NOW AND THEN

The Southampton examination schedule for the diagnosis of musculoskeletal disorders of the upper limb

Keith Palmer, Karen Walker-Bone, Cathy Linaker, Isabel Reading, Samantha Kellingray, David Coggon, Cyrus Cooper

Abstract

Objectives—Following a consensus statement from a multidisciplinary UK workshop, a structured examination schedule was developed for the diagnosis and classification of musculoskeletal disorders of the upper limb. The aim of this study was to test the repeatability and the validity of the newly developed schedule in a hospital setting.

Method—43 consecutive referrals to a soft tissue rheumatism clinic (group 1) and 45 subjects with one of a list of specific upper limb disorders (including shoulder capsulitis, rotator cuff tendinitis, lateral epicondylitis and tenosynovitis) (group 2), were recruited from hospital rheumatology and orthopaedic outpatient clinics. All 88 subjects were examined by a research nurse (blinded to diagnosis), and everyone from group 1 was independently examined by a rheumatologist. Between observer agreement was assessed among subjects from group 1 by calculating Cohen's κ for dichotomous physical signs, and mean differences with limits of agreement for measured ranges of joint movement. To assess the validity of the examination, a pre-defined algorithm was applied to the nurse's examination findings in patients from both groups, and the sensitivity and specificity of the derived diagnoses were determined in comparison with the clinic's independent diagnosis as the reference standard.

Results—The between observer repeatability of physical signs varied from good to excellent, with κ coefficients of 0.66 to 1.00 for most categorical observations, and mean absolute differences of 1.4°–11.9° for measurements of shoulder movement. The sensitivity of the schedule in comparison with the reference standard varied between diagnoses from 58%–100%, while the specificities ranged from 84%–100%. The nurse and the clinic physician generally agreed in their diagnoses, but in the presence of shoulder capsulitis the nurse usually also diagnosed shoulder tendinitis, whereas the clinic physician did not.

Conclusion—The new examination protocol is repeatable and gives acceptable diagnostic accuracy in a hospital setting. Examination can feasibly be delegated to a trained nurse, and the protocol has the benefit of face and construct validity as well as consensus backing. Its performance in the community, where disease is less clear cut, merits separate evaluation, and further refinement is needed to discriminate between discrete pathologies at the shoulder.

(*Ann Rheum Dis* 2000;**59**:5–11)

Musculoskeletal disorders of the upper limb and neck are a common cause of morbidity,[1–10] but their exact frequency and burden on health in the community are difficult to ascertain. This difficulty arises in part because they comprise a heterogeneous group of clinical disorders and non-specific regional pain syndromes. Disagreement exists about case definition and about the distinction, relation and overlap between conditions; this lack of consensus hampers meaningful comparison between studies.[11 12]

To address this issue, a workshop convened by the UK Health and Safety Executive (HSE) used a "Delphi" technique to develop consensus criteria for some of the more common upper limb disorders.[13] By collating, analysing and re-discussing opinions in a structured manner, a broadly constituted group of rheumatologists, orthopaedic surgeons, occupational physicians, epidemiologists, physiotherapists and ergonomists were able to agree diagnostic criteria for nine categories of upper limb disorder (table 1).

These criteria provide a useful starting point for surveys of upper limb and neck complaints in the general population, but do not include full detail of the relevant procedures and definitions.[12] It is not apparent, for example, whether two observers would agree where the boundaries of each anatomical region lie, what procedures should be adopted to elicit pain on resisted movement, where the pain should be

The MRC Environmental Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton SO16 6YD

Correspondence to: Dr Palmer

Table 1 Diagnostic criteria for upper limb disorders proposed by the HSE Workshop (adapted from Harrington et al, 1998[13])

