Author's response to reviews

**Title:** Systematic Reviews: A Cross-Sectional Study of Location and Citation Counts

**Authors:**
- Dr Victor M Montori (montov@mcmaster.ca)
- Nancy L Wilczynski (wilczyn@mcmaster.ca)
- Douglas Morgan (dmorgan@mcmaster.ca)
- Prof Brian Haynes (bhaynes@mcmaster.ca)

**Version:** 2 **Date:** 17 Sep 2003

PDF covering letter
POINT-BY-POINT RESPONSE TO REVIEWERS COMMENTS

We were fortunate to get insightful reviews that included useful suggestions to improve the quality of our submission. We are indeed grateful. Following, please find our point-by-point response to their comments and suggestions for revisions. Because both reviewers raised important and similar concerns about our brief methods section we have expanded this section providing the details of that expansion only as we responded to the first reviewer.

Reviewer: Michael Callaham

In response to discretionary revisions

1. Apparent contradiction between the spectrum of patients in trials and the application to a “wide array of patients” of the results of systematic reviews (p 3, para 2).
   >> We have clarified the sentence to indicate the following: “Clinicians can apply the results of meta-analyses to a wide array of patients -- certainly wider than those included in each of the primary studies -- that do not differ importantly from those enrolled in the primary studies.” This is now consistent with current doctrine that seeks to apply the results of studies to patients unless there are compelling reasons not to do so.

In response to minor compulsory revisions

2. No page numbers – we have added page numbers.
3. Reference missing to statement about the “prevalence” of systematic reviews and their authors (p 4 para 1). We have added reference to our paper published in the Mayo Clinic Proceedings were we estimated the total number of systematic reviews published in the period 1980-2000 to exceed 5000. Since the Cochrane Collaboration has published just in excess of 1000 reviews, a conservative estimate is that other authors collectively have produced 3-5 times that number.

In response to major compulsory revisions

Both reviewers demanded more details about the selection of journals, the classification scheme for papers, and the reliability training and testing. In response to these requests, we have rewritten the METHODS section as follows (pages 5-9):

Methods
The Hedges Team of the Health Information Research Unit (HIRU) at McMaster University is conducting an expansion and update of our 1991 work on search filters or ‘hedges’ to aid clinicians, researchers, and policymakers harness high-quality and relevant information from MEDLINE.[14] We planned to conduct the present work within the larger context of the Hedges Project prior to the onset of data collection and analyses.

Journal selection
The editorial group at HIRU prepares 4 evidence-based medical journals, the ACP Journal Club, Evidence-based Medicine, Evidence-based Nursing, and Evidence-based Mental Health. These journals help keep healthcare providers up-to-date. To produce these secondary journals, the editorial staff has identified 170 journals that regularly publish clinically-relevant research in the areas of focus of these evidence-based journals (i.e., general internal medicine, family practice, nursing, and mental health). We evaluated journals for inclusion into this set that have the highest Science Citation Index Impact Factors in each field and journals that clinicians and librarians that collaborate with HIRU recommended based on their perceived yield of important papers. The editorial staff then monitors the
yield of original studies and reviews of scientific merit and clinical relevance (criteria below) for each of these journals, to determine if we should add new ones, keep those on the list, or remove low-yield journals from the list.

**Study identification and classification**

On an ongoing basis, 6 research associates review each of these journals and apply methodological criteria to each item to determine if the article is eligible for inclusion in the evidence-based publications. For the purpose of the Hedges Project (i.e., to develop search filters), we expanded the data collection effort and began intensive training and calibration of the research staff in 1999. In this manuscript, we report the $\kappa$ statistic measuring chance-adjusted agreement between the 6 research assistants for each classification procedure.

We reported the training and calibration process in detail elsewhere.[15] Briefly, prior to the first inter-rater reliability test research staff met to develop the data collection form, and to develop a document outlining the coding instructions and category definitions using examples from the 1999 literature. Meetings involving the research staff revealed differences in interpretation of the definitions (early $\kappa$ were as low as 0.54). Intensive discussion periods and practice sessions using actual articles were used to hone definitions and thus remove ambiguities ($\text{goal } \kappa > 0.8$). The 6 research associates received the same articles packaged with the data collection form and the instructions document (this document is available from the authors on request) and each independently and blindly reviewed each article and recorded their classification in the data collection forms. We conducted three reliability tests during 1999. We conducted the fourth and final inter-rater reliability test approximately 14 months after the process commenced using a sample of 72 articles randomly selected across the 170 journal titles. In calculating the $\kappa$ statistic for methodological rigor raters had to agree on the purpose category for the item to be included in the calculation (Table 1 describes the purpose categories and the criteria for methodological rigor for each one). We analyzed data using PC-agree (software code written by Richard Cook; maintained by Stephen Walter, McMaster University, Hamilton, Canada).

