Clinical application of findings of clinical trials and systematic reviews

R G Newcombe

The outcome obtained by a clinical team in a therapeutic trial may not be a reliable guide to management of their future patients

The late Professor Archie Cochrane is probably best remembered for enunciating several key principles which are now recognised as fundamental in all areas of clinical practice and associated research. He insisted on the central role of the randomised controlled trial (RCT) in seeking to evaluate the efficacy of interventions, particularly (though not exclusively) in the clinical context. Since the resources available will never be limitless, efficiency is a key issue as well as effectiveness. Furthermore, Cochrane advocated more systematic application of the findings of research. Several trials bearing on the same clinical issue may have yielded divergent findings—what is the clinician then to do? He argued that it is incumbent on the profession to assemble systematically the evidence bearing on an issue. Systematic review uses objective, reproducible criteria to determine which studies yield the most reliable information. The results of the selected studies are then combined, giving due weight to the larger and hence more informative ones, using a mathematical process known as meta-analysis. Systematic review is now accepted as the cornerstone of evidence based clinical practice.

How are we to identify all relevant trials, appraise their quality, and combine their results? The methodological issues that arise in doing so are not trivial. In particular, the meta-analysis process has been the subject of much research by statisticians. Most of this research is far from transparent, and is familiar only to statisticians and others involved in the now rather large industry of producing and regularly updating systematic reviews on a wide range of therapeutic issues.

In an article in this issue, Dr Ulrich Helfenstein presents an interesting and informative account of alternative modelling approaches used in meta-analysis of results from clinical trials (see p 131). This article is addressed to the clinician who has some familiarity with published systematic reviews and with treatment comparisons expressed in terms of log odds ratios. Algebraic details are commendably absent! It explores the key

issue of pooling of effect sizes from different studies, and the implications for clinical practice, particularly from the standpoint of a clinician who has contributed to one of the studies. This issue is important for clinical practice as well as for the biostatistical, evidence based medicine research and systematic reviews communities.

Dr Helfenstein notes that the choice of model may be of vital relevance, and shows how which of the three models is chosen can substantially alter the interpretation of an individual trial within the set of trials included in the meta-analysis. It is clear that the choice of model should not be purely a data-driven one, based solely on statistical considerations. The judicious statistician can collaborate with the clinical research team in assessing the background knowledge relevant to the context, leading to a choice of model that goes considerably beyond the mathematical. Indeed, this is the appropriate form of statistical collaboration in any research project.

As a statistician working in a medical school, but not particularly in the meta-analysis field, I was asked to referee Dr Helfenstein's manuscript. In doing so, the main reservation I expressed relates to the premise that there is a direct relationship between a clinical team's experiences in an RCT and how they should subsequently manage patients similar to those who were eligible for the trial. In other words, should we ever seek to interpret the results of an individual trial, once this has been taken together with others into a meta-analysis process? To broaden the issue somewhat, the team may be contributing to a single centre or a multicentre study, and in either situation the results from this RCT may subsequently be combined with others by the usual meta-analytic processes. Irrespective of which of these possibilities applies, the premise that the team's results in the RCT should directly translate into their preference for future patient management seems highly questionable, on several grounds.

Firstly, a meta-analysis is a retrospective exercise, which includes studies that

have different protocols. For example, in a set of placebo controlled trials of β -blockers, different trials will involve different proprietary β -blockers—all "state of the art", but more distinct than to be classifiable as generic equivalents. They will involve different eligibility criteria in terms of factors such as blood pressure at screening and age. The centres contributing patients to these studies may have caseloads that differ with respect to ethnic composition, a range of possible occupational or environmental exposures, etc. All these factors may potentially alter the balance between two treatment options, and cause divergence (A) between the centre's results in the trial and the conclusions of the meta-analysis and (B) between the conclusions of the meta-analysis and the centre's subsequent results of treating their caseload accordingly.

