Evidence-based clinical guidelines: a new system to better determine true strength of recommendation

Edward Roddy MRCP (Specialist Registrar in Rheumatology),[1] Weiya Zhang PhD (Senior Lecturer in Musculoskeletal Epidemiology),[1] Michael Doherty MA MD FRCP (Professor of Rheumatology),[1] Nigel K. Arden MD MSc MRCP (Senior Lecturer in Rheumatology),[2] Julie Barlow PhD (Professor of Health Psychology),[3] Fraser Birrell MA PhD MRCP (Senior Lecturer in Rheumatology),[4] Alison Carr PhD (Special Lecturer in Musculoskeletal Epidemiology),[1] Kuntal Chakravarty FRCP (Consultant Rheumatologist),[5] John Dickson FRCP MRCGP (Community Specialist in Rheumatology),[6] Elaine Hay MD FRCP (Professor of Community Rheumatology),[7] Gillian Hosie FRCP (General Practitioner),[8] Michael Hurley PhD (Reader in Physiotherapy & ARC Research Fellow),[9] Kelsey M. Jordan MRCP (Rheumatology Research Fellow),[2] Christopher McCarthy PhD (Research Physiotherapist),[10] Marion McMurdo MD FRCP (Professor of Ageing and Health),[11] Simon Mockett MPhil (Senior Lecturer),[12] Sheila O'Reilly MD MRCP (Consultant Rheumatologist),[13] George Peat PhD MCSP (Research Fellow),[7] Adrian Pendleton MD MRCP (Specialist Registrar in Rheumatology),[14] Selwyn Richards MA MSc FRCP (Consultant Rheumatologist)[15]

[1]Academic Rheumatology, Clinical Sciences Building, Nottingham City Hospital, Hucknall Road, Nottingham, UK
[2]MRC Epidemiology Resource Centre, Southampton General Hospital, Tremona Road, Southampton, Hampshire, UK
[3]Interdisciplinary Research Centre in Health, School of Health and Social Sciences, Coventry University, Priory St, Coventry, UK
[4]Musculoskeletal Research Group, University of Newcastle upon Tyne, UK
[5]Haroldwood Hospital, Gubbins Lane, Romford, Essex, UK
[6]Langbaurgh PCT, Langbaurgh House, Bow Street, Guisborough, Cleveland, UK
[7]Primary Care Sciences Research Centre, Keele University, Staffordshire, UK
[8]Primary Care Rheumatology Society, Northallerton, North Yorkshire, UK
[9]King's College London, Rehabilitation Research Unit, Dulwich Hospital, East Dulwich Grove, London, UK
[10]The Centre for Rehabilitation Science, University of Manchester, Oxford Road, Manchester, UK
[11]Department of Medicine, University of Dundee, Ninewells Hospital, Dundee, UK
[12]Division of Physiotherapy Education, School of Community Health Sciences, University of Nottingham, Nottingham City Hospital, Hucknall Road, Nottingham, UK
[13]Derbyshire Royal Infirmary, London Road, Derby, UK
[14]Craigavon Area Hospital, 68 Lurgan Road, Portadown, Co Armagh, UK
[15]Poole Hospital, Longfleet Road, Poole, UK

Correspondence
Edward Roddy
Academic Rheumatology
Clinical Sciences Building
Nottingham City Hospital
Hucknall Road
NG5 1PB
UK
E-mail:
edward.roddy@nottingham.ac.uk

Keywords: clinical guidelines, evidence-based medicine, strength of recommendation