Disorder	Diagnostic criteria
Rotator cuff tendinitis	History of pain in the deltoid region *and* pain on resisted active movement (abduction - supraspinatus; external rotation - infraspinatus; internal rotation - subscapularis)
Bicipital tendinitis	History of anterior shoulder pain *and* pain on resisted active flexion or supination of forearm
Shoulder capsulitis (frozen shoulder)	History of pain in the deltoid area *and* equal restriction of active and passive glenohumeral movement with capsular pattern (external rotation > abduction > internal rotation)
Lateral epicondylitis	Epicondylar pain *and* epicondylar tenderness *and* pain on resisted extension of the wrist
Medial epicondylitis	Epicondylar pain *and* epicondylar tenderness *and* pain on resisted flexion of the wrist
De Quervain's disease of the wrist	Pain over the radial styloid *and* tender swelling of first extensor compartment *and either* pain reproduced by resisted thumb extension *or* positive Finkelstein's test
Tenosynovitis of wrist	Pain on movement localised to the tendon sheaths in the wrist *and* reproduction of pain by resisted active movement
Carpal tunnel syndrome	Pain *or* paraesthesia *or* sensory loss in the median nerve distribution *and one of*: Tinel's test positive, Phalen's test positive, nocturnal exacerbation of symptoms, motor loss with wasting of abductor pollicis brevis, abnormal nerve conduction time
Non-specific diffuse forearm pain	Pain in the forearm in the absence of a specific diagnosis or pathology (sometimes includes: loss of function, weakness, cramp, muscle tenderness, allodynia, slowing of fine movements)

felt, and what degree of restricted movement is important.

Using the Workshop criteria as a basis for case definition, we have developed a detailed schedule of diagnostic procedures that could be followed in epidemiological field studies, together with a pre-defined algorithm by which the findings are translated into diagnoses. As a first step towards determining the validity and repeatability of our schedule, we have investigated its performance in a hospital setting among relatively clear cut cases of upper limb disease.

Method
The repeatability of the schedule was assessed by comparing the physical signs elicited in a group of hospital outpatients who were examined independently by two observers—a research nurse (CL) and a rheumatologist (KWB). In addition, the face validity of diagnoses derived from the nurse's examination was assessed in comparison with diagnoses made independently in the hospital clinic as the reference standard.

The nurse and the rheumatologist had been trained in the examination schedule through a series of examinations carried out jointly on normal subjects and outpatients with soft tissue rheumatism, followed later by independent assessments with a comparison of findings. Training continued over a six week period (12 sessions) until a high level of consistency was achieved between the observers.

The examination entailed recording the location of pain at the shoulder, elbow, wrist and hand; eliciting signs of tenderness and pain on resisted movement at these sites; conducting three standard clinical provocation tests (Finkelstein's test, Phalen's test and Tinel's test), and searching for tender spots, as described in the American College of Rheumatology criteria for diagnosis of fibromyalgia.[14] The range of shoulder movement was measured in accordance with the methods described by Norkin and White,[15] using a goniometer for external rotation and a plane inclinometer (pleurimeter) for other shoulder movements.[16] Lateral flexion, forward flexion and extension of the neck were also measured using the plane inclinometer. A hand diagram similar to that used by Katz *et al*[17] was used to record and classify the pattern of sensorineural complaints in the hand. The full schedule also includes a procedure for field measurement of median nerve conduction latencies using a portable electroneurometer: this method has been evaluated elsewhere,[18] and is not reported here. (Further particulars of the schedule are available on request).

The study sample came from two sources. Subjects for the repeatability element of the survey were drawn from consecutive referrals to the soft tissue rheumatism clinic at Southampton General Hospital (group 1). Recruitment occurred over a seven month period between November 1997 and May 1998. Everyone who had been referred because of neck or upper limb symptoms was eligible, and all agreed to participate. Each subject underwent two examinations, one by the nurse and one by the rheumatologist. The examinations were conducted during the patient's visit to the clinic and were performed independently, spaced by an interval of a few minutes.

For the assessment of validity, information was separately abstracted on the diagnosis or diagnoses made by the clinic. In addition, to supplement this part of the investigation, a second group of subjects was recruited from patients attending rheumatology and orthopaedic outpatient clinics at Southampton General Hospital and two district general hospitals (group 2). Consecutive cases with one or more of a specified list of upper limb diagnoses were identified by the doctor in the clinic and invited to undergo examination by the research nurse. The conditions included were: adhesive capsulitis, bicipital tendinitis, rotator cuff tendinitis, lateral epicondylitis, medial epicondylitis, carpal tunnel syndrome, de Quervain's disease of the wrist, and tenosynovitis of the wrist. Everyone who was eligible agreed to be examined. Examinations were performed during the clinic visit with the nurse blind to the diagnosis that had been made in the clinic.

