

## Measuring up to shoulder pain

Shoulder pain is common, affecting 15–30% of adults at any one time,<sup>1</sup> of whom 1 in 20 will visit a general practitioner in the course of a year.<sup>2</sup> What is the most useful way to categorise such a common symptom to measure its impact, study its aetiology, and determine the efficacy and effectiveness of treatment?

The clinical literature gives primacy to classifications based on presumed pathology. There is little concrete evidence that exalting most shoulder pain with terms such as tendonitis, bursitis or impingement syndrome is either reliable or useful, and such classifications cannot anyway provide a measure of outcome. Clinical measures of shoulder function, such as range of movement, provide a means to classify shoulder problems into subgroups and to assess change over time. However, they do not necessarily reflect patient well being or the ability to carry out usual activities. So the field is open for self completed questionnaires that assess symptom severity and the impact of shoulder pain on everyday living.

At least four instruments that measure function in this way have appeared recently.<sup>3–7</sup> They include one (the Dutch Shoulder Disability Questionnaire or SDQ), which has received further study in this issue,<sup>6,7</sup> two American schedules: the SPADI,<sup>3</sup> and the Shoulder Rating Questionnaire or SRQ,<sup>4</sup> and a British disability questionnaire.<sup>5</sup> A number of others, such as the Dutch Shoulder Function Assessment,<sup>8</sup> incorporate range of movement measurements by an observer, in addition to items completed by the patient. Each, like many new scales accepted for publication these days, have met some or all of the technical criteria of a “good” instrument (repeatability, validity, and responsiveness to change). Readers who are seeking a simple standard tool for use in clinical practice, audit or research, may justifiably be bewildered at the number, range, and persuasiveness of these instruments. Can we ask some simple questions of such measures to choose between them?

A first question is whether instruments specific to the shoulder are superior to measures of general function and health status such as the Short-Form 36.<sup>9</sup> The latter can be used to compare the effects of different musculoskeletal conditions on daily functioning in different patients or to summarise the aggregate effect of different health problems in one person. However, studies that have compared these two types of measure (in osteoarthritis for example<sup>10</sup>) have concluded that outcome studies are more powerfully served by specific measures. Winters and colleagues have also pointed out the practical constraints on using long generic measures in routine clinical practice and to the redundancy of much of the information obtained from them in subjects with isolated shoulder problems.<sup>11</sup> More difficult to judge, in the absence of comparative studies, is the place of instruments that relate to a wider area than the shoulder itself, such as the Disabilities of the Arm, Shoulder and Hand schedule (DASH).<sup>12</sup>

The four questionnaire schedules referred to above<sup>3–7</sup> were all developed as specific instruments for measuring the impact of shoulder problems on daily life. Beyond the jargon and technicalities of questionnaire assessment, what validation is likely to be useful? We want such instru-

ments to reflect real life. The Dutch group behind the SDQ for example<sup>6,7</sup> considered this by selecting items from real clinical histories, consulting with 273 physiotherapists to identify what they perceived to be the 15 most important items, and finally piloting the 15 item schedule with a new group of patients.

The real life that is reflected, however, will depend on the setting chosen for developing and testing the questionnaire. The American schedules, for example, relate to outpatients and specialist office populations in America, where, for example, in the SRQ study<sup>4</sup> 56 of the 100 patients studied had an operation in the year after the first administration of the questionnaire. This is clearly a different context to the two settings in which the British questionnaire<sup>5</sup> was developed (the general population and primary care) or to the outpatient and primary care fields in which the Dutch SDQ has now been tested.<sup>6,7</sup> Other types of shoulder assessment have been tested in disease specific patient groups, such as those with rheumatoid arthritis,<sup>8</sup> or in occupational settings.<sup>12</sup>

Another potentially desirable quality is the range of “domains” covered by the instrument. The British study for example was developed from a generic disability questionnaire, drawing questions from 11 of 12 categories of daily living,<sup>5</sup> and the American SRQ was based on six different domains, including one specifically considering patient satisfaction.<sup>4</sup> By contrast the SPADI covered only two domains,<sup>3</sup> and this meant that sleep disturbance for example did not appear in the SPADI despite it being a frequently reported problem in other studies. Although the inclusion of a wide range of activities potentially affected by shoulder pain seems an attractive quality, we need evidence that it is relevant and useful. Some symptoms, such as irritability or bad temper, may be weak or redundant measures of the overall impact of shoulder pain on daily life.

