Mapping the Cochrane evidence for decision making in health care

Regina P. El Dib PhD,[1] Álvaro N. Atallah MD PhD[2] and Regis B. Andriolo[3]

[1]Speech, Language and Hearing Pathologist and Scientific Research Assistant of The Brazilian Cochrane Centre, PhD Postgraduate student in Internal and Therapeutic Medicine at Universidade Federal de São Paulo, The Brazilian Cochrane Centre – Universidade Federal de São Paulo, São Paulo, Brazil
[2]Director of The Brazilian Cochrane Centre, Full Professor of Urgency Medicine and Evidence-Based Medicine at Universidade Federal de São Paulo, Brazil., The Brazilian Cochrane Centre – Universidade Federal de São Paulo, São Paulo, Brazil
[3]Biologist and Scientific Research Assistant of The Brazilian Cochrane Centre, Postgraduate student (master) in Internal and Therapeutic Medicine at Universidade Federal de São Paulo, The Brazilian Cochrane Centre – Universidade Federal de São Paulo, São Paulo, Brazil

Keywords
clinical medicine, clinical trials, evidence-based medicine, meta-analysis, research, review literature

Correspondence
Álvaro Nagib Atallah
The Brazilian Cochrane Centre
Rua Pedro de Toledo, 598 – Vl. Clementino
São Paulo – SP 04039-001
Brazil
E-mail: atallahmbe@uol.com.br

Accepted for publication: 3 April 2007

Abstract

Rationale and aim Over the past 12 years, thousands of authors working with the Cochrane Collaboration around the world have produced systematic reviews to reduce uncertainty in health care decision making. We evaluated the conclusions from Cochrane systematic reviews of randomized controlled trials in terms of their recommendations for clinical practice and research.

Methods In our cross-sectional study of systematic reviews published in the Cochrane Library, we randomly selected and analysed completed systematic reviews published across all 50 Cochrane Collaborative Review Groups.

Results We analysed 1016 completed systematic reviews. Of these, 44% concluded that the interventions studied were likely to be beneficial, of which 1% recommended no further research and 43% recommended additional research. Also, 7% of the reviews concluded that the interventions were likely to be harmful, of which 2% did not recommend further studies and 5% recommended additional studies. In total, 49% of the reviews reported that the evidence did not support either benefit or harm, of which 1% did not recommend further studies and 48% recommended additional studies. Overall, 96% of the reviews recommended further research.

Conclusions Cochrane systematic reviews were about evenly split between those in which the authors concluded that at least one of the interventions was beneficial and those in which the evidence neither supported nor refuted the intervention tested. The Cochrane Collaboration needs to include clinical trial protocol summaries with a study design optimized to answer the relevant research questions.

Introduction

The aims of the Cochrane Collaboration are to make readily available up-to-date, accurate information about the effects of health care, to produce and disseminate systematic reviews of health care interventions, and to promote the search for evidence in the form of clinical trials and other intervention studies.

Systematic reviews, however, are criticized for frequently offering inconsistent evidences and absence of straightforward recommendations [1]. Their value seems to be depreciated when the conclusions are uncertain or based on less than the highest grading of evidence [2]. Moreover, both readers and authors of systematic reviews usually, but erroneously, use to conceive 'absence of effect' or 'absence of differences between treatments' instead of 'there is insufficient evidence either to support or to refute' [3].

We analysed a random sample of Cochrane systematic reviews to evaluate the percentage of Cochrane reviews classified by their authors as showing either a beneficial or harmful intervention, or those reviews with insufficient evidence to make a judgement whether the treatment is beneficial or harmful. We recorded whether further research was recommended and we also noted the number of studies and meta-analyses performed in each review.

Methods

In this cross-sectional study, we selected systematic reviews from the Cochrane Library issue 4, 2004, excluding, withdrawn reviews and protocols.

We randomly selected up to 23 systematic reviews from all 50 Cochrane Collaborative Review Groups. First, we allocated the

50 Review Groups to one of two investigators. Then, we printed and placed in a bag all review titles from each Group. The allocated investigator selected 23 titles at random from each bag. Where there were less than 23 titles in one bag all the titles were selected.

We analysed the conclusions of each selected review and allocated the conclusions to one of six categories describing the implications for practice and research: (1) beneficial interventions, for which the authors did not recommend further research (treatment is more beneficial/effective than control for the primary outcome); (2) interventions likely to be beneficial, for which the authors recommended further research (treatment may have a positive effect, but a major unresolved methodology issue, such as all studies being very low quality, or findings based on only one study, precluded making a definitive statement); (3) harmful interventions and the authors did not recommend further research (treatment does more harm than good); (4) interventions likely to be harmful, for which the authors did suggest more research (treatment may have a negative effect, but a major unresolved issue, such as all studies being very low quality, or findings based on only one study, precluded making a definitive statement [4]; (5) insufficient evidence for which the authors did not suggest further research (there was insufficient evidence to assess effectiveness but the clinical question does not stilly further research); and (6) insufficient evidence, and the authors asked for further research (there is insufficient evidence to assess effectiveness).

