
Editorial

Narrative or systematic reviews: can we be more 'evidence-based'?

Although *Reviews in Clinical Gerontology* aims to provide comprehensive, authoritative and up-to-date appraisals of the literature on the main topics of interest to those working for the health of elderly people, the vast majority of our authors have used the traditional 'narrative' approach to literature review. Indeed, the article by MacKnight and Rockwood in the Rehabilitation section of this issue is the first to make substantial use of the systematic, 'meta-analytic' approach favoured by the gurus of evidence-based medicine. Are we hide-bound by tradition, merely lazy, or is meta-analysis not all it is cracked up to be?

The techniques of systematic review have mainly been used to assess the effectiveness of treatments evaluated in controlled trials. In order to reach valid conclusions based on a 'weighted sum' of the results of several studies, at least four conditions must be satisfied: (1) *all* available evidence must be considered; (2) only *unbiased* studies should be included in the analysis; (3) the studies should address the *same question*; and (4) the results must be presented in a *compatible format*. When these rules are observed, meta-analyses of well designed but small trials usually predict the findings of subsequent larger studies fairly well.¹ On the other hand, incomplete or flawed reviews can produce misleading conclusions when compared to a more comprehensive meta-analysis.^{2,3}

The first condition is often extremely difficult to meet. The biomedical literature is so vast and

even specialist areas expand so fast that it is literally impossible for any clinician or scientist to keep up to date simply by regular reading. Even in skilled hands, computerized searches using databases such as MEDLINE pick up only about half of the relevant published studies,⁴ and many are not published at all. There is no escaping the fact that studies which are harder to find, less often cited, or unpublished are more likely to be 'negative'.⁵

Nowadays it is recognized that the only way to make an unbiased comparison of different treatments is to use random allocation, so most meta-analyses are confined to randomized controlled trials. Nevertheless, the quality of such trials can vary widely and there are many other ways of introducing bias, especially if not all randomized patients are included in the analysis or if the outcome assessor is not blind to treatment allocation (a particular problem in many rehabilitation studies⁶).

Thirdly, although different trials may look as if they are addressing the same general question, there may be subtle but important differences. This could be one explanation of the much discussed discrepancy between meta-analysis of trials of intravenous magnesium in acute myocardial infarction and the results of the ISIS-4 megatrial,⁷⁻⁸ where the treatment was mostly started somewhat later than in the smaller trials. Most meta-analyses include a test for homogeneity of the results of the component studies. If there are wide differences it may not be

wise to combine them.⁹

Finally, the studies in a meta-analysis must not only ask the same general question but also give their answer in the same general form. This is not much of a problem where simple outcomes like death are the main concern, but morbidity and disability are often more important in studies involving elderly people and there are any number of different ways of measuring these things. This problem is particularly acute in rehabilitation, with its plethora of clinical assessment scales.¹⁰ Since the concepts used in rehabilitation are much more subtle and varied than the simple measures used in most other branches of medicine, it is tempting to design specific measures for each new study, and few investigators have managed to resist this temptation.

The result is a veritable Babel, where our efforts to reach the heights of scientific enlightenment by building systematically on the work of others are confounded by the lack of a common language or understanding. MacKnight and Rockwood have explored one small corner of this polyglot tumult, using some of the techniques of systematic review to identify many similar, but different, clinical instruments for measuring balance, gait and mobility in elderly people.

The results of controlled trials can still be combined even if they all use different clinical measures. Within each trial the same instruments are used to assess outcome in all patients, so the comparison between treatment groups can be described in terms of a standard measure, the 'effect-size'. Clearly it is much harder to pool the results of studies other than controlled trials, but the purpose of MacKnight and Rockwood's review was not to do this but to apply standardized criteria to each of the studies in an attempt to identify the most useful clinical measures. How well did they succeed in this?

First, although their search strategy was well planned and clearly described, many readers will probably know of apparently suitable measures which have been omitted from the review. These include some well tested and fairly widely used instruments such as the Rivermead Motor Assessment and the Rivermead Mobility Index.^{11,12} This probably reflects the North American bias of databases such as MEDLINE which do not index some important British- and European-based journals.