For the purposes of the Hedges Project, we defined **review** as any full text article that was bannered as a review, overview, or meta-analysis in the title or in a section heading, or that indicated in the text that the intention of the authors was to review or summarize the literature on a particular topic.[15] To be considered a **systematic review** the authors had to clearly state the clinical topic of the review, how the evidence was retrieved and from what sources (i.e., name the databases), and provide explicit inclusion and exclusion criteria. The absence of any one of these 3 characteristics would classify a review as a **narrative review**. The inter-rater agreement for this classification was almost perfect ($\kappa = 0.92$, 95% confidence interval 0.89-0.95).

Then, we classified all reviews by whether they were concerned with the understanding of healthcare in humans. Examples of studies that would not have a direct effect on patients or participants (and, thus, are excluded from analysis) include studies that describe the normal development of people; basic science; gender and equality studies in the health profession; or studies looking at research methodology issues. The inter-rater agreement for this classification was almost perfect ($\kappa = 0.87$, 95% confidence interval 0.89-0.96).

A third level of classification placed reviews in purpose categories (i.e., what question(s) are the investigators addressing) that we defined for the Hedges Project and included etiology (causation and safety), prognosis, diagnosis, treatment, economics, clinical prediction guides, and qualitative (Table 1).[15] The inter-rater agreement for this classification was almost perfect ($\kappa = 0.81$, 95% confidence interval 0.79-0.84).

A fourth level of classification graded reviews for methodological rigor placing them in pass and fail categories. To pass, the review should include a statement of the clinical topic (i.e., a focused review question); explicit statements of the inclusion and exclusion criteria; a description of the search strategy and study sources (i.e., a list of the databases); and at least 1 included study that satisfied methodological rigor criteria for the purpose category (Table 1). For example, reviews of treatment interventions had to have at least one study with random allocation of participants to comparison groups and assessment of at least one clinical outcome. All narrative reviews were included in the fail category. We refer to systematic reviews that passed this methodological rigor evaluation as **rigorous systematic reviews**. Again, the inter-rater agreement for this classification was almost perfect ($\kappa = 0.89$, 95% confidence interval 0.78-0.99).

For this report, we retrieved data on review articles including a complete bibliographic citation (including journal title), the pass/fail methodological grade, and the review type (narrative or systematic review).

**Impact factor and citation counts**

To collect impact factor data for all 170 journals in the database we used the ISI Journal Citation Reports (http://isiknowledge.com). We also queried the ISI Web of Science database to ascertain, as of February 2003, the number of citations to each one of the reviews in an arbitrary subset of 5 journals that published the most systematic reviews and are indexed journals in the ISI database.
Data analysis
Data were arrayed in frequency tables. We conducted nonparametric univariate analysis (Kruskal-Wallis) to assess the relationship between the number of citations and the type of review. We assessed the correlation between journal impact factor and citation counts. Then, using multiple linear regression, we determined the ability of the independent variables - methodological quality of the reviews and journal source - to predict the dependent variable, the number of citations (after log transformation). Thus, this analysis was stratified by journal to adjust not only for impact factor, but also for other journal-specific factors not captured by this measure.

4. Citation query –the papers included in the database were published in 2000. Thus the reviewer’s concern about immediacy and time bias is not as relevant as when one considers, as he did, that we were looking at a very brief time frame (2002 to February of 2003). As the reviewer comments we are including up to 3 years of follow-up for citations to accrue for papers published during 2000.

5. P 9 para 2: We now describe our choice of 5 top journals publishing most systematic reviews as arbitrary. To some extent this practical choice was informed by the paucity in general of systematic reviews and the need to include journals that published both types of reviews (narrative and systematic). In response to other reviewers’ comments we now highlight this choice in the discussion section as a limit to the pertinence of our findings.

6. In response to the reviewer’s comments we downplay the importance of the VERY weak association between impact factor and citations. In response to the other reviewer’s comments we have a note pointing this out at the bottom of page 12: “Our results are consistent with another study that also documented a weak association between journal impact factor and the methodological quality of published studies.[16]”.