Accepted for publication:
27 April 2005

Abstract

Rationale, aims and objectives Clinical practice guidelines often grade the 'strength' of their recommendations according to the robustness of the supporting research evidence. The existing methodology does not allow the strength of recommendation (SOR) to be upgraded for recommendations for which randomized controlled trials are impractical or unethical. The purpose of this study was to develop a new method of determining SOR, incorporating both research evidence and expert opinion. **Methods** A Delphi technique was employed to produce 10 recommendations for the role of exercise therapy in the management of osteoarthritis of the hip or knee. The SOR for each recommendation was determined by the traditional method, closely linked to the category of research evidence found on a systematic literature search, and on a visual analogue scale (VAS). Recommendations were grouped A-D according to the traditional SOR allocated and the mean VAS calculated. Difference across the groups was assessed by one-

way ANOVA variance analysis. **Results** Mean VAS scores for the traditional SOR groups A-D and one proposition which was 'not recommended' showed significant linearity on one-way ANOVA. However, certain recommendations which, for practical reasons, could not assessed in randomized controlled trials and therefore could not be recommended strongly by the traditional methodology, were allocated a strong recommendation by VAS. **Conclusions** This new system of grading strength of SOR is less constrained than the traditional methodology and offers the advantage of allowing SOR for procedures which cannot be assessed in RCTs for practical or ethical reasons to be upgraded according to expert opinion.

Introduction

Clinical guidelines have been defined as 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical conditions' (Field & Lohr 1990). Guidelines that employ an evidence-based format currently grade each recommendation in two ways: first, by classifying the 'category of evidence' and, second, by giving a 'strength of recommendation'. Although several methods of producing such grades are described, in most of these, including the method most commonly used by clinical guidelines in rheumatology (Pendleton *et al*. 2000; Jordan *et al*. 2003; Dougados *et al*. 2004; Zhang *et al*. 2004; Roddy *et al*. 2005), the latter is strongly dependent on the former (Shekelle *et al*. 1999) (Table 1). That is, the strength of recommendation (SOR) primarily reflects the robustness of the research evidence, with evidence from randomized controlled trials (RCTs) and systematic reviews automatically conferring the strongest recommendation. However, although this traditional method allows a downgrading of the SOR for reasons including side effects or inconsistent studies, it does not allow an upgrading of recommendations in situations where RCTs are impractical or unethical, e.g. total joint replacement, but effectiveness is not in doubt. Furthermore, the practice of evidence-based medicine requires the integration of clinical expertise with the best available evidence from systematic research (Sackett *et al*. 1996). During the development of recent recommendations for the role of exercise in the management of osteoarthritis (OA) of the hip or knee (Roddy *et al*. 2005), we found that the SOR allocated by this method was often discordant with the consensus opinion of the

Table 1 Traditional hierarchy for category of evidence and strength of recommendation (Shekelle *et al*. 1999)

Categories of evidence
1A. meta-analysis of RCT
1B. at least one RCT
2A. at least one CT without randomization
2B. at least one type of quasi-experimental study
3. descriptive studies (comparative, correlation, case-control)
4. expert committee reports/opinions and/or clinical opinion of respected authorities
Strength of recommendation
A. Directly based on category 1 evidence
B. Directly based on category 2 evidence or extrapolated recommendation from category 1 evidence
C. Directly based on category 3 evidence or extrapolated recommendation from category 1 or 2 evidence
D. Directly based on category 4 evidence or extrapolated recommendation from category 1, 2 or 3 evidence

RCT, randomized controlled trial; CT, controlled trial.

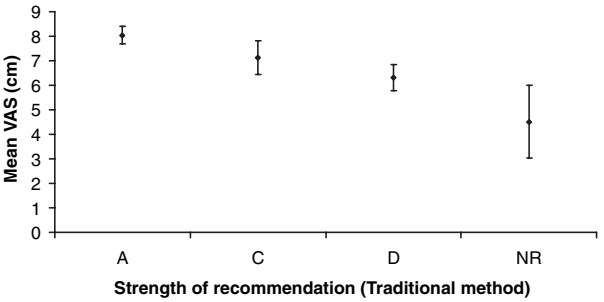
panel. Therefore, we developed an alternative method to better capture the true opinion of the panel, whilst still requiring them to consider the research evidence. We examined whether this alternative approach affected the support afforded to each recommendation.