Observations on subjects from group 1 contributed to the assessment of repeatability, while those on subjects from both groups were used to assess the validity of the schedule. The between observer repeatability of physical signs was assessed among patients from group 1 by calculating a Cohen's κ coefficient (or weighted κ coefficient, as appropriate) for categorical variables,[19] and mean differences and limits of agreement (mean (+2SD) difference)[20] for the continuous variables—range of shoulder movement and range of neck movement. The

Table 2　Clinical diagnoses in the study sample

Clinic diagnosis	Group 1	Group 2	All subjects
Single diagnosis			
Adhesive capsulitis	9	5	14
Bicipital tendinitis	1	0	1
Rotator cuff tendinitis	8	4	12
Lateral epicondylitis	4	4	8
Medial epicondylitis	0	0	0
Carpal tunnel syndrome	1	13	14
De Quervain's disease	3	3	6
Tenosynovitis of the wrist	1	0	1
Two diagnoses			
Lateral and medial epicondylitis	1	1	2
De Quervain's and Carpal tunnel syndrome	0	1	1
Three diagnoses			
Adhesive capsulitis and lateral and medial epicondylitis	1	0	1
None of the above	14	14	28
Total	43	45	88

repeatability of the diagnoses derived from the examination was also compared between observers. Finally, for all subjects (that is, both group 1 and group 2), the nurse's derived diagnoses were compared with those reached in the clinics, and the sensitivity and specificity of the nurse's diagnoses were determined, with those of the clinician as the reference standard.

Results

Altogether, 88 subjects were examined (43 patients in group 1 and 45 in group 2). Participants had a median age of 49 years (IQ range 40–57) and 46 (52%) were women. Table 2 records the diagnoses made by the clinics in these subjects. A total of 56 subjects were identified as having one of the specific upper limb disorders covered by the schedule, three cases had two, and one case had three specific diagnoses. Twenty eight subjects had none of the disorders, but were suffering from other rheumatic complaints. The most common diagnoses were adhesive capsulitis (15 subjects altogether), carpal tunnel syndrome (15), rotator cuff tendinitis (12) and lateral epicondylitis (11). Among subjects considered to have none of the scheduled disorders, the most common diagnoses were cervical spondylosis or brachial neuralgia *(*14), diffuse arm pain (6), fibromyalgia (2), and seronegative polyarthritis (2).

REPEATABILITY
The assessment of repeatability was based on the 43 subjects (86 limbs) from group 1. Table 3 records the extent of agreement on physical signs between the two observers. κ Coefficients varied between 0.54 and 0.93 at the shoulder, between −0.02 and 0.79 at the elbow and between 0 and 1 in the forearm and hand. There were particularly high levels of agreement at the shoulder for presence of a painful arc (κ = 0.93) and pain on resisted external rotation (κ = 0.90), flexion (κ = 0.83) and abduction (κ = 0.81); at the elbow for signs of lateral epicondylitis (κ = 0.75 for tenderness over the lateral epicondyle and for pain on resisted wrist extension); and in the hand for abnormal sensation of light touch affecting the little finger (κ = 1.0) and Finkelstein's test (κ = 0.79). Tenderness of the medial epicondyle and resisted movements of the fingers and

Table 3　Between observer repeatability of physical signs included in the schedule

