



# VU Research Portal

## Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement

de Winter, A.F.; Jans, M.P.; Scholten, R.J.P.M.; Deville, W.L.; van Schaardenburg, D.; Bouter, L.M.

### **published in**

Annals of the Rheumatic Diseases  
1999

### **DOI (link to publisher)**

[10.1136/ard.58.5.272](https://doi.org/10.1136/ard.58.5.272)

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

de Winter, A. F., Jans, M. P., Scholten, R. J. P. M., Deville, W. L., van Schaardenburg, D., & Bouter, L. M. (1999). Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. *Annals of the Rheumatic Diseases*, 58, 272-277. <https://doi.org/10.1136/ard.58.5.272>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## EXTENDED REPORTS

## Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement

Andrea F de Winter, Marielle P Jans, Rob J P M Scholten, Walter Devillé, Dirkjan van Schaardenburg, Lex M Bouter

### Abstract

**Objectives**—To assess the interobserver agreement on the diagnostic classification of shoulder disorders, based on history taking and physical examination, and to identify the determinants of diagnostic disagreement.

**Methods**—Consecutive eligible patients with shoulder pain were recruited in various health care settings in the Netherlands. After history taking, two physiotherapists independently performed a physical examination and subsequently the shoulder complaints were classified into one of six diagnostic categories: capsular syndrome (for example, capsulitis, arthritis), acute bursitis, acromioclavicular syndrome, subacromial syndrome (for example, tendinitis, chronic bursitis), rest group (for example, unclear clinical picture, extrinsic causes) and mixed clinical picture. To quantify the interobserver agreement Cohen's  $\kappa$  was calculated. Multivariate logistic regression analysis was applied to determine which clinical characteristics were determinants of diagnostic disagreement.

**Results**—The study population consisted of 201 patients with varying severity and duration of complaints. The  $\kappa$  for the classification of shoulder disorders was 0.45 (95% confidence intervals (CI) 0.37, 0.54). Diagnostic disagreement was associated with bilateral involvement (odds ratio (OR) 1.9; 95% CI 1.0, 3.7), chronic complaints (OR 2.0; 95% CI 1.1, 3.7), and severe pain (OR 2.7; 95% CI 1.3, 5.3).

**Conclusions**—Only moderate agreement was found on the classification of shoulder disorders, which implies that differentiation between the various categories of shoulder disorders is complicated. Especially patients with high pain severity, chronic complaints and bilateral involvement represent a diagnostic challenge for clinicians. As diagnostic classification is a guide for treatment decisions, unsatisfactory reproducibility might affect treatment outcome. To improve the

reproducibility, more insight into the reproducibility of clinical findings and the value of additional diagnostic procedures is needed.

(Ann Rheum Dis 1999;58:272-277)

Shoulder disorders are associated with pain, restricted range of motion and disability, which in some cases may last for several years.<sup>1-4</sup> History taking and physical examination are the cornerstones of the diagnosis of shoulder disorders. Differentiation between various shoulder disorders might be an important prerequisite for effective treatment.<sup>5</sup> However, differential diagnosis of shoulder disorders is often difficult, as various extrinsic and intrinsic conditions may underlie shoulder pain.<sup>6</sup> The complexity of the diagnosis of shoulder disorders is illustrated by the lack of consensus on the appropriate diagnostic criteria<sup>5</sup> and the fact that several diagnostic classifications have been proposed.<sup>7-13</sup>

In 1990, the Dutch College of General Practitioners developed guidelines for the diagnosis and management of shoulder pain,<sup>11</sup> which are largely based on the concepts of Cyriax.<sup>12, 13</sup> According to Cyriax, the anatomical site of the lesion can be identified by a systematic examination of the shoulder joint and the cervical spine. Although his concepts are well known and widely used, formal evaluation is still scarce.

The reproducibility of a diagnostic classification determines, to a large extent, its usefulness for clinical practice and research. Existing data on the interobserver agreement on diagnostic classification of shoulder disorders have mainly been derived from small studies, and have yielded contradictory results.<sup>14-16</sup> Therefore, in this study the interobserver agreement of the classification of shoulder disorders was assessed in a large population, according to the diagnostic criteria recommended in the guidelines of the Dutch College of General Practitioners.<sup>11</sup> The second objective of this study was to identify the determinants of diagnostic disagreement.