Methods

A multi-disciplinary panel employed a Delphi technique to produce 10 recommendations relating to the role of exercise in the management of OA of the hip or knee (Roddy *et al*. 2005). Following a literature search and summary analysis of results, the evidence for each recommendation was assessed. The category of evidence and SOR was assigned for each according to the method previously described (Shekelle *et al*. 1999). In addition, each participant was asked to indicate how strongly they rated each recommendation, based not just on research evidence but also on all aspects relating to their knowledge and clinical opinion. This was recorded using a 10-cm visual analogue scale (VAS) anchored with two descriptors labelled 'not recommended at all' at the far left (0 cm) and 'fully recommended' at the far right (10 cm). The mean VAS and standard deviation for each recommendation were calculated. The recommendations were then grouped according to their original SOR (A-D) and the mean VAS and 95% confidence interval calculated for each group. A one-way ANOVA variance analysis was performed to assess the difference between the groups.

Results

The recommendations and the categories of evidence, SOR and VAS for each, are shown in Table 2. Figure 1 shows the mean VAS and 95% confidence interval for recommendation groups A, C and D in addition to one recommendation which was contradicted by the research evidence and could not therefore be graded according to the traditional method ('not recommended'). No recommendations were allocated a grade B SOR. The one-way ANOVA variance analysis identified a significant difference across the groups ($P < 0.001$) and significant linearity ($P < 0.001$).



VAS = visual analogue scale, NR = not recommended

Figure 1 Comparison of mean VAS (95% confidence intervals) and traditional strength of recommendation. VAS, visual analogue scale; NR, not recommended.

Discussion

There was similarity between the SOR produced by this method and the traditional methodology (Shekelle *et al*. 1999). The mean VAS for each recommendation group (A, C, D) increased with the traditional SOR, and therefore the category of evidence, and the lowest mean was seen for the recommendation which could not be recommended by the research evidence ie was based solely on expert opinion.

This new system has the advantage of allowing the SOR to be upgraded or downgraded based on expert opinion relating to global aspects of health care delivery, such as generalizability, safety, cost-effectiveness and patient preference, and common sense. It therefore gives an additional dimension and weighting to guideline recommendations other than just the support from research evidence alone. In the traditional system, the term 'strength of recommendation' is almost a misnomer as it directly relates to the category of evidence and provides little extra information beyond that afforded by the 'category of evidence'. This is an important limitation of currently practised evidence-based guideline methodology that was overlooked in a recent critique of the methodology of OA guidelines (Pencharz *et al*. 2002).

During the development of guidelines there are many situations for which the existing SOR methodology (Shekelle *et al*. 1999) is not ideal. Interventions for which placebo-controlled trials are impractical or unethical (e.g. total joint replacement) cannot score highly on the existing hierarchy and yet clearly may

Table 2 Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee: category of evidence, strength of recommendation (Shekelle *et al*. 1999) and visual analogue score (VAS)

Recommendation	*Category of Evidence (1–4)*	*Strength of Recommendation (A-D)*	*Strength of recommendation (VAS) – Mean (SD) cms*
1. Both strengthening and aerobic exercise can reduce pain and improve function and health status in patients with knee and hip OA.	Knee 1B Hip 4	A C (extrapolated from knee OA)	8.9 (1.1) 6.3 (2.1)
2. There are few contra-indications to the prescription of strengthening or aerobic exercise to patients with hip or knee OA.	4	C (extrapolated from adverse event data)	8.0 (1.5)
3. Prescription of both general (aerobic fitness training) and local (strengthening) exercises is an essential, core aspect of management for every patient with hip or knee OA.	4	D	7.1 (2.5)
4. Exercise therapy for OA of the hip or knee should be individualized and patient-centred taking into account factors such as age, co-morbidity and overall mobility.	4	D	7.7 (1.9)
5. To be effective, exercise programmes should include advice	4	D	6.1 (2.6)
and education to promote a positive lifestyle change with an increase in physical activity.	1B	A	7.7 (1.4)
6. Group exercise and home exercise are equally	1A	A	8.0 (1.5)
effective and patient preference should be considered.	4	D	7.6 (2.3)
7. Adherence is the principal predictor of long-term outcome from exercise in patients with knee or hip OA.	4	D	5.1 (2.4)
8. Strategies to improve and maintain adherence should be adopted, e.g. long-term monitoring/review and inclusion of spouse/family in exercise.	1B	A	7.6 (1.5)
9. The effectiveness of exercise is independent of the presence or severity of radiographic findings.	4	Not recommended	4.5 (2.8)
10. Improvements in muscle strength and proprioception gained from exercise programmes may reduce the progression of knee and hip OA.	4	D	4.2 (2.5)