Data extraction

We allocated categories based on the reviewer's conclusions; we scrutinised other sections of the reviews when this was necessary to clarify the understanding and consistency of the statements. If there was indecision about the judgements to be made, the investigators met to reach a consensus.

Sample size

To estimate the sample size, we assumed that, across all the systematic reviews analysed, 30% would show insufficient evidence. An error of 3% within a 95% confidence interval was accepted. According to these assumptions, it would be necessary to analyse approximately 900 systematic reviews, according to the following formula:

$$E = Z\sqrt{pq/n}$$

where, E is the sample error; Z is the constant relative to 95% confidence interval (1.96); p corresponds to the expected proportion of systematic reviews showing insufficient evidence; q is the complementary of p regarding the totality of systematic reviews (1 – p).

Statistical analysis

We randomized 100 systematic reviews for consistency checking by two investigators, and calculated the interobserver agreement rates using the Kappa test in relation to the main outcomes.

The occurrence of each 'implications for practice and research' category was represented as a natural number, percentages and 95% confidence intervals for all the systematic reviews analysed. To calculate the 95% confidence intervals, we used the finite correction factor $(N - n)/(N - 1)$, considering that $(n \cdot N - 1 \geq 0.05)$ [5], where 'n' is the sample and 'N' is the total of systematic reviews issue 4, 2004.

We expressed the number of meta-analyses performed and studies included as totals, means and standard deviations; and ranges, medians and modes.

Results

We analysed 1016 (46%) of the completed systematic reviews published in the Cochrane Library, issue 4, 2004, in terms of their authors' conclusions relating to implications for practice and research.

The interobserver concordance for reviews classified as presenting insufficient evidence to support or refute indication of the intervention of interest was modest (kappa coefficient: 0.35). In the same way, for reviews classified as presenting insufficient evidence to support or refute indication of the intervention of interest, in which the authors did not recommend future research, the interobserver concordance was low, as expected because of the small percentage with this outcome (kappa coefficient: –0.01) (Table 2).

The statistics relating to the numbers of studies included and numbers of meta-analyses found in the reviews, and the totals, are shown in Table 1. The reviews included a median of eight randomized trials (range: 0–292) and two meta-analyses (range: 0–177).

The main outcomes were as follows:

1.38% – beneficial interventions (and the authors did not recommend further research);

43.01% – interventions likely to be beneficial (and the authors did recommend further research);

1.67% – harmful interventions (and the authors did not recommend further research);

5.12% – interventions likely to be harmful (and the authors did suggest more research);

0.98% – insufficient evidence for clinical practice (and the authors did not suggest further research);

47.83% – insufficient evidence for clinical practice (and the authors did ask for further research).

In total, 95.96% of the reviews recommended further studies.

The percentages and respective 95% confidence interval limits relating to the six categories of conclusions on 'implications for practice and research' are shown in Fig. 1.

The levels of interobserver concordance relating to each category of conclusion, as expressed by kappa coefficients, are shown in Table 2.

Table 1 Statistics for the numbers of studies included and meta-analyses performed in 1016 systematic reviews

Statistics	Studies included	Number of meta-analyses
Mean (SD)	13.61 (19.92)	6.36 (13.25)
Range	0–292	0–177
Median	8	2
Mode	2	0
Total	13 830	6461

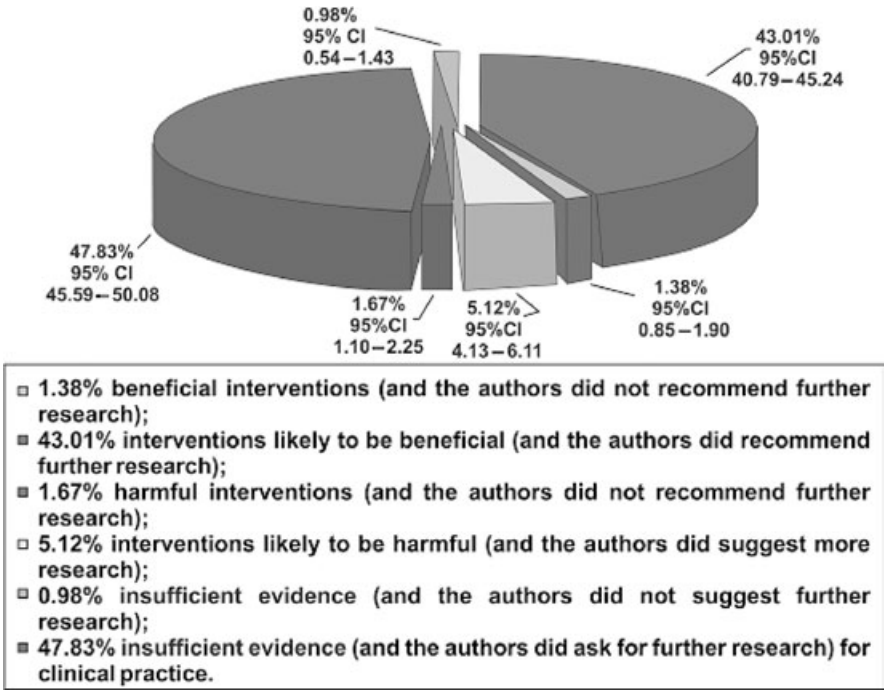

Figure 1 Graph showing percentage of reviews in each category of conclusion regarding implications for practice and research.