Institute for Research in Extramural Medicine, Vrije Universiteit, Amsterdam, the Netherlands  
A F de Winter  
M P Jans  
R J P M Scholten  
W Devillé  
L M Bouter

Department of Epidemiology and Biostatistics, Faculty of Medicine, Vrije Universiteit, Amsterdam, the Netherlands  
W Devillé  
L M Bouter

Jan van Breemen Institute, Centre for Rheumatology and Rehabilitation, Amsterdam, the Netherlands  
D van Schaardenburg

Correspondence to:  
Mrs A F de Winter, Institute for Research in Extramural Medicine, Faculty of Medicine, Vrije Universiteit, Van der Boechorststraat 7, 1081 BT Amsterdam, the Netherlands.

Accepted for publication  
18 February 1999

Table 1 Diagnostic classification of shoulder disorders

Diagnostic categories	Main criteria
1 Capsular syndrome (adhesive capsulitis, arthrosis, frozen shoulder, etc)	Capsular pattern: proportionally larger passive restriction of external rotation to glenohumeral abduction and internal rotation. Pain in C5 dermatome
2 Acute bursitis	Restriction of active and passive abduction, severe pain in C5 dermatome. Acute onset, no evident trauma
3 Acromioclavicular syndrome	Restriction of horizontal adduction. Pain in area of acromioclavicular joint and/or C4 dermatome
4 Subacromial syndrome (chronic bursitis, tendinitis, rotator cuff tears)	Painful arc during abduction. Pain in C5 dermatome. No restriction in passive range of motion. At least one positive resistance test. Bursitis: variable/slight pain, no or slight muscle weakness Tendinitis: pain, no or slight muscle weakness. Cuff tears: little pain, moderate to severe muscle weakness
5 Rest group	
Unclear clinical picture	Signs and symptoms do not meet the criteria of one of the other diagnostic categories
Extrinsic causes	Extrinsic causes such as cervical spine disorders, thoracic outlet syndrome, shoulder-hand syndrome, referred pain
Other intrinsic causes	Glenohumeral instability
No shoulder problems	No signs and symptoms indicating intrinsic or extrinsic shoulder disorders*
6 Mixed clinical picture	Signs and symptoms do meet the criteria for two diagnostic categories†

\*Between the inclusion by the general practitioners, physicians or rheumatologists and the physical examination by the examiners recovery of the shoulder problems might have occurred. †If the signs and symptoms completely meet the criteria of one diagnostic category and to a lesser extent a second category, the first diagnostic category was chosen.

## Methods

### PATIENTS

During a 20 month period, all consecutive eligible patients with incident or prevalent shoulder pain were invited to participate in this study by 20 general practitioners, two physicians working in an orthopaedic practice, and 20 secondary care rheumatologists. Patients were eligible for participation if they gave informed consent, were between 18 and 75 years of age, and were sufficiently competent to complete questionnaires (for example, no dementia). Patients with shoulder problems attributable to neurological, vascular or internal disorders, systemic rheumatic diseases, fractures or dislocations were not invited to participate.

### DESIGN

Two examiners (MPJ and AFW), both physiotherapists with three years and 10 years of clinical experience, respectively, performed the diagnostic procedure, which consisted of standardised history taking, physical examination, and subsequent diagnostic classification. One of the examiners was leading the history taking in the presence of the other. Subsequently, both examiners independently performed a physical examination. In each case, the history taking examiner performed the first

physical examination, and within one hour the other examiner performed the second physical examination, after which each examiner independently registered the diagnosis. The sequence of the examiners was randomly assigned. Before the study, the physical examination was standardised and trained and the criteria for the diagnostic classification were established. Moreover, during a pilot study among four patients the feasibility of the diagnostic procedure was tested.

Before undergoing the diagnostic procedure, the participants completed several questionnaires. The examiners were blinded for the results, because the answers given by the participants might have influenced the diagnostic assessment of the shoulder complaints.

### DIAGNOSTIC PROCEDURE

Demographic characteristics (age, sex, ethnicity) and clinical characteristics (for example, cause, nature and duration of the shoulder complaints, previous episodes of shoulder complaints and comorbidity) were recorded by history taking. The physical examination consisted of assessment of the active movements of the neck, and the active, passive and resisted movements of the shoulder.<sup>12 13</sup> The clinical findings recorded included the presence or absence of

Table 2 Main characteristics of participating patients recruited in different settings