How important are the technical characteristics of a questionnaire? Repeatability for example can be crucial to observer administered items, such as range of movement measures, but may be a rather over-hyped characteristic for self reported symptoms and daily activities. Reports of poorly repeatable questionnaires in the literature are rare and fluctuations over time in an episodic condition such as shoulder pain will tend to be attributed to change in the condition rather than to poor repeatability.

In the absence of a “gold standard” measure of disability, criterion validity asks: “Do the results using the new instrument reflect results from other measures which might also reflect severity or change?” For example, the Dutch SDQ was compared with a pain scale and a single question about function,<sup>6,7</sup> the American SRQ with a generic measure of arthritis impact,<sup>4</sup> and the other two with range of movement at the shoulder.<sup>3,5</sup> Criterion validity carries the danger of circularity, particularly in the musculoskeletal literature. For example you might read one week that a range of movement scale of shoulder problems is valid because it correlates with a pain score; the next week you read about a pain severity scale that has been validated by its ability to reflect range of movement. Such exercises are in the end a comfort rather than a proof of validity.

Most interest nowadays concerns the use of questionnaires as outcome measures, and in particular their responsiveness to change over time. Of the measures considered here, the Dutch SDQ can now claim to have the most meticulous evidence about this,<sup>7</sup> although the American SRQ provides data on preoperative and postoperative comparisons.<sup>4</sup> Responsiveness can be judged in relation to another measure of change, such as, in the SDQ study, the patient's own judgement of progress on a simple six point scale. Such a simple scale is an appealing yardstick, but why not use it anyway instead of the more complex instrument? The arguments in favour of complexity are both practical and technical. A complex scale allows change in specific activities to be explored, more variation between individuals across a wider range of domains to be summarised, and individual responses to be standardised.

One solution to the problem of the patient's own perception being the main gold standard of change was proposed by Guyatt.<sup>13</sup> This entails asking the patient at baseline to select the item they perceive to be their most important restriction and then measuring changes in it over time as the main outcome. Two of the reviewed questionnaires<sup>4, 6, 7</sup> incorporate a version of this idea. One reported difficulty with it is that patients change their perception of their most important problem by the time of the follow up visit.

Some of the studies considered the concept of "minimal important change".<sup>13</sup> This is the actual change in a questionnaire score that reflects clinically important change in the patient's condition. This is a nice idea, getting away from statistical analyses and presenting actual changes. But how is it estimated? Well once again the patient's own estimate of their progress is trundled out to validate it. The added value of this new measure remains to be clarified.

The final questions to raise about these instruments are more basic. Questionnaires should be simple, easy to use, and look like common sense. There should be evidence that they will work in the setting in which you would like to use them, and so the paper in this journal<sup>7</sup> that takes an instrument developed in an outpatient department<sup>6</sup> and re-examines it in a general practice setting is to be applauded. Different instruments may be required for different patient groups, but this must be balanced against the need to standardise results from different studies by the use of a single instrument.

This last point is one reason why we should attempt to identify a first choice. There are at least four reasonable shoulder specific questionnaires in the literature, many more when generic and clinical instruments are included.<sup>14</sup> The questionnaires considered here seem sensible and incorporate efforts to reflect real life. They are also intriguingly different. They have been developed in different languages, different cultures, and, most importantly, different patient groups, and they emphasise different aspects of the "shoulder experience". One focuses on the hurt that is associated with various shoulder movements,<sup>6, 7</sup> another on the difficulty that is experienced in doing the various tasks,<sup>5</sup> while the American schedules consider both pain and activity restriction.<sup>3, 4</sup> Two have more obvious credentials for use in the community or in primary care<sup>5-7</sup>; one is well tested in a group undergoing surgery.<sup>4</sup> They all go some way to meeting the technical demands of questionnaire design.