Secondly, although the criteria for evaluating clinical measures used in the review were also clearly described, it is still not easy to pick out the 'best buys'. Nowadays some attempt is usually made to test the reliability of new clinical assessments, but sensitivity and responsiveness to change are less often documented. The term 'validity' is often used almost as a slogan, without a clear idea of how or whether it should be tested in practice. For example, there can be little doubt that the 1001 different ADL scales that assess the same items in slightly different ways are measuring the same basic constructs, or that the scores on each scale will be strongly intercorrelated, so is it always helpful to 'validate' a new scale by comparing it with the Barthel or the Katz? Similarly, the 'sensibility' of an assessment is clearly an important property, but can it really be measured objectively?

None of these practical problems detract from the major theoretical advantages of systematic literature review, but they should make us wary of accepting the conclusions of any kind of retrospective 'meta-analysis' as definitive. As with most scientific studies, the real value of systematic review is less in the answers it provides than in the questions it raises. For instance, meta-analysis has shown us that 'stroke unit care' saves lives and reduces the risk of long-term institutionalization,¹³ but it has not told us what it is about stroke unit care that is beneficial. The next step must therefore be to address this issue in a prospective collaboration, using a standardized approach to describing and measuring the contents of the stroke care 'black box'.

Similarly, MacKnight and Rockwood's review demonstrates the need to work together to bring some order to the chaos of rehabilitation assessment, to agree on which measures are best for which situations, which should be dropped, and which areas require new instruments to be developed. The purpose of literature review is not just to reach a consensus view based on previous work, but to identify major areas of uncertainty for future study. This will allow us to tackle these burning questions in an organized way, creating a collaborative framework to enable future reviewers to find all the relevant studies and pool the results in a much more effective way than is now possible.¹⁴ We will then have moved beyond retrospective 'meta-analysis' into

the era of organized prospective collaboration, 'epi-analysis'.

Professor David Barer, University Department of Medicine (Geriatric Medicine), Newcastle General Hospital, Westgate Road, Newcastle upon Tyne NE4 6BE, UK.

References

- 1 Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995; **345**: 772–76.
- 2 De PedroCuesta J, WidenHolmqvist L, BachyRita P. Evaluation of stroke rehabilitation by randomized controlled studies: a review. *Acta Neurologica Scandinavica* 1992; **86**: 433–39.
- 3 Langhorne P, Williams BO, Gilchrist W, Howie K. Do stroke units save lives? *Lancet* 1993; **342**: 395–98.
- 4 Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; **309**: 1286–91.
- 5 Easterbrook PJ, Berlin JA, Gopalan R, Mathews DW. Publication bias in clinical research. *Lancet* 1991; **337**: 867–72.
- 6 Ottenbacher KJ, Jannel S. The results of clinical trials in stroke rehabilitation research. *Arch Neurol* 1993; **50**: 37–44.
- 7 Egger M, Davey Smith G. Misleading meta-analysis. *BMJ* 1995; **310**: 752–54.
- 8 Woods KL, Barnett DB. Magnesium in acute myocardial infarction. *BMJ* 1995; **310**: 1669–70.
- 9 Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994; **309**: 1351–55.
- 10 Barer D. Assessment in rehabilitation. *Rev Clin Gerontol* 1993; **3**: 169–86.
- 11 Collen F, Wade DT, Bradshaw GM. Mobility after stroke: reliability of measures of impairment and disability. *Int Disabil Stud* 1990; **12**: 6–9.
- 12 Collen F, Wade DT, Robb GF, Bradshaw GM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Stud* 1991; **13**: 50–54.
- 13 Dennis M, Langhorne P. So stroke units save lives: where do we go from here? *BMJ* 1994; **309**: 1273–77.
- 14 Barer D, Ellul J. From meta-analysis to epi-analysis. In: Fracchia GN, Haavisto KH eds. *European medicines research*. European Conference Publications, 1995.