviews, we anticipate that the substantive experts conducting those research reviews will want instruction on methodology for systematic reviews, from problem formulation through the reporting of results. The C2 Training Group will serve as a clearinghouse for review groups who need training in any aspect of systematic reviews. Review groups could request training in a specific area or set of areas, and the C2 Training Group would enable or assist those reviewers in locating a trainer (or group of trainers) to respond to the need.

We expect that the C2 Training Group will provide both formal training courses and shorter term instruction on other issues involved in systematic reviews.

The membership of the Training Group will draw from the international pool of scholars associated with C2 who have extensive experience in conducting systematic reviews, and those who actively conduct research on methods for meta-analysis. To date over 30 scholars from four countries have indicated interest in the Training Group.

The C2 Training Group expects to eventually establish and provide:

1. a clearinghouse for training opportunities,
2. specialized methods training,
3. training for trainers and a forum for interaction among trainers,
4. training workshops presented at C2 meetings and other venues,
5. a collection of training materials, and
6. follow-up and evaluation of training opportunities.

...the C2 Training Group will help researchers to make the best use of existing evidence.

We expect that the C2 Training Group will provide both formal training courses and shorter term instruction on other issues involved in systematic reviews. Members of the group will also receive instruction on new techniques for systematic reviews, so that a core group of trainers is available across the countries represented in C2 who could provide instruction on a variety of review topics. As part of the training process, the C2 Training Group also plans to set up a listserv for trainers and to compile a set of training materials to be used for instructional purposes.

The formation of the C2 Training Group should enable more researchers to become proficient in the methods for systematic review and the analysis of data gathered in the review process. Through the six activities outlined here (and others that may develop in the future), the C2 Training Group will help researchers to make the best use of existing evidence. In the long run this should lead to more informed decisions about what new or additional evidence is needed to inform critical policy decisions across the social sciences. Also this group will facilitate communication among scholars interested in providing the best educational opportunities for persons interested in learning about systematic reviews and meta-analysis.

If you are interested in becoming active in the C2 Training Group please contact either Betsy Becker (bjbecker@msu.edu) or Therese Pigott (tpigott@luc.edu) for more information.

The Implementation Process Methods Group, January - July 2002

By Jennie Popay

Six months since we last reported on our work and looking back, time seems to have flown! The group's priority for the first couple of years is to begin to clarify the role of research evaluating implementation processes in Campbell Systematic Reviews and to contribute to the development of methods for review and synthesis of this type of evidence. This has involved members of the group running workshops to stimulate debate and discussion and developing proposals for methodological research.

Annual Methods Group Meeting - February 2002

Members of the IP Methods Group met at the Campbell Colloquium in February this year in the USA and continued a long running debate about the precise focus of the group and, in particular, the extent to which it included methodological work on qualitative research appraisal and synthesis. As the registration protocol makes clear, the original decision of the Steering Group was that the Collaboration should have a methods group concerned with the methodological implications of including evidence on implementation process issues in systematic reviews. The registration document recognizes that evaluative research on implementation processes is multi-method, including but not restricted to qualitative research. The Cochrane Collaboration had an existing network of people interested in developing methods for the synthesis of qualitative research evidence and they are shortly to seek registration as a methods group. Jennie Popay is one of the conveners of this proposed Cochrane Methods Group and in earlier discussions of the role of the Campbell Implementation Process Group it has been assumed that in the future these two groups would become a joint methods group across the Cochrane and Campbell Collaborations in acknowledgement that the issues facing the two groups and the membership are overlapping.

The group's priority... is to clarify the role of research evaluating implementation processes in Campbell Systematic Reviews and to contribute to the development of methods for review and synthesis of this type of evidence.

Unfortunately, Jennie Popay could not be at the Campbell Colloquium in February so this approach as a possible resolution of the issues raised at the meeting was not discussed. The conveners of the IP Methods Group will therefore be producing a position paper on the focus of the group and its relationship to the proposed Cochrane Qualitative Methods Group. This will be circulated to members of the Campbell IP Methods group and discussed at a meeting of the Methods Group at the 2003 Collaboration Colloquium in Sweden. Recommendations from this meeting will then be taken to the Campbell Methods Steering Group.

A New Website

The protocol for the proposed Cochrane Qualitative Methods Group is posted on their new website: http://mysite.freereserve.com/Cochrane_Qual_Method/index.htm. This is to be a joint website with the Campbell IP Methods Group and our registration protocol is already posted on the site. In the autumn, members of the Campbell IP Methods Group will be surveyed to ask them to provide contact details, to identify their interests and to say which aspects of the Campbell I&P Methods Groups activities they would be willing to contribute to.