	General practice (n=75)	Orthopaedic practice (n=33)	Clinic for rheumatology and rehabilitation (n=93)	Total (n=201)
Female (%)	61	52	69	66
Mean age in years (SD)	44 (13)	49 (11)	50 (11)	48 (12)
Dominant shoulder affected (%)	57	67	43	54
Precipitating cause (%):				
unknown	43	24	41	39
strain/overuse	16	33	34	27
(minor) injury	17	12	12	14
sport injury	11	9	7	9
other causes	13	22	6	11
Previous episode(s) of shoulder pain (%)	40	30	43	40
Duration of current episode (%):				
≤3 months	45	9	18	26
3–6 months	13	12	20	16
6–12 months	19	27	22	21
>12 months	23	52	40	34
Sleep disturbances (%):				
unable to lie on the involved shoulder	44	39	55	51
waking up; cannot fall asleep	65	48	78	69
Pain at rest (%)	63	58	73	67
Mean pain score* (SD):				
at night	54 (32)	36 (32)	54 (27)	51 (30)
during the day	55 (27)	41 (27)	57 (23)	54 (26)
Mean SDQ score† (SD)	63 (25)	61 (24)	75 (19)	68 (23)

\*Visual analogue scale (0–100). †SDQ score (0–100).

Table 3 Interobserver agreement of the diagnostic classification of shoulder disorders

	Examiner A				Rest group	Mixed clinical picture	Total	$\kappa^*$ (95% CI)
	Capsular syndrome	Acute bursitis	Acromioclavicular syndrome	Subacromial syndrome				
Examiner B								
Capsular syndrome	32	—	—	2	2 (1)	10	46	0.63 (0.50, 0.76)
Acute bursitis	—	1	—	—	—	—	1	0.50 (-0.10, 1.0)
Acromioclavicular syndrome	2	—	2	1	2 (1)	—	7	0.24 (-0.06, 0.53)
Subacromial syndrome	1	—	3	57	12 (7)	9	82	0.56 (0.45, 0.68)
Rest group	5 (4)	2 (1)	2 (2)	5 (5)	22 (5)	3 (2)	39 (19)	0.39 (0.24, 0.54)
Mixed clinical picture	4	—	1	8	7 (6)	6	26	0.14 (-0.03, 0.30)
Total	44	3	8	73	45 (20)	28	201	0.45 (0.37, 0.54)

Between parentheses: number of cases scored as unclear clinical picture. \*Dichotomised and overall  $\kappa$ .

restriction of active or passive motion, range of motion (in degrees), presence or absence of a painful arc, presence or absence of a capsular pattern, degree of pain (none, slight, moderate, severe) and degree of muscle weakness (none, slight, moderate, severe). Subsequently, the shoulder complaints were classified into one of six diagnostic categories: capsular syndrome, acute bursitis, acromioclavicular syndrome, subacromial syndrome, rest group, and mixed clinical picture. Table 1 gives the main criteria for the six diagnostic categories. In addition, both examiners estimated independently the severity of the pain on a 100 mm visual analogue scale (VAS) ranging from 0 "no pain" to 100 "very severe pain". Detailed information on the diagnostic procedure is available on request from the first author.

#### QUESTIONNAIRES

All patients recorded the severity of their pain, both at night and during the day, in the preceding week on a VAS ranging from 0 "no pain" to 100 "very severe pain". Furthermore, they filled in the Shoulder Disability Questionnaire (SDQ), which consists of 16 questions pertaining to difficulties in performing various daily activities on the previous day.<sup>17, 18</sup> The total score ranges from 0 "no disability" to 100

"difficulty with all applicable items". Personality traits (anxiety, anger, depression and optimism) were measured by means of the Self-Assessment Questionnaire-Nijmegen (SAQ-N).<sup>19-22</sup>

#### ANALYSES

Percentage of agreement and Cohen's  $\kappa$ , including 95% confidence intervals (CI), were calculated to quantify the interobserver agreement.<sup>23, 24</sup> The  $\kappa$  statistic was computed for the overall classification of shoulder disorders in the six categories and for each diagnostic category separately (dichotomous  $\kappa$ s). Multivariate logistic regression analysis was applied to determine whether demographic and clinical characteristics, pain, functional status, and personality traits were determinants of overall diagnostic disagreement. Variables with  $p \leq 0.25$  for the  $\chi^2$  test were considered as candidates for multivariate logistic regression. The logistic regression model was fitted by backward selection of variables (removal criterion  $p > 0.10$ ). The predictive performance of the logistic model was assessed by means of the Hosmer-Lemeshow test (goodness of fit test; calibration) and the receiver operating characteristic (ROC) curve area (discrimination).<sup>25, 26</sup> Odds ratios (OR) and 95% CI were calculated.