On balance, this diversity of protocols and caseload characteristics is probably a strength rather than a weakness, as it tends to widen the applicability of the conclusions reached. It leads to a strong prior expectation that an equal effects model is unlikely to be satisfactory, even if the test for heterogeneity of effect size does not reject the null hypothesis. But how does this affect the conclusions to be reached by participating centres? Of the above factors, there is a reasonable expectation that caseload characteristics will remain the same, and it may be worth taking these into account. But this is counterbalanced by the highly selected nature of the patients who get recruited to any RCT. The patients whom the centre recruited to the trial may be importantly different to the present caseload, and the way in which they differ may depend on the particular trial protocol that the centre followed. Consequently, the centre's results in the RCT may be quite unreliable as a guide to management of even ostensibly similar future patients.

Secondly, in surgical studies, an additional complicating feature arises—individual skill and habituation. It is commonly observed, both in RCTs and in routine practice, that surgeon X gets better results with technique A compared with B, whereas surgeon Y gets better results with B than with A. This can be a real difference—different people, even those selected to have similar high levels of skill, nevertheless have different aptitudes. Such differences at the level of the individual clinician imply that the performance of a centre is liable to change over time, with the natural turnover of staff. Furthermore, skills can develop over time—indeed, a key element in introducing a new, potentially beneficial surgical technique such as axillary sentinel node biopsy is habituation.

Finally, above all, there is also chance variation. By chance variation we simply

mean variation that we cannot adequately explain—we are not taking a fatalistic, external locus of control view regarding patient response. The chance element is best thought of in terms of just which individuals from a large, potentially eligible population of patients get recruited and which get allocated to which treatment, rather than postulating a random element in how they respond. Sample size plays an important part here, of course—whatever the context, the degree of uncertainty associated with the conclusion from a series of patients is strongly inversely related to sample size. Accordingly, chance variation affects the overall conclusion of a meta-analysis to some degree, the imprecision is expressed by the width of the confidence interval reported. It affects the conclusions from each constituent trial to a greater degree, and the individual centre's results much more still. For example, the number of patients that an individual centre enrolls in a multicentre RCT is unlikely to be sufficient to justify drawing a firm conclusion using that centre's data alone. Maybe trial 19's pessimistic results were largely contributed to by the play of chance. If so, the clinical teams involved should not take too much

notice of the poor performance of the active treatment in their study. In recognition of the sample size issue, in figure 1 (see p 132) the larger and hence more reliable studies are identified by the use of bold lines.

With all these issues in mind, then, it seems that the correct sequence of events should be as follows. Each centre enrolled in an RCT contributes data to it. This is for the purpose of the trial as a whole—the centre should not be regarded as retaining intellectual property rights to publish their data alone. The trial is written up, agreed by all centres as a fair summary, and published. The systematic reviewers then incorporate this trial into a systematic review, along with other trials that address a sufficiently similar question. This then reaches the clinical team who contributed the data, in either printed or electronic form. And they are issued with clinical practice recommendations and guidelines by the Royal Colleges, NICE (National Institute for Clinical Excellence), etc. It is then the clinician's responsibility to translate this information into the care of his/her patients, taking due account of knowledge of the patient's individual characteristics, and also any relevant local background

knowledge about the underlying population, its circumstances, environment, etc. If the local team can identify just why their results were so discrepant from others, and if this is because of factors that are still likely to operate, then it is right for this to influence their practice. Otherwise, they should accept that they will probably never be able to explain why their results were so atypical. They should then accept the majority consensus as a guide to practice, at any rate for the "typical patient" (whatever this means). And they should use their clinical expertise, judgment, and experience to tailor these recommendations to the individual characteristics and needs of their patients.

The view I have enunciated above is merely my summary of what I regard as fair. It is incumbent upon the clinical research and practice community to seek to reach a consensus on this very important issue.

Postgrad Med J 2002;**78**:129–130

.....

Correspondence to: Dr Robert G Newcombe, Department of Epidemiology, Statistics and Public Health, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK; newcombe@cf.ac.uk