An editorial should leap off the fence, but my only rational conclusion is to urge those of you who want to design another shoulder questionnaire to reconsider. A systematic review of shoulder schedules would be useful in laying out a baseline of all published instruments, in helping decisions and in designing changes or comparative studies; and there is at least one excellent example in the occupational literature that reviewed 52 instruments related to the neck and upper limb.<sup>14</sup> The authors of the paper in this issue suggest that more comparisons should be carried out<sup>7</sup> and this seems good advice: more empirical testing of published questionnaires to build up our practical knowledge: how useful are they? how easy to use? how relevant to different situations? While admitting to one seductive but totally unacceptable argument in favour of using a particular questionnaire ("I'm one of the authors"<sup>5</sup>), in the name of science and sanity I would encourage critical comparative studies of published schedules to determine a standard measurement scale for studies of shoulder problems.

I am grateful for the many discussions with the shoulder research teams at the Arthritis and Rheumatism Council's Epidemiology Research Unit, at the EMGO Institute of the Free University in Amsterdam, and at the University Hospital in Leiden. Sue Willson for the manuscript. Mike Doherty for his patience.

PETER CROFT

*Industrial and Community Health Research Centre,  
North Staffordshire Medical Institute, Hartshill,  
Stoke on Trent ST4 7NY*

- 1 Pope DP, Croft PR, Pritchard CM, Silman AJ. Prevalence of shoulder pain in the community: the influence of case definition. *Annals Rheum Dis* 1997;56:308-12.
- 2 McCormick A, Fleming D, Charlton J. *Morbidity statistics from general practice*. 4th National Study 1991-2. Office of Population Censuses and Surveys. Series MB5 No 3. London: HMSO, 1995.
- 3 Roach K E, Budiman-Mak E, Songsrirdej N, Lertratanakul Y. Development of a shoulder pain disability index. *Arthritis Care and Research* 1991;4:143-9.
- 4 L'Insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MGE. A Self-Administered Questionnaire for Assessment of Symptoms and Function of the Shoulder. *J Bone Joint Surg* 1997;79A:738-48.
- 5 Croft P, Pope D, Zonca M, O'Neill T, Silman A. Measurement of shoulder related disability: results of a validation study. *Ann Rheum Dis* 1994;53:525-8.
- 6 Van der Heijden JMG. Shoulder disorder treatment: efficacy of ultrasound therapy and electrotherapy. [PhD Thesis]. Maastricht: University of Maastricht University Press, 1996.
- 7 van der Windt DAWM, van der Heijden GJMG, de Winter AF, Koes BW, Devillé W, Bouter LM. The responsiveness of the Shoulder Disability Questionnaire. *Ann Rheum Dis* 1998;57:82-7.
- 8 van den Ende CHM, Rozing PM, Dijkmans BAC, Verhoef JAC, Voogt-Van der Harst EM, Hazes JMW. Assessment of Shoulder Function in Rheumatoid Arthritis. *J Rheumatol* 1996;23:2043-8.
- 9 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
- 10 Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol* 1995; 22:1193-6.
- 11 Sobel JS, Winters JC. Shoulder complaints in general practice - diagnosis and treatment. [PhD Thesis]. MediTekst, Amsterdam: Rijksuniversiteit Groningen, 1996.
- 12 Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH. The Upper Extremity Collaborative Group. *Am J Ind Med* 1996;29:602-8.
- 13 Guyatt GH, Kirschner B, Jaeschker. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol* 1992;45: 1341-5.
- 14 Stock SR, Cole DC, Tugwell P, Streiner D. Review of applicability of existing functional status measures to the study of workers with musculoskeletal disorders of the neck and upper limb. *Am J Ind Med* 1996;29:679-88.