7. The reviewer requests further discussion about the results of citation counts, the meaning of our study in terms of editorial practices, and the explanations for maldistribution. Also the reviewer brings up a “limitation” of our study in that we did not measure the clinical impact of systematic review. We have added paragraphs (p 13, para 2 and 3) to the discussion section with limited speculation about these issues that we did not set out to study:

We can only speculate about the causes of the maldistribution of rigorous systematic reviews among a few journals, since exploration of such causes was not an objective of our study. Journal policy and author preferences may contribute to this maldistribution. The lack of systematic reviews and meta-analysis published in the New England Journal of Medicine is evidence of the effect of journal policy. Other journals, such as JAMA, Lancet, BMJ, and Annals of Internal Medicine, have published articles about systematic review methodology and reporting, and enthusiastically publish rigorous reviews of clinical importance. Authors of such reviews, naturally, may prefer to submit their reviews to journals with large circulation and impact. The relative contributions of these sources to the observed maldistribution constitute hypotheses that remain to be tested.

Given that our research design does not support causal inferences, it is unwise to derive recommendations to journal editors based on our findings. We think that journal editors interested in publishing rigorous research should prefer systematic reviews over narrative reviews. Furthermore, our research generates the hypothesis that a choice of systematic over narrative reviews may contribute to increase a journal’s impact factor. However, editors of traditional journals have other competing priorities that rely less on citation counts and more on popularity (e.g., attract and maintain readership, attract advertisement and generate revenue) which may direct their choice of reviews to publish (i.e., if they perceive narrative reviews as easier to read and more attractive to their readership than systematic reviews and meta-analyses).

And we added a statement to the limitations paragraph (p 12, para 2): “We did not set out to evaluate the impact of these reviews on clinical practice.”
8. We share the reviewers concern about the paucity of high quality systematic reviews and share his curiosity about the readership of these reviews, including the Cochrane reviews. Again, we did not set out to understand this and we cannot comment further on these issues.

Again, we are indebted to this reviewer’s astute observations and we hope to have responded satisfactorily to his helpful comments and insightful suggestions.

Reviewer: Lisa Bero

In response to discretionary revisions

1. We have defined in the abstract what we mean by ‘review’ in this study as follows: “We defined ‘review’ as any full text article that was bannered as a review, overview, or meta-analysis in the title or in a section heading, or that indicated in the text that the intention of the authors was to review or summarize the literature on a particular topic.”

In response to minor compulsory revisions

1. We appreciate the suggestion to point out the difficulty in finding systematic reviews within MEDLINE. We have now added several references that speak to the difficulty in finding systematic reviews within MEDLINE (p4, para 1).
2. We have changed the phrase in question to link clinical importance with frequent citation, the hypothesis of our work (p4, para 2): “If systematic reviews in fact represent the best available evidence, they are likely to have great clinical importance. It follows that they would be cited often in the literature.”
3. We have added an important reference to highlight the limitation of impact factors (ref 13, p 4, para 2).
4. We decided to define systematic review and narrative review later in the Methods section because it flows better this way. This is particularly the case now that in response to both reviewers we have expanded the Methods section regarding article classification. We did not change the manuscript in response to this suggestion.

In response to major compulsory revisions

Points 1, 2, 3, 4, and 5 are now responded in the expanded description in the Methods section (p 5-9). We have also added a reference to the instruction manuals that the research assistants used to judge the methodological quality of reviews and identify systematic reviews. Furthermore, we have explicitly stated that they were looking for statements about the data sources that were consistent with a systematic and thorough approach (p 7 para 2): “To be considered a systematic review, the authors had to clearly state the clinical topic of the review, how the evidence was retrieved and from what sources (i.e., name the databases), and provide explicit inclusion and exclusion criteria.”

6. The sample for analysis of citation counts only included the 5 journals. We have now added frequent statements about the limitations of this selection and the conclusions derived from these analyses, including an explicit statement added to the abstract: “We
obtained citation counts for review articles in the 5 journals that published the most systematic reviews.” P 9 para 1: “We also queried the ISI Web of Science database to ascertain, as of February 2003, the number of citations to each one of the reviews in an arbitrary subset of 5 journals that published the most systematic reviews and are indexed journals in the ISI database.” P12 para 1: “Among the 5 journals publishing most systematic reviews, and after stratifying by journal, the type of review (rigorous systematic vs. narrative) was independently associated with the number of citations.”

7. We appreciate the reviewer indicating the results of the study linking journal characteristics to study quality. We have now quoted it in the discussion section (p12 para 1): “Our results are consistent with another study that also documented a weak association between journal impact factor and the methodological quality of published studies. [16]” with ref 16 being Lee et al. JAMA 2002; 287: 2805-2808. We have also incorporated it in the discussion of journal policies and systematic reviews (p 13 para 2, 3).

We are again indebted to the contributions that these two experts have made to the quality of our paper. We sincerely hope to have justified their efforts by responding to them thoughtfully and thoroughly. We think that, as a result, we have a more detailed and, perhaps, more useful contribution. We appreciate their repeated expressions of importance and relevance that they generously attribute to our work.