Signs	Number of pairs examined	Observer 1 / Observer 2				κ (SE)
		−/−	−/+	+/−	+/+	
Shoulder tenderness	86	68	2	3	13	0.80 (0.11)
Shoulder pain on resisted:						
* elbow flexion	86	70	3	1	12	0.83 (0.11)
* forearm supination	86	73	3	3	7	0.66 (0.11)
* external rotation	86	66	2	1	17	0.90 (0.11)
* internal rotation	86	74	4	3	5	0.54 (0.11)
* abduction	86	67	0	5	14	0.81 (0.11)
* Painful arc	86	78	1	0	7	0.93 (0.11)
(*) +ve a–c joint stress test	86	76	3	0	7	0.80 (0.11)
* Elbow tenderness						
- lateral	86	72	5	0	9	0.75 (0.10)
- medial	86	81	4	1	0	−0.02 (0.09)
* Pain lateral elbow, resisted wrist extension	86	78	2	1	5	0.75 (0.11)
* Pain lateral elbow, resisted wrist flexion	86	83	1	0	2	0.79 (0.11)
(*) Swelling posterior elbow	86	85	0	0	1	1.00 (0.11)
* Radial wrist tenderness	86	84	1	0	1	0.66 (0.10)
* Tenderness forearm, dorsal or palmar wrist	86	82	1	1	2	0.65 (0.11)
* Pain on resisted movement:						
* radial wrist	86	84	0	1	1	0.66 (0.10)
* medial wrist	86	84	0	1	1	0.66 (0.10)
* finger extension	86	82	0	4	0	0 (–)
* finger flexion	86	84	0	2	0	0 (–)
* Muscle wasting (thenar)	86	85	1	0	0	0 (–)
(*) Dupuytren's contracture	86	78	1	1	6	0.84 (0.11)
Abnormal light touch:						
* thumb	86	82	3	0	1	0.39 (0.09)
* index finger	86	82	0	2	2	0.66 (0.10)
* little finger	86	82	0	0	4	1.00 (0.11)
* +ve Phalen's test	86	85	0	0	1	1.00 (0.11)
* +ve Tinel's test	86	85	1	0	0	0 (–)
Weakness of thumb:						
* abduction	86	85	1	0	0	0 (–)
* opposition	86	82	2	2	0	−0.02 (0.11)
* Pain, resisted thumb extension	86	81	1	2	2	0.55 (0.11)
* +ve Finkelstein's test	86	83	1	0	2	0.79 (0.11)
* Katz hand diagram, classic	84	83	0	0	1	1.00 (0.15)

Analysis is based on the 86 limbs (43 patients) examined both by the nurse and the rheumatologist. Observer 1 = nurse observer; 2 = rheumatologist. *Form part of the Delphi criteria; (*) other items we have added.

Table 4 *Repeatability of measurements of shoulder and neck movement included in the examination schedule*

	Number of pairs	Between observer						
		Mean (obs 1)	Mean (obs 2)	Mean difference (obs 1—obs 2)	SD	Limits of agreement (mean (2SD) difference)	% pairs within 10°	% pairs within 20°
Shoulder:								
Range of active movement (°):								
abduction	86	143.3	137.0	6.3	14.7	(−23.1, 35.7)	64	87
forward flexion	86	142.5	130.6	11.9	19.5	(−27.1, 50.9)	56	70
extension	86	62.0	56.6	5.4	14.0	(−22.6, 33.4)	72	88
external rotation	86	54.2	55.6	−1.4	16.2	(−33.8, 31.0)	73	87
internal rotation	86	106.4	100.4	6.2	5.8	(−5.4, 17.8)	98	100
Range of passive movement (°):								
abduction	86	147.9	143.1	4.8	16.9	(−29.0, 38.6)	64	88
forward flexion	86	145.3	134.4	11.0	18.9	(−26.8, 48.8)	57	71
extension	86	62.8	56.7	6.1	13.7	(−21.3, 33.5)	70	88
external rotation	86	54.9	56.3	−1.4	16.3	(−34.0, 31.2)	73	87
internal rotation	86	106.6	100.2	6.4	5.1	(−3.8, 16.6)	99	100
Neck:								
Range of active movement (°):								
rotation (right)	43	65.6	65.5	0.1	10.0	(−19.9, 20.1)	86	98
rotation (left)	43	65.6	70.5	−4.9	10.5	(−25.9, 16.1)	77	95
flexion	43	53.8	47.7	6.2	12.9	(−9.6, 32.0)	60	91
extension	43	48.5	49.0	−0.5	13.1	(−26.7 ,25.7)	67	98
lateral flexion (right)	43	42.3	39.1	3.3	11.0	(−18.7, 25.3)	72	98
lateral flexion (left)	43	41.3	32.3	2.2	14.9	(−27.6, 32.0)	65	88

Analysis is based on the 86 limbs (43 patients) examined both by the nurse and the rheumatologist.

thumb proved to be the least repeatable signs (κ −0.02 to 0.55).

An analysis based on limbs (rather than subjects) could be biased towards agreement through a lack of independence in individuals between paired observations from their right and left arms. To allow for within subject concordance, we recalculated the κ values by examining agreement within patients rather than limbs, but the findings were little changed. Thus, for abnormal light touch in the thumb, the κ value fell from 0.39 to 0.38, while smaller differences than this were found for all of the other estimates of agreement.