Table 2 Interobserver concordance relating to each category of conclusion, as expressed by kappa coefficients

Category of conclusion	A1	A2	B1	B2	C1	C2
Kappa coefficient	0.80	0.65	0.71	1.00	0.84	–0.01

A1, beneficial interventions (and the authors did not recommend further research); A2, interventions likely to be beneficial (and the authors recommend more research); B1, harmful interventions (and the authors did not recommend further research); B2, interventions likely to be harmful (and the authors did suggest more research); C1, insufficient evidence (and the authors did not recommend more research); C2, insufficient evidence (and the authors ask for further research).

Discussion

Good clinical research aims to reduce uncertainty in order to help to make uniform clinical decisions. This study has found that the majority of Cochrane Reviews highlight the absence or poor evidence around the questions on health care that has been covered by them. We found that Cochrane Systematic Reviews were about evenly split between those in which the authors concluded that intervention was beneficial and those in which the evidence neither supported nor refuted the intervention tested.

The random selection, large sample size and independent dual data extraction are strong features of the present study. These help to ensure both that the results are applicable to all the reviews in The Cochrane Database of Systematic Reviews and also that our findings are reproducible. The large sample size helps to ensure the precision of our estimates and reduce the risk of random errors [6]. In this way, we believe that the results have importance not only with regard to their implications for clinical practice, but also with regard to scientific research.

Around half of the reviews analysed in this study (47.83%) did not offer enough evidence for clinical practice, and the authors asked for further research. In only 0.98% of the 1016 reviews did the authors find insufficient evidence to support or refute the indication, yet did not consider it necessary to undertake further investigation, probably because there was a limited rationale for pursuing such questions. Even when the reviews found evidence to support the intervention for use in clinical practice, for a large percentage of these the authors considered this evidence was limited, and took the view that further research would be worthwhile (43.01%). The same type of uncertainty was also present in reviews in which the intervention of interest were classified as likely to be harmful, but the authors considered that the matter deserved more research (5.12%). In only 1.67% of the 1016 systematic reviews analysed, in which the evidence suggested that the interventions of interest were harmful, did the authors of the reviews discourage further research (Fig. 1). Overall, in 95.96% of all the reviews analysed, the authors recommended more research. This important finding was similar to that found by Vlassov (2004) who investigated how frequently recommendations such as 'more research is needed' were made, and how these related to the results from the reviews [7]. This author evaluated 100 Cochrane reviews and found that 93% of them concluded by making recommendations of this type.

Even having systematic reviews as Level I of evidence for decision making with regard to therapy in the field of health care, it was observed that most authors of systematic reviews concluded that there was insufficient evidence to answer the questions around therapeutic strategies for treatment and prevention of diseases. This is perhaps because the primary studies do not have minimal methodological quality to be included or, if so, did not generate consistent data, or data suitable for meta-analysis. Both conditions are clearly demonstrated in Table 1, from the mode data (mode = 2) for studies included in the systematic reviews and absence of meta-analysis for the majority of them (mode = 0). It can be seen from this that, for most of the 1016 systematic

reviews, it was only possible to identify two clinical trials that satisfied the inclusion criteria.

To raise the awareness of the need for having higher-quality primary studies, Cochrane Reviews could include protocols for relevant clinical trials. Such an initiative would help to increase the quality of primary studies for inclusion in systematic reviews, thereby reducing the widespread uncertainties that still exist in medical science.

Acknowledgements

We would like to thank Ian Chalmers and Fiona Milne for their help during the corrections and review of this article.

References

1. Browman, G. P. (1999) Essence of evidence-based medicine: a case report. *Journal of Clinical Oncology*, 17, 1969–1973.
2. Egger, M., Smith, G. D. & Phillips, N. A. (1997) Meta-analysis: principles and procedures. *British Medical Journal*, 315, 1533–1537.
3. Alderson, P. & Chalmers, I. (2003) Survey of claims of no effect in abstracts of Cochrane reviews. *British Medical Journal*, 326 (7387), 475.
4. Ezzo, J., Bausell, B., Moerman, D. E., *et al.* (2001) Reviewing the reviews. How strong is the evidence? How clear are the conclusions? *International Journal of Technology Assessment in Health Care*, 17 (4), 457–466.
5. Berquó, E. S., Souza, J. M. P. & Gotlieb, S. L. D. (1980) Bioestatística, 5th edn. São Paulo: Editora Pedagógica e Universitária Ltda.
6. Clarke, D. D. (1983) Sample-size nomograms. *Annals of Internal Medicine*, 99 (6), 875–876.
7. Vlassov, V. (2004) Further research is needed? *Cochrane Collaboration*. Colloquia Abstracts. Ottawa, 2004. p. 155. Available from: http://www.cochrane.org/colloquia/abstracts/ottawa/P-155.htm (accessed 26 July 2007)