decades of reports of trials of spinal manipulation and mobilisation for back and neck pain will be needed to show an effect of treatment; once of treatment (p=0.009). Target group sizes in allocated, 'but in Koes and colleagues' paper they reviewed. Considering its application to one of the trials such trials. 'This deficiency may best be seen by the methods score used seems not to have been SIR.

Spinal manipulation and mobilisation for back and neck pain

SIR,—In B W Koes and colleagues' review of two decades of reports of trials of spinal manipulation the methods score used seems not to have been sufficiently adapted to the special requirements of such trials. This deficiency may best be seen by considering its application to one of the trials reviewed.

Hadler et al stated that patients were randomly allocated, but in Koes and colleagues' paper they lost all four available points for randomisation by not having stated the method. All 12 points available for adequacy of group sizes were lost as a figure of 50 patients per group was not reached. But the trial showed a highly significant effect of treatment (p=0.009). Target group sizes in protocols are merely estimates of the numbers that will be needed to show an effect of treatment; once a positive result of high significance has been obtained the estimate is superseded by reality. To penalise this trial for a demonstrably adequate group size is illogical.

The trial is not awarded the five points available for use of a placebo control as the sham manipulation employed involved laying on of hands and may thus have had some beneficial effect. If this had been so it would have led to an underestimate of the benefit from manipulation and therefore could not invalidate a positive result.

Ten points were available for using five different outcome measures, and yet Pocock et al, whom Koes and colleagues quote, advise deciding a priori on a small number of outcome measures and end points to avoid invalidating the significance tests used. Other measures may be made as an exploratory feature of the design to compare the utility of different outcome measures, but this secondary function provides data for use in designing further trials and is not relevant when the trial's primary function of assessing outcome is being considered. Credit in this section should be given for the authors stating prospectively a small number of appropriate outcome measures. Hadler et al used one: the disability score designed by Roland and Morris and shown by them to be a more reliable and sensitive index of disability in back pain than measures such as pain experienced or spinal mobility, for which Koes and colleagues would have awarded points.

Ten points were available if the five suggested outcome measures were measured blind. Hadler et al relied on a patient questionnaire administered over the telephone by someone unaware of the treatment, and their control patients had received the most realistic 'sham' treatment of all reported trials. Therefore blind assessment was probably optimum and yet no points were awarded.

If the same categories and weighting that Koes and colleagues used were applied, the adjustments suggested above would increase the score for Hadler et al's trial from 53% to 90% of the maximum. This confirms my view that this trial had the most sophisticated design reported to date and was greatly undervalued in the review.

R S MACDONALD
London College of Osteopathic Medicine, London NW1 8QH

3 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of controlled trials II. The presentation of results, particularly where (as in our case) minimisation is used to permit analyses for different groupings, an advantage that Koes and colleagues did not recognise.

SIR,—B W Koes and colleagues' review of trials of manipulation for back pain has two serious limitations. Firstly, many of the criteria and methods are arbitrary and illogical. A less than homogeneous study population reflects the subjects seen in real life and may increase the replicability of results, particularly where (as in our case) minimisation is used to permit analyses for different groupings, an advantage that Koes and colleagues did not recognise.

It is not mandatory to avoid 'cointerventions' (other treatments) in a pragmatic trial, as resort to other treatments may in day to day practice be the consequence of the approaches under comparison. Indeed, insistence on post-randomisation and the simultaneous avoidance of cointerventions, which is what Koes and colleagues imply by their criteria, make it impossible to recognise the full implications of different policies. These inconsistencies and illogicalities should not be penalised.

It is not mandatory to avoid "cointerventions" (other treatments) in a pragmatic trial, as resort to other treatments may in day to day practice be the consequence of the approaches under comparison. Indeed, insistence on post-randomisation and the simultaneous avoidance of cointerventions, which is what Koes and colleagues imply by their criteria, make it impossible to recognise the full implications of different policies. These inconsistencies and illogicalities should not be penalised.

It is not mandatory to avoid "cointerventions" (other treatments) in a pragmatic trial, as resort to other treatments may in day to day practice be the consequence of the approaches under comparison. Indeed, insistence on post-randomisation and the simultaneous avoidance of cointerventions, which is what Koes and colleagues imply by their criteria, make it impossible to recognise the full implications of different policies. These inconsistencies and illogicalities should not be penalised.

The serious message is that epidemiologists and statisticians may not be qualified to assess the merits of clinical papers and that their pronouncements may be misleading. I believe that such assessments should use appropriate clinicians.