The extent of between observer agreement for measurements of neck and shoulder movement is shown in table 4. Mean absolute differences were comparatively small: 1.4–11.9° for active shoulder movements; 1.4–11.0° for passive shoulder movements; and 0.1–6.2° for active cervical movements. Internal rotation was the least repeatable of the three measurements at the shoulder used in the algorithm to assess for adhesive capsulitis.

In keeping with the broad agreement between observers on physical signs, application of the diagnostic algorithm produced complete between observer concordance on the diagnoses derived from the examination schedule.

SENSITIVITY AND SPECIFICITY

The analysis of sensitivity and specificity was based on 176 limbs in all 88 subjects. Table 5

records the number of people who had each of the specified diagnoses according to the clinic doctors, the number according to the nurse, the number for whom the nurse and clinic doctors agreed on the diagnosis, and the sensitivity and specificity of the examination schedule based on the nurse's assessment relative to that of the clinic as the standard.

Altogether, 65 specific diagnoses were made by the clinic and 71 by the nurse. For the more common conditions the schedule had a high overall specificity (84%–100%), and a somewhat lower sensitivity—adhesive capsulitis (87%), rotator cuff tendinitis (58%), lateral epicondylitis (73%), carpal tunnel syndrome (67%) and De Quervain's disease (71%).

The diagnosis of rotator cuff tendinitis was found to be the main area of disagreement between the nurse and clinic physician. The nurse diagnosed this condition more often (19 subjects versus 12 for the clinic); and in only seven of 24 cases diagnosed by either the nurse or clinic was there complete agreement between them. Among 12 other cases that were diagnosed by the nurse but not the clinic, adhesive capsulitis was also diagnosed both by the nurse and by the clinic. This suggests that in cases of capsulitis the schedule also makes a diagnosis of shoulder tendinitis, but practising clinicians do not. Five cases were diagnosed only by the clinic, and in these cases the patient did not report shoulder pain on the nurse's inquiry (the algorithm does not allow tendinitis to be diagnosed in the absence of shoulder pain). Similarly, absence of reported pain rather than restricted shoulder movements accounted for the discrepancy in two cases of adhesive capsulitis that were diagnosed by the clinic but not by the nurse.

Discussion

There have been many investigations of the frequency,[1–10] causes[24–27] and treatment[28–31] of soft tissue disorders of the upper limb and neck, particularly in occupational health settings, but interpretation of these studies has

Table 5 *The sensitivity and specificity of the examination schedule*

Disorder	Diagnosed by clinic	Diagnosed by nurse	Diagnosed by both	Nurse versus clinic	
				Sensitivity (%)	Specificity (%)
Adhesive capsulitis	15	20	13	87	90
Bicipital tendinitis	1	3	1	100	98
Rotator cuff tendinitis	12	19	7	58	84
Lateral epicondylitis	11	10	8	73	97
Medial epicondylitis	3	0	0	0	100
Carpal tunnel syndrome	15	10	10	67	100
De Quervain's disease	7	5	5	71	100
Tenosynovitis	1	4	1	100	97

Analysis based on 176 limbs in all 88 subjects.

been hampered because of differences in nomenclature and approaches to case definition.[12][13][32] Some studies have focused only on regional pain symptoms, but these health end points are too non-specific to have direct clinical value. Orthodox approaches to classification, based on classic text book descriptions of expected clinical features[33] provide an alternative framework for assessment, but one that has not been standardised or validated. Differences exist in taxonomy and the current use of diagnostic labels,[32] and in most cases no reliable standard exists against which to resolve disagreements.

Several researchers have adopted a more structured approach to the classification of soft tissue disorders of the upper limb and neck. A systematic review by Buchbinder *et al*[32] identified four systems—two from Finland[33][34] and two from North America[1][35]—in which explicit criteria were proposed, intended to classify all, or a significant proportion of soft tissue disorders into distinct categories. All four schemes were developed to investigate neck and upper limb disorders in epidemiological surveys in the occupational health setting. However, all were criticised because of their failure to demonstrate satisfactory between and within observer repeatability, or to demonstrate construct validity against alternative systems in the same domain of enquiry. In addition, the protocols of Viikari-Juntura[34] and Silverstein[35] were considered to be elaborate, requiring one or more trained specialists with special skills to undertake the assessment; while in the case of McCormack *et al*[1] and Waris *et al*[33] there had been a failure to demonstrate that a delegated examination was adequate and accurate. Another concern, was the failure to cater for patients who partially fulfilled the criteria or who had patterns of complaint that fell outwith the specified categories. To these concerns may be added the practical observation that no system specified a method for conducting the assessment that could be described and followed by third parties.