Other Activities

With financial support from the UK based Social Care Institute of Excellence (SCIE) Jennie Popay also organised an introductory seminar in London in July for people interested in developing expertise in the synthesis of evidence from process evaluations and qualitative research. This was well attended and there was a lively discussion of a range of policy, practice and methodological issues. A second SCIE seminar is planned for the winter and the International Seminar on evidence synthesis, described in our January progress report, has been postponed so that too will be going ahead this coming winter. Also earlier on this year Jennie Popay spoke about the work of the Campbell IP Methods Group at an informal workshop for social work directors in Sydney Australia hosted by the Methods Group's co-convenor Roger Dunston. In September we are also hoping to organise a plenary session at the Campbell Collaboration Methods Group focusing on the role of qualitative research and process evaluations in Campbell Reviews and we are looking for offers of papers to be included in this.

Methodological Research

As we reported previously, exploratory methodological work is already underway in the UK looking at the feasibility of extending two existing systematic reviews to include evidence from evaluative studies of implementation processes and qualitative studies of the experience of recipients of an intervention: one focusing on smoke alarms and the other on community oriented interventions. Additionally, a group of researchers in the England and Scotland have written a proposal for work aiming to develop more systematic and transparent methods for the narrative synthesis of quantitative and qualitative evidence from studies using a diverse range of methods. This proposal was short-listed for funding in the ESRC's research methods programme but ultimately, although well reviewed it was not funded. Undeterred, however, we will be submitting it elsewhere in the next few months!

The Future

As noted above, our priority for the first couple of years has to be to demonstrate the value of including diverse types of evidence within Campbell Reviews as a complement to evidence on effectiveness drawn from experimental and quasi-experimental studies. It seems clear that there is currently a lack of clarity about this role within the Collaboration. It is also not clear how many (if any) protocols for systematic reviews going through the editing process describe plans to include evidence from qualitative studies and/or evaluation studies of implementation and other processes. We certainly haven't been asked to comment on such plans! To push forward this agenda it is important that members of the group are able to show practical examples of the benefits that flow from extending the scope of systematic reviews in this way. This in turn means that we have to develop a programme of methodological work and create and maintain a database of systematic reviews that include such evidence. The group has agreed to produce a policy brief for the Steering Group setting out some initial guidelines to help those wishing to extend the lens of Campbell Systematic Reviews to include evidence from a range of study designs - this will be developed over the coming year and offers of support for this would be gratefully received. Finally, we are still looking for conveners for the group from countries other than the UK and Australia so if there is anybody out there interested please contact Jennie Popay.

If you want to get involved with the work of the group or simply find out more about us then contact one of the conveners, preferably by e-mail.

Methods convener

Professor Jennie Popay (Primary contact)
Institute for Health Research
Lancaster University, Lancaster LA1 4YT
Telephone (+44) 1524 59-2127
Fax (+44) 1524 59-2401
E-mail j.popay@lancaster.ac.uk

Policy/practice convener

Roger Dunston
Deputy Head, Division of Allied Health Services
C/O Royal North Shore Hospital
St Leonards, NSW 2065 Australia
Telephone 61 (02) 9926 7580
Email Roger.dunston@pgrad.arts.usyd.edu.au

What do we know about the performance of quasi-experimental evaluation methods? A systematic review in progress

By Steven Glazerman, Dan Levy, and David Myers
Randomized experiments, while desirable, are often perceived as infeasible. The result is a heavy imbalance in the literature on the effectiveness of social interventions towards quasi-experimental (or "non-experimental") evidence. Thus, many practitioners of research synthesis need guidance to determine when to include and how to treat non-experimental findings in their reviews.

Under what conditions can non-experimental methods produce findings similar to those of a well designed and executed randomized experiment?

Our study reviews a body of methodological research that tries to address this need using empirical evidence. The evidence is drawn from case studies, known as design replication studies, that directly test non-experimental designs against an experimental benchmark. The test is conducted by taking a randomized experiment and trying to replicate its result using a variety of non-experimental methods. The difference between the experimental and each non-experimental program impact estimate is itself

an estimate of the bias. Our study asks the question: Under what conditions can non-experimental methods produce findings similar to those of a well designed and executed randomized experiment?