be very efficacious and warrant strong recommendation for clinical practice. The new method allows the recommendation for such interventions to be upgraded beyond that afforded by the category of research evidence. Furthermore, when recommendations are not easily assessed in the setting of a clinical trial yet have clear face validity, as with our third and fourth recommendations (Table 2), the panel may feel a much stronger recommendation is warranted than that permitted by the current research-linked method. For example, the mean VAS for both propositions 4 and 5B (Table 2) was 7.7, yet the SOR according to the traditional methodology were D and 1B respectively. This reflects that although proposition 4 would be impractical to assess in the setting of a RCT, it was highly supported by the expert panel

whereas robust evidence from RCTs exists to support proposition 5B. Finally, the traditional hierarchy does not accommodate the scenario where research evidence contradicts a recommendation, as with our ninth recommendation (Table 2).

Other guideline methodology groups have attempted to overcome these limitations and reduce the dependence of the SOR on the category of research evidence. However, the grading systems, produced by American College of Cardiology/American Heart Association (ACC/AHA) Task Force (ACA/AHA 2004), the US Preventive Services Task Force (2003), the National Institute for Clinical Excellence (NICE) (NICE 2004) and the New Zealand Guidelines Group (New Zealand Guidelines Group 2004), derive the SOR primarily from the category of research evidence. The ACC/AHA guidelines state that any combination of classification of recommendation and level of evidence is possible and that a recommendation can be strongly supported even if it is based entirely on expert opinion and no research studies have ever been conducted on the recommendation (ACA/AHA 2004). However, this system does not provide for the incorporation of factors such as cost-effectiveness and safety, and the descriptive and quantitative criteria for assigning the classification and evidence ratings weight research evidence and clinical expertise equally, which may not be appropriate for some modalities, e.g. total joint replacement. The guideline development methods of NICE state that when the evidence is very strong, this should translate directly into a recommendation, yet when the literature search finds no evidence to answer the clinical question, the guideline development group should consider using consensus methods to identify current best practice, suggesting that consensus methods are only needed when there is no robust evidence (NICE 2004). Furthermore, NICE produces guidance on the role of individual treatments rather than disease-orientated recommendations on global treatment strategies. The recently published GRADE collaboration (Atkins *et al*. 2004), although highlighting the difficulties in producing clinical guidelines and grading strength of recommendation, has not produced a simple, practical solution. The VAS, on the other hand, has the advantage of being simple to apply and allows all facets to be incorporated, e.g. category of research evidence, safety, cost-effectiveness, generalizability and expert opinion.

A limitation of the VAS-SOR methodology is that as the basis for the VAS is not based on explicit criteria, it cannot be examined and assessed readily by external groups. However, we recommend that the VAS method should be used alongside the traditional method of determining the category of research evidence supporting each recommendation. Any discrepancy between the category of evidence and SOR would therefore be highlighted and should then be justified in the ensuing discussion. A further limitation is that this method has only been used in the setting of recommendations for exercise in osteoarthritis by a single group of experts, so evidence of its generalizability to other fields and other groups is required.

Other possible methods for grading SOR include the development of an ordinal scale. A numerical scale, however, is commonly used to assess self-reported pain and disability in clinical trials, and applying this principle to SOR seemed preferable. Although the numerical scale scores themselves do not have intrinsic comparability between different sets of guidelines, there is at least scope for grading or even ranking of different recommendations within each set of guidelines. Other groups that prefer verbal scales may wish to develop an ordinal scale with descriptors to help guide practice in a clinical setting.