Buchbinder *et al* recommended that future work should be directed toward improving the systems of classification or developing new approaches that fulfilled basic measurement criteria. Our own study is aimed to establish a practical examination schedule that has a measurable standard of construct validity and repeatability, and that can feasibly be delegated to research nurses in field epidemiological investigations. The criteria that underlie the protocol were based on the views of a multi-disciplinary workshop, and have the benefit of face validity and consensus backing.

Although consensus was reached at the HSE workshop, the number of participants was too small for it to be represented as a UK view, much less an international view. Nevertheless, the level of agreement achieved has enabled us to develop, test and endorse a practical measuring instrument that performs well in relation to clinical opinion. Assessment was also made of the between observer repeatability of elements of the examination and of its diagnostic conclusions. The data indicate that the components of the examination schedule are generally repeatable, as judged by Fleiss's criteria (a κ greater than 0.75 is said to denote excellent agreement and that of 0.4–0.75 a good agreement).[36] These findings were based on an analysis of limbs, and might have been biased towards agreement if paired observations on the right and left side were not independent; but the κ values changed little in an analysis based on individual subjects rather than limbs. Furthermore, after a period of training, no systematic differences were evident between observers in the elicitation of signs such as tenderness and pain on resisted movement. The derived diagnoses were also found to be repeatable, although numbers in several of the diagnostic groups were small.

Our findings on reliability follow a period of training, provided to promote consistency between the observers. In this context, it should be noted that between observer agreement is not a fixed quantity, but varies according to the degree of training undertaken and how recent it is (and so periodic refresher training and re-evaluation would be warranted when using the instrument in a longitudinal investigation).

There was a high overall level of specificity with an acceptable sensitivity. The reference standard in our analysis of sensitivity and specificity was the diagnosis made in the clinics. We recognise that other clinicians in other clinics might have reached different diagnoses. However, the level of agreement achieved demonstrates face validity in relation to the everyday opinions of a group of rheumatologists and orthopaedic surgeons. In addition, attempting to assess a large number of diagnostic entities, in contrast with studies that consider the presence or absence of a single disorder, poses an extra challenge in interpretation. The imperative for this inclusive approach arises from the need to distinguish between competing diagnostic possibilities, but it has also enabled us to explore the overlap between disorders.

Although the schedule performed well in many respects, there were several areas of difficulty. Our observations suggest that less stringent criteria are used by hospital doctors than required by the HSE Workshop when faced with patients complaining of elbow pain. The criteria for epicondylitis proposed by the Workshop require local pain plus two physical signs, but relaxing the criteria to require only one of the two signs would increase the sensitivity for a diagnosis of lateral epicondylitis from 73% to 91% while only reducing the specificity from 97% to 96%. A similar adjustment to the algorithm would improve the sensitivity for a diagnosis of medial epicondylitis from 0% to 67% but leave the specificity little changed at 98% (down from 100%). In keeping with others,[37] we found that patients with physical signs of carpal tunnel syndrome did not always shade a hand diagram in a way that indicated a classic or probable distribution of symptoms. Relaxing the criteria in the presence of supporting signs to permit a diagnosis based on a "possible" pattern of symptoms (as

defined by Katz *et al*[17]) would have resulted in complete agreement between the nurse and clinics. Adjustments to the algorithm might improve its performance in this setting, but could lead to an unacceptable rate of false-positive diagnoses when it was applied in the general population.