Table 4 Determinants of diagnostic disagreement

Variables*	Univariate analysis†		Multivariate logistic regression analysis‡	
	OR	95% CI	OR	95% CI
Sex (male = 0, female = 1)	1.0	0.6, 1.9		
Age ( $\leq 55 = 0, > 55 = 1$ )	1.1	0.6, 2.2		
Ethnicity (white = 0, non-white = 1)	1.0	0.4, 2.1		
Bilateral involvement (no = 0, yes = 1)	2.2¶	1.2, 4.2	1.9	1.0, 3.7
Duration of complaints ( $\leq 6$ months = 0, > 6 months = 1)	1.9¶	1.0, 3.4	2.0	1.1, 3.7
Previous episode(s) of shoulder pain (no = 0, yes = 1)	0.7¶	0.4, 1.2		
Neck pain present (no = 0, yes = 1)	1.6¶	0.9, 2.9		
Normal activities possible (yes = 0, no = 1)	1.0	0.6, 1.8		
Able to lie on the involved shoulder (yes = 0, no = 1)	0.9	0.5, 1.6		
Waking up (no = 0, yes = 1)	1.6¶	0.9, 3.0		
Cannot fall asleep (no = 0, yes = 1)	1.5¶	0.8, 2.7		
Pain during rest (no = 0, yes = 1)	1.2	0.7, 2.1		
Pain at night (VAS score $\leq 76 = 0, > 76 = 1$ )‡	1.8¶	0.9, 3.4		
Pain during the day (VAS score $\leq 75 = 0, > 75 = 1$ )‡	1.7¶	0.9, 3.2		
Functional status (score on SDQ $\leq 87 = 0, > 87 = 1$ )	1.2	0.6, 2.3		
Severity of pain (VAS score $\leq 72 = 0, > 72 = 1$ )§	2.7¶	1.4, 5.2	2.7	1.3, 5.3
Depression (score on SAQ-N $\leq 41 = 0, > 41 = 1$ )	1.6¶	0.8, 3.3		
Anxiety (score on SAQ-N $\leq 47 = 0, > 47 = 1$ )	1.6¶	0.8, 3.3		
Anger (score on SAQ-N $\leq 21 = 0, > 21 = 1$ )	1.7¶	0.9, 3.2		
Optimism (score on SAQ-N $> 27 = 0, \leq 27 = 1$ )	0.8	0.4, 1.6		

\*As the continuous variables were not linearly related to disagreement, the value of the 75th percentile was used to form dichotomised sub-groups. †The dependent variable was defined to be 0 if the observers agreed on the diagnosis, and 1 if they did not. ‡Severity of pain assessed by the patient. §Severity of pain assessed by both examiners on a VAS. The 75th percentile of the mean score of both examiners was included in the analysis. ¶Variables with  $p < 0.25$  for the  $\chi^2$  test were considered as candidates for multivariate logistic regression.

#### Results

Table 2 shows the main characteristics of the 201 patients. The severity of the shoulder problem was greater in patients recruited by rheumatologists than in patients recruited in other settings, expressed by more frequent sleep disturbances and pain during rest, and a lower functional status (higher SDQ-score).

Table 3 presents the diagnostic classification of shoulder disorders according to both examiners. The percentage of agreement was 60%, and the overall  $\kappa$  was 0.45 (95% CI 0.37, 0.54). In cases of disagreement (81 patients; 40%), the examiners frequently disagreed on whether the shoulder pain should be classified as a distinct category or a mixed clinical picture (42 patients; 21%; dichotomous  $\kappa = 0.14$ ).

Univariate analyses showed that most indicators of high severity of the complaints were associated with disagreement. Multivariate logistic regression analysis revealed that bilateral involvement, long disease duration ( $> 6$  months), and high pain severity (mean pain score according to both examiners  $> 7.2$ ) were independently associated with diagnostic disagreement (table 4). The model fitted the data

well (goodness of fit test:  $p = 0.46$ ) and the discrimination was acceptable (area under the ROC curve: 0.68).

### Discussion

In this study, the interobserver agreement of the diagnostic classification of shoulder disorders was evaluated in a large number of consecutive patients ( $n=201$ ) with varying severity and duration of complaints. Only a moderate interobserver agreement was found (percentage of agreement = 60%,  $\kappa = 0.45$ ). Especially for patients with high pain severity, chronic shoulder complaints and bilateral involvement, it seems to be difficult to define the anatomical site of the lesion.

The patients for this study were enrolled in three different health care settings. An association between type of health care setting and diagnostic disagreement might be present when patients are more likely to be referred by general practitioners to other health care settings when shoulder disorders are difficult to diagnose. Before the study we expected that the disagreement would vary between the setting of the general practitioners (GPs) and the other two settings. However, the level of disagreement was similar for the patients recruited by the GPs (37.3% disagreement) and the clinic for rheumatology and rehabilitation (39.8%), whereas for the patients ( $n=33$ ) recruited in the orthopaedic practice a higher level of diagnostic disagreement was found (54.5%). We have no clear explanation for this finding. The results of the multivariate logistic regression showed the same pattern when the patients of the orthopaedic practice were excluded from the analysis (data not shown).