Our guideline development group concludes that, in comparison to existing traditional methodology, this new system of grading SOR is less constrained and offers the advantage of allowing the SOR for procedures which cannot be assessed in RCTs to be upgraded according to expert opinion consistent with the principles of evidence-based medicine (Sackett *et al*. 1996). We would encourage other groups that develop management recommendations or guidelines to try this approach, so that its clinical applicability and usefulness can be determined more widely.

Acknowledgements

We are grateful for an educational grant from MOVE (http://www.move.uk.net) and are also indebted to the Arthritis Research Campaign, UK for financial support (ICAC grant D0593; WZ Senior

Lectureship D0565). We would also like to thank Dr Jinying Lin, a visiting scholar from The People's Hospital of Guangxi Province, China, for assistance with data entry.

John Dickson during the past 5 years has received support to attend or organize symposia, or has received a speaker's honoraria, or a board member's honoraria from one or more of the following companies – MSD, Wyeth, Pfizer, BI, TSB Chemedica, GSK. He has received research funding from Q-med. He has shares in Merck and Pfizer.

Marion McMurdo is a Director of D D Developments Limited, a University of Dundee company whose mission is to provide exercise opportunities for older people. Profits go to ageing research.

References

American College of Cardiology/American Heart Association (2004) *Manual for ACC/AHA Guideline Writing Committees. Section II: Tools and Methods for Creating Guidelines. Step Six: Assign Classification of Recommendations and Level of Evidence.* Available at: http://www.acc.org/clinical/manual/manual_Iistep6.htm (accessed 13 July 2004).

Atkins D., Best D., Briss P.A. *et al.* (2004) Grading quality of evidence and strength of recommendations. *BMJ* **328**, 1490.

Dougados M., Betteridge N., Burmester G.R. *et al.* (2004) EULAR standardised operating procedures for the elaboration, evaluation, dissemination, and implementation of recommendations endorsed by the EULAR standing committees. *Annals of the Rheumatic Disease* **63**, 1172–1176.

Field M. & Lohr K. (1990) *Clinical Practice Guidelines: Directions for a New Program*. National Academy Press, Washington DC.

Jordan K.M., Arden N.K., Doherty M. *et al.* (2003) EULAR recommendations 2003: an evidence based approach to the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Annals of the Rheumatic Disease* **62**, 1145–1155.

National Institute for Clinical Excellence (2004) Guideline development methods. Chapter 11: creating guideline recommendations. Available at: http://www.nice.org.uk/pdf/GDM_Chapter11.pdf (accessed 13 July 2004).

New Zealand Guidelines Group (2004). *Grading Systems for Guidelines.* Available at: http://www.nzgg.org.nz (accessed 13 July 2004).

Pencharz J.N., Grigoriadis E., Jansz G.F. & Bombardier C. (2002) A critical appraisal of clinical practice guidelines for the treatment of lower-limb osteoarthritis. *Arthritis Research* **4**, 36–44.

Pendleton A., Arden N., Dougados M. *et al.* (2000) EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Annals of the Rheumatic Disease* **59**, 936–944.

Roddy E., Zhang W., Doherty M. *et al.* (2005) Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee – the MOVE consensus. *Rheumatology* **44**, 67–73.

Sackett D.L., Rosenberg W.M., Gray J.A., Haynes R.B. & Richardson W.S. (1996) Evidence based medicine: what it is and what it isn't. *BMJ* **312**, 71–72.

Shekelle P.G., Woolf S.H., Eccles M. & Grimshaw J. (1999) Clinical guidelines: developing guidelines. *BMJ* **318**, 593–596.

US Preventive Services Task Force Ratings (2003) *Strength of Recommendations and Quality of Evidence. Guide to Clinical Preventive Services*, 3rd edn. Periodic Updates, 2000–2003. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/clinic/3rduspstf/ratings.htm (accessed 13 July 2004).

Zhang W., Doherty M., Arden N. *et al.* (2004) EULAR evidence-based recommendations for the management of hip osteoarthritis – report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Annals of the Rheumatic Disease.* Published Online First: 7 October 2004, 10.1136/ard.2004.028886.