The discrimination between disorders of the shoulder was particularly difficult. In this small sample the doctors in the clinics never diagnosed rotator cuff tendinitis and adhesive capsulitis in the same person, whereas the schedule nearly always diagnosed the former in the presence of the latter. The data in table 4 indicate an element of disagreement between observers in the estimated range of shoulder movements, (albeit similar to that found elsewhere[16]), but altering the cut points used to define restriction did not affect the extent of overlap, which is not likely to be explained by this source of imprecision. Alternatively, differences might have occurred because the conditions frequently coexist but doctors fail to consider or record a second diagnosis—assuming capsulitis to be the main underlying pathology. However, a more probable explanation is that subjects with capsulitis also experience pain on resisted movement, and thereby fulfil the definition for tendinitis proposed by the HSE Workshop. Refinement of the diagnostic criteria may therefore be appropriate, and the problem merits more detailed inquiry, including observations on a larger sample, studies of natural history, and perhaps investigation with magnetic resonance imaging.

The schedule was tested in a panel of patients from hospital clinics, many of whom were known to be suffering specific disorders of the upper limb severe enough to present for secondary health care. The results of our study should be interpreted in this context. The specificity of a test in any population depends on the group chosen to represents those without the disorder. In a hospital outpatient setting, subjects who were seen in clinic but not considered to have the disease in question form the natural comparative group, and in this context it seems that a trained nurse can achieve an acceptable level of diagnostic accuracy. However, in the community, where non-specific regional pain syndromes may be more common and specific upper limb and neck disorders less severe, the outcome is less assured, and further assessment of the schedule would be required.

As with other initiatives in this area, the criteria that have been developed are not comprehensive as to all possible upper limb pathologies. They also suffer the common weakness of aggregating potentially dissimilar disorders into a single category of "non-specific upper limb pain" to accommodate presentations that do not satisfy the criteria of discrete clinical disorders. This reflects the extent of consensus at the HSE workshop, where there was no agreement, for example, on how to classify arm pain of cervical origin, and no discussion about acromioclavicular joint dysfunction or olecranon bursitis. However, the schedule provides a base from which further expansion is possible. We have added to it some modifications of our own in areas of evident deficiency: notably, the recording of neck pain and restricted neck movement so that its association with arm pain can be mapped. Provided that case definitions can be agreed for other specific disorders of the upper limb and neck, then a similar approach of validation can be followed, permitting the schedule's utility to be extended. A major challenge will be to determine whether patients currently labelled as having non-specific upper limb pain can be subdivided further, into groups defined by clusters of symptoms and signs that predict prognosis and response to treatment: a reproducible physical examination of the kind tested here will be indispensable to this task.

1 McCormack RR Jr, Inman RD, Wells A, Berntsen C, Imbus HR. Prevalence of tendinitis and related disorders of the upper extremity in a manufacturing workforce. J Rheumatol 1990;17:958–64.
2 Bergenudd H, Lindgarde F, Nilsson B, Petersson CJ. Shoulder pain in middle age. A study of prevalence and relation to occupational workload and psychosocial factors. Clin Orthopaed 1988;231:234–8.
3 Makela M, Heliovaara M, Sievers K, Impivaara O, Knekt P, Aromaa A. Prevalence, determinants, and consequences of chronic neck pain in Finland. Am J Epidemiol 1991;134:1356–67.
4 Allander E. Prevalence, incidence and remission rates of some common rheumatic diseases or syndromes. Scand J Rheumatol 1974;3:145–53.
5 Dimberg L. The prevalence and causation of tennis elbow (lateral humeral epicondylitis) in a population of workers in an engineering industry. Ergonomics 1987;30:573–80.
6 Luopajarvi T, Kuorinka I, Virolainen M, Holmberg M. Prevalence of tenosynovitis and other injuries of the upper extremities in repetitive work. Scand J Work Environ Health 1979;5 (suppl 3):48–55.
7 Lawrence JS. Disc degeneration, its frequency and relationship to symptoms. Ann Rheum Dis 1969;28:121–38.
8 Cunningham LS, Kelsey JL. Epidemiology of musculoskeletal impairments and associated disability. Am J Pub Health 1984;74:574–9.
9 de Krom MCTFM, Kester ADM, Knipschild PG, Spaans F. Risk factors for carpal tunnel syndrome. Am J Epidemiol 1990;132:1102–10.
10 Stevens JC, Sun S, Beard CM, OFallon WM, Kurland LT. Carpal tunnel syndrome in Rochester, Minnesota, 1961 to 1980. Neurology 1988;38:134–8.
11 Harrington JM, Gompertz D, MCA RH. Repetitive strain injuries. Lancet 1997;349:1701–2.
12 Palmer K, Coggon D, Cooper C, Doherty M. Work-related upper limb disorders: getting down to specifics. Ann Rheum Dis 1998;57:445–6.
13 Harrington JM, Carter JT, Birrell L, Gompertz D. Surveillance case definitions for work-related upper limb pain syndromes. Occup Environ Med 1998;55:264–71.
14 Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, *et al*. The American College of Rheumatology 1990. Criteria for the classification of fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 1990;33:160–72.
15 Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry. Philadelphia: Davies FA, 1985.
16 Green G, Buchbinder R, Forbes A, Bellamy N. A standardized protocol for measurement of range of movement of the shoulder using the Plurimeter-V Inclinometer and assessment of its intrarater and interrater reliability. Arthritis Care and Research 1998;11:43–52.
17 Katz JN, Larson MG, Fossel AH, Liang MH. Validation of surveillance case definition of carpal tunnel syndrome. Am J Public Health 1991;81:189–93.
18 Dunne CA, Thompson PW, Cole J, Dury J, Martyn CN, Coggon D, *et al*. Carpal tunnel syndrome: evaluation of a new method of assessing median nerve conduction at the wrist. Ann Rheum Dis 1996;55:396–8.
19 Cohen S, Karmarck T, Mermelstein R. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;20:37–46.
20 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;i:307–10.