Methods

Our study is being conducted much like a Campbell Review, but instead of the effectiveness of an intervention, we are synthesizing estimates of the bias associated with different non-experimental estimators. Design replication studies are those for which at least one bias estimate can be extracted. We conducted an extensive literature search, but restricted the search to studies in which the intervention was related to education, training, or employment services. We extracted over 1,300 bias estimates from the 16 completed studies we identified. We identified two more design studies in progress.

Preliminary Findings

A very preliminary review of the evidence suggests that the 16 design replication case studies, even taken together, will not resolve any enduring debates about the use of non-experimental methods. The range of methods we examined - mostly nonequivalent comparison group designs with either regression or propensity score matching methods - were effective in isolated cases, but often were very inaccurate in approximating the experimental result.

Half of the design replication studies we identified found no support for the validity of the non-experimental methods they tried. In the other half, the authors concluded that one or more of the non-experimental approaches was promising, but the conditions under which a randomized control group can be successfully mimicked are still largely a matter of speculation. The size of the bias estimates did not appear to depend on the type of intervention.

Our preliminary analyses were able to identify some factors that were associated with lower predicted bias. These include the comparability of the comparison group -- whether it was drawn from within the evaluation itself or based on a national dataset, and whether it was locally matched to the treatment population - and the availability of good background data, such as pre-intervention measures of the outcome, to adjust for differences. We did not find

that matching approaches (most of which used propensity scores) were significantly more effective than simple regression in reducing bias. Specification tests, sometimes used to select among alternative non-experimental estimators, provided inconsistent guidance on selecting the best estimators.

The bias estimates were both positive and negative, and their distribution across the literature as a whole was centered roughly around zero. Within any one context, however, the estimates tended to be either mostly positive or mostly negative, with no obvious way to predict the direction of bias.

We caution that these findings are based on some preliminary analyses of a still sparse design replication literature. We are continuing to refine the analysis and will update the database as more design replication studies are completed.

Content Analysis of C2 Protocols

By Jeff Valentine

The following is a report of a content analysis of C2 protocols, undertaken in March 2002. Fifteen protocols were included in the content analysis. These protocols were reviewed by 10 different Methods Group critiquers. The following summary was undertaken to identify the most common comments made by critiquers.

As is common in scientific peer review, critiquers tended to be problem focused. That is, they focused much more on what was wrong with the protocols than on what was right with them. Thus, there was little consensus about the positive aspects of the protocols as a whole. In general however, protocols were judged to be well conceived and well written, with very good prospects for informing public policy.

Positive Comments

- Description of the background clear and thorough (47% of protocol critiques)
- Protocol was clear, well-written, largely complete (33% of protocol critiques)
- Search strategy well-specified (33% of protocol critiques)
- Objectives of the review are reasonably clear (27% of protocol critiques)
- Data extraction sheet was clear and detailed

- Protocol is an excellent exemplar for future protocols
- Double-coding all studies is an excellent way to address coding reliability
- Worthwhile project with excellent potential to affect public policy
- Interventions are clearly defined
- Methodological categories well defined

Critical Suggestions

- Data synthesis methods inadequately described (100% of protocol critiques)
- Inclusion / exclusion criteria not clear or not justified (87%)
- Criteria for determining independent findings not specified (67%)
- Intervention and/or participant moderators not sufficiently explicated (67%)
- Search strategy not clear or incomplete (53%)
- Details on study coding not clear (53%)
- Review background needs to be more thorough (53%)
- Study quality addressed in a problematic fashion (53%)
- Specific protocol element missing (53%)

Research Design Policy Brief

*Executive Summary

By William Shadish and David Myers

This Brief addresses the following key question for Campbell Collaboration (C2) reviews: *What should be C2 policy concerning acceptable methodologies used in primary studies when a systematic review concerns the effectiveness of an intervention?* The Brief:

- identifies the key issues that are confronted by C2 systematic reviewers who find a variety of study designs in their literature;
- outlines possible ways to represent this diversity in their work;
- proposes agreed-upon guidelines that C2 may wish to promulgate; and
- provides exemplars that demonstrate how these guidelines might be implemented in practical ways.

The heart of the Brief is a set of eight key issues, along with proposals for C2 policies for each. A summary of those issues and a summary of the proposals is:

1. Should the C2 reference database be limited to randomized experiments?

Proposal: The C2 database(s) should not be limited to randomized experiments.

2. If nonrandomized studies are included, should one C2 reference database include both randomized and nonrandomized studies, or should separate databases be constructed for the two kinds of studies?

Proposal: C2 should maintain two databases, one for randomized experiments and one for nonrandomized studies.