J A MATHEWS
Department of Rheumatology, St Thomas's Hospital, London SE1 7EH


SIR,—In their review of the most worthy papers on manipulation for back and neck pain Koes and colleagues have shown the unsatisfactory nature of almost all previous work. We do not, however, agree with the suggestion that further attempts should be made to answer the same question by using this format. Of the papers reviewed by the authors, four achieved methodological scores of 50-60, and these were all comparatively recent. We think it unlikely that the quality of this type of study could improve dramatically, and this is
supported by a report by the Department of Health and Social Security's working group on back pain, which analysed many papers on this subject in terms of scientific validity and found only three on manipulative therapy that fulfilled its criteria. In general these studies are akin to asking "Is appendicitis an effective treatment for abdominal pain?"

Breen, a chiropractor, suggested that "for those who use manipulative therapy in the management of back pain patients, a clear idea of the nature of pain and its possible sources is of paramount importance." Furthermore, "without rational hypotheses to address these questions treatment becomes incoherent and irrelevant to the problem." By way of example, mechanical and local inflammatory sources of pain may be amenable to manual therapy, but infective, metabolic, neoplastic, and systemic inflammatory causes most probably are not. Indeed, it is quite possible that acute and chronic conditions, whether in the neck or in the back, will respond in different ways. In keeping with Breen's view we believe strongly that any future attempt to rationalise the role of manipulative therapy in the treatment of back pain should be preceded by an attempt to classify and segregate both site and cause.

S M Hay
B Todd

Department of Orthopaedic Surgery, Northern General Hospital, Sheffield S5 7AU


Authors' reply. We are pleased that three of the letters discussing our review are written by authors of trials that we listed in our "top 10." The detailed comments on the methods we used to assess published trials show the considerable advantage of this method of review. Whereas traditionally authors of review articles implicitly apply their (unknown) criteria, we explicitly formulated and used our criteria, which obviously facilitates discussion.

R S MacDonald states that the trial of Hadler et al. was greatly undervalued in the review, but this trial scored 53 points, which was the second best score. MacDonald applied our criteria and weighting to Hadler et al.'s trial and ended up with 90 points. J A Mathews assessed a study of which we were a coauthor and also obtained 90 points. Both of these scores seem to be based on a different use of our criteria and the use of additional information that was not included in the original articles.

The statement that random allocation has been carried out is, in our opinion, not sufficient to warrant all four points for this criterion. All studies included in our review were randomised controlled trials. Studies could earn points if there was a clear description of how the randomisation procedure had been carried out. For readers this is important. The points given for sample size are not meant as a reward for sufficient power. Consequently we do not agree that points should be given depending on whether a certain difference in outcome was significant. Our main reason for giving points for large sample sizes was the assumption that with larger numbers one has more assurance that unknown prognostic indicators will be equally divided between the study groups.

MacDonald's suggestion that all 10 points should be given for relevant outcome measures when an author reports only one or two primary outcome measures is certainly interesting. Of course these primary outcome measures should be chosen (and preferably published) before the data from the trial are analysed. Otherwise the reader has no assurance that the choice of primary outcome measures was not guided by the data analysis.

T W Meade and colleagues performed one of the best studies, but it still showed severe shortcomings. They think that we insisted on pragmatic trials, but this is obviously not the case because studies were also rewarded if they included an adequate placebo treatment. Meade and colleagues claim points for having carried out an analysis based on intention to treat. In our review, however, studies could earn points for this criterion only if in cases of more than 10% loss to follow up (which was the case after two years' follow up) and an alternative analysis had also been carried out correcting for withdrawals and missing values.

We do not agree with Mathews that these kind of reviews can be carried out only by clinicians (although three of us are clinicians). As long as explicit criteria are used that can be applied and checked by most readers we see no ground for this statement.

We agree with S M Hay and B Todd that the design of studies of this treatment is probably difficult to improve. A recent study from our department scored 55 points when assessed by the two reviewers (GJMGH and WJJA, in this case not blinded) who had made the assessments in our meta-analysis. We still think, however, that further trials are needed to determine the efficacy of spinal manipulation and mobilisation for well defined subgroups of patients with back and neck pain. Although the methods of reviewing that we used clearly needed to be developed further, we recommend them for reviewing past and future studies.

Bart Koos
Geert van der Heijden
Pim Assendelft
Lex Bouter
Paul Knipschild

Department of Epidemiology and Biostatistics, University of Limburg, PO Box 616, 6200 MD Maastricht, The Netherlands