21 Hagberg M, Wegman DH. Prevalence rates and odds ratios of shoulder-neck diseases in different occupational groups. Br J Ind Med 1987;44:602–10.
22 Herberts P, Kadefors R, Hogfors C, Sigholm G. Shoulder pain and heavy manual labour. Clin Orthopaed 1984;191: 166–78.
23 Silverstein BA, Fine LJ, Armstrong TJ. Hand wrist cumulative trauma disorders in industry. Br J Ind Med 1986;43:779–84.
24 Withrington RH, Girgis FL, Seifert MH. A comparative study of the aetiological factors in shoulder pain. Br J Rheumatol 1985;24:24–6.
25 Viikari-Juntura E, Vuori J, Silverstein BA, Kalimo R, Kuosma E, Videman T. A life-long prospective study on the role of psychosocial factors in neck-shoulder and low-back pain. Spine 1991;16:1056–61.
26 Dimberg L, Olafsson A, Stefansson E, Aagaard H, Oden A, Andersson GBJ, *et al*. The correlation between work environment and the occurrence of cerviobrachial symptoms. J Occup Med 1989;31:447–53.
27 Stock SR. workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. Am J Ind Med 1991;19:87–107.
28 Kilbom A. Intervention programmes for work-related neck and upper limb disorders: strategies and evaluation. Ergonomics 1988;31:735–47.
29 Neustadt DH. Local corticosteroid injection therapy in soft tissue rheumatic conditions of the hand and wrist. Arthritis Rheum 1991;34:923–6.

30 England S, Farrell AJ. Low power laser therapy of shoulder tendonitis. Scand J Rheumatol 1989;18:427–31.
31 Hollingworth GR, Ellis RM, Hattersley TS. Comparison of injection techniques for shoulder pain: results of a double blind, randomised study. BMJ 1983;287:1339–41.
32 Buchbinder R, Goel V, Bombardier C, Hogg-Johnson S. Classification systems of soft tissue disorders of the neck and upper limb: Do they satisfy methodological guidelines? J Clin Epidemiol 1996;49:141–9.
33 Waris P, Kuorinka I, Kurppa K. Epidemiological screening of occupational neck and upper limb disorders. Scand J Work Environ Health 1979;6:25–38.
34 Viikari-Juntura E. Neck and upper limb disorders among slaughterhouse workers: an epidemiologic and clinical study. Scand J Work Environ Health 1983;9:283–90.
35 Silverstein BA. The prevalence of upper extremity cumulative disorders in industry [Thesis]. Ann Arbor: The University of Michigan, Occupational Health and Safety, 1985.
36 Fleiss JL. The measurement and control of misclassification error. In: Fleiss JL, ed. *Statistical methods for rates and proportions.* Chichester: Wiley, 1981:140–53.
37 Ferry S, Silman A, Pritchard T, Keenan J, Croft P. The association between different patterns of hand symptoms and objective evidence of median nerve compression: A community-based survey. Arthritis Rheum 1998;41: 720–4.