Proposal: C2 should address how to handle the borderline case of haphazard assignment, that is, assignment methods that appear to be functionally random in how they distribute bias over conditions.

Proposal: Group randomized trials with discrepancies between unit of assignment and unit of analysis or with small samples of aggregate units should not be combined with other randomized experiments.

3. What searchable fields should be available for users to identify research designs in the database(s)?

Proposal: The searchable fields initially included to identify research designs in the databases should be kept to the minimum that potential reviewers might need to select studies for their review. The list should include:

- Randomized Design (include group randomized designs here unless they meet criteria for the following code)
- Group Randomized Design with Discrepant Units of Analysis or with Inadequate Number of Aggregate Units Assigned to Conditions
- Quasi-Experiment: Interrupted Time Series Design
- Quasi-Experiment: Regression Discontinuity Design
- Quasi-Experiment: Nonequivalent Comparison Group Design
- Case Control Design
- Other Designs

4. Should C2 systematic reviews be limited to randomized experiments?

Proposal: With some exceptions described in more detail in the Brief, C2 reviews should not be conducted unless randomized experiments are available within the body of evidence to be reviewed.

However, such reviews need not be limited to randomized experiments.

5. If nonrandomized studies are included in C2 systematic reviews, what procedures should be required or recommended for determining what nonrandomized designs are legitimate for inclusion?

Proposal: Where both randomized and nonrandomized experiments are included, C2 reviews *must* separate estimates of intervention effects for randomized versus nonrandomized studies in important analyses. The Brief contains more specific recommendations about how to treat different kinds of designs in many other respects, including differentiating among better and worse randomized experiments, or better and worse quasi-experiments.

6. Should a standard set of design coding categories be used in C2 systematic reviews?

Proposal: C2 should develop a standard set of codes to be used in reviews for purposes of coding study design and related features. Some such codes are presented and discussed in the Appendix to this Brief.

7. Of these design codes, which should systematic reviewers be required to code and which should be recommended (optional)?

Proposal: Specific recommendations are made for which codes are required and which are optional in the text of the Brief.

8. Should C2 consider grading the quality of either randomized or nonrandomized experiments, or both, as to level of defensibility of results?

Proposal: Aspects of methodology that are related to the validity of a study's conclusions should be assessed individually rather than being summed into a total quality score.

* Full brief available at:

<http://www.missouri.edu/~c2method/>

Statistical Analysis Policy Brief

*** Executive Summary**

By Betsy Becker, Larry Hedges and Therese Pigott
Systematic reviews of the effects of interventions and relations among variables often rely on statistical summaries of the results of primary studies. Because Campbell Collaboration (C2) systematic reviewers are

likely to face a variety of statistical issues in conducting reviews, this policy brief attempts to:

1. Identify the key issues that are confronted by C2 systematic reviewers who want to synthesize the results of studies statistically,
2. Outline possible ways that statistical procedures might be used, and
3. Provide examples of how these methods might be used.

In this brief we address six key issues concerning statistical analysis, and make proposals for C2 policies for each. A summary of the issues and our proposals follows.

1. When conducting a research synthesis, is it ever appropriate for a C2 reviewer to do a review *without* statistically integrating the results of studies? If yes, what are the characteristics of the literature that make this permissible?

Proposal: Study findings should be represented as effect sizes (i.e., indices of treatment impact or relationship strength) in C2 reviews whenever the studies being summarized present quantitative findings. Statistical integration should only be used in any C2 review (or any part of a C2 review) where a summary conclusion from at least two studies is desired, the studies and effect sizes are sufficiently similar to justify integration, and the number of studies is sufficient to support the analysis used in that statistical integration.

2. When statistical integration is used in a C2 review, are there certain statistical procedures that should routinely be carried out? If so, what are they?

Proposal: Statistical summaries of average effects and variation in effects should be computed (and reported) for fixed-effects, random-effects or both types of analyses. The specific statistics used will depend on whether the review is aimed at (a) estimating a mean effect across studies, (b) examining the variation in effect-size estimates across studies, or (c) fitting a model of effect-size variation.

3. When systematic reviews retrieve and code characteristics of statistical analyses, what characteristics of the analyses should routinely be coded, and, if possible, examined for their impact on the outcomes of studies?

Proposal: Reviewers should code (a) characteristics of the statistical analyses used in the primary study and (b) details about the computations used for the effect size derived from that study. C2 takes the position that it is important to document specific statistical procedures and methods for computing an effect size just as it is important to code study design differences. Coding of statistical procedures allows the use of sensitivity analyses as a method for examining how differences in statistical methods of studies or effect-size computations influence the results of the systematic review.

4. Should multiple (nonindependent) effect-size estimates from the same study ever be used in a C2 synthesis?

Proposal: Reviewers should not ignore dependence among study outcomes. They should use *some* procedure to deal with dependence, describing and giving a justification for that procedure, even if it is ad hoc. Simple approaches such as dropping or combining outcomes or using sensitivity analyses may make sense if the amount of dependent data is small. More sophisticated analyses may be needed if multivariate data are prevalent in the review. In such cases the reviewer must assess the similarity of studies and availability of reliable information on the extent of dependence.

5. Should C2 have a role in advancing cross-design synthesis methods (e.g., propensity scoring and alternatives)? What must be considered if/when reviewers combine estimates of effect from randomized trials with estimates of effect based on other designs, such as surveillance systems, passive observational studies, etc?

Proposal: In some syntheses results from subsets of studies in the synthesis will not be comparable. In such cases reviewers should not summarize across the designs, but rather should report both sets of results separately. In other cases where effects are more comparable, the reviewer may wish to summarize

across designs as well as provide separate results by design. Assumptions underlying such comparisons should be made explicit, and the reviewer should critically examine the data for the possibility of design-related differences in effects. Further, when such comparisons are made, the type of design should be tested as a moderator variable and separate results should be reported.

Furthermore, while the primary focus of C2 is on matters directly related to research cumulation, the study and careful application of methods of cross-design synthesis is consistent with the goals of C2.

6. What should be the role of C2's Social, Psychological, Educational and Criminological Trials Register (SPECTR) in supporting or informing the statistical research that might be done in the Campbell context?

Proposal: The Steering Committee should endorse the use of SPECTR for research on normative methodological and reporting practice in relevant research domains, improving information for imputation in effect-size computation, and studying associations between synthesis methods and results.

* Full brief available at:
<http://www.missouri.edu/~c2method/>

List of Protocols Critiqued (Jan 1 - July 31, 2002)

Preventing Repeat Victimization: A Systematic Review
Design replication studies of interventions in education, training, and employment service
Cost-benefits of sentencing
Curfews for preventing juvenile delinquency and victimization
Systematic Review of the Impact of Welfare Reform on Family Structure
Early parent training to prevent disruptive behavior problems and delinquency in children
The effectiveness of school-based problem-solving interventions on aggressive behavior: A Campbell Collaboration systematic review



Thanks are extended to the following methods group members for providing critiques:

Tom Cook	Larry Hedges
Mark Lipsey	Gilbert Ramirez
Will Shadish	Darcy Strouse
Helen Thomas	Jeff Valentine
David Wilson	



CONDENSED AGENDA*
THE 1st CAMPBELL COLLABORATION METHODS GROUP CONFERENCE

September 17-19, 2002

The Admiral Fell Inn Hotel
Baltimore, Maryland, USA

*To view full agenda, including presenter's names, please visit:
<http://www.missouri.edu/~whyten/Agenda.htm>

DAY 1: TUESDAY, SEPTEMBER 17

Afternoon Sessions	2:00 - 3:15	Meeting of the Center for Research Synthesis Board of Advisors (Open)
	3:30 - 5:00	What do I need to know to help others do C2 reviews?
	5:00 - 6:00	Reception
		Dinner (On your own or at Camden Yards)

DAY 2: WEDNESDAY, SEPTEMBER 18

Morning Sessions	8:00 - 9:00	Breakfast
	9:00 - 9:30	Welcome and Charge
	9:30 - 10:30	Does research design really matter?
	10:45- 11:45	Can a study's quality be expressed as a single score?
	Noon - 1:00	Lunch



Afternoon Sessions	1:00 - 2:00	What is the role of qualitative research in C2 systematic reviews?
	2:15 - 3:15	Given publication bias, missing data, and coder unreliability, does C2 data really resemble population data?
	3:30 - 4:30	What methodological work is being done by some other organizations?
	6:00 - 8:30	Dinner

DAY 3: THURSDAY, SEPTEMBER 19

Morning Sessions	8:00 - 9:00	Breakfast
	9:00 - 10:00	When is meta-analysis inappropriate?
	10:15- 11:15	What will the next generation of synthesis methods look like?
	11:15- Noon	Farewell and Charge

The Campbell Collaboration Methods Group
Center for Research Synthesis Methodology
University of Missouri-Columbia
217 McAlester Hall, Columbia, Missouri 65211
Tel: 573-882-5824 Fax: 573-884-4537 email: crsm@missouri.edu