In interobserver agreement studies varying experience of examiners might influence the level of reproducibility.<sup>27</sup> In this study the examiners had a different level of clinical experience and this could have influenced the results. To lessen this influence, before the study, the examiners had already achieved theoretical consensus and the diagnostic procedure had been standardised. In routine daily practice less attention will be paid to standardisation of the physical examination. Therefore, it might be argued that, on average, the level of reproducibility in clinical practice would not reach the level in this study.

One earlier study also evaluated interobserver agreement on the basis of the diagnostic classification recommended in the guidelines of the Dutch College of General Practitioners.<sup>14</sup> In this study, which compared the diagnosis of general practitioners with the diagnosis of physiotherapists in routine daily practice, frequent discrepancies were found ( $\kappa = 0.32$ ). This was a disappointing result, considering the fact that the physiotherapists were not blinded for the diagnoses of the GPs.

Two other interobserver studies based on a slightly different diagnostic classification show varying results. In the study carried out by Bamji *et al.*,<sup>15</sup> three consultant rheumatologists agreed on the diagnosis of shoulder disorders in less than 50% of the cases involved. In contrast, high agreement ( $\kappa = 0.88$ ) between

experienced physical therapists who examined 19 patients was found by Pellecchia *et al.*<sup>16</sup> However, the clinical characteristics of their study population were not presented, so it is unclear whether the study population consisted of consecutive patients or a selected group of patients. Therefore, it is difficult to establish an explanation for the high level of reproducibility found in their study.

The unsatisfactory reproducibility reported in the various studies might be explained by the fact that the diagnostic categories are insufficiently mutually exclusive. If clinical findings are not clearly attributable to one single diagnostic category, clinicians have to decide which clinical findings are most prominent to differentiate between shoulder disorders. In this study, the examiners frequently had difficulties in classifying the shoulder disorders into distinct categories, given the number of cases classified as "mixed" or "unclear clinical picture". Based on the same diagnostic classification, Sobel and Winters<sup>28</sup> showed that with strict application of the criteria only 3% of the cases could be distinctly classified, whereas with less stringent application of the criteria and additional tests 50% of the cases could be classified into distinct categories.

Insufficient mutual exclusiveness might also explain why patients with high pain severity, chronic complaints, and bilateral involvement represent a diagnostic challenge. These patients probably meet the diagnostic criteria for more than one category. This is understandable, because for patients with severe pain many of the test results will be positive, which makes it difficult to assess the relation between local factors and complaints. The complexity will also increase if various shoulder disorders, or a combination of a shoulder disorders with extrinsic conditions, underlie the shoulder complaints. This increased complexity might explain why it is more difficult to classify patients with bilateral involvement and chronic complaints. Moreover, for chronic complaints it has been suggested that local factors might determine the initial location of complaints, but that reasons for persistence and recurrence may be more general, such as previous episodes and psychosocial factors.<sup>29</sup>

What efforts should be made to improve the reproducibility of diagnostic classification of shoulder disorders? It has been postulated that diagnostic injections with a local anaesthetic<sup>30</sup> or additional tests during physical examination, such as the Neer impingement test,<sup>7, 28</sup> are helpful in establishing a diagnosis. If those diagnostic procedures do offer a solution is unclear. Additional skills will be required to perform certain diagnostic procedures, and some procedures might result in an increase in patient discomfort. It can also be questioned whether imaging techniques, such as ultrasound and magnetic resonance imaging (MRI), are beneficial in the selection of a diagnosis. Ultrasound might be an important diagnostic procedure because it is non-invasive and the costs are low.<sup>31</sup> Recently, a meta analysis was conducted to assess the

accuracy and reliability of ultrasound for shoulder disorders.<sup>32</sup> After evaluation of 58 studies the authors concluded that the accuracy of ultrasound was acceptable and that the reliability of ultrasound is unknown. However, often the accuracy of ultrasound was assessed for the detection of partial or complete tears of the rotator cuff. Consequently, it remains unclear what the accuracy of ultrasound will be in a population with varying severity of shoulder complaints. Although in the medical literature MRI is considered to be a useful diagnostic procedure in the evaluation of shoulder pain,<sup>33-35</sup> this procedure is time consuming and would increase health care costs.

It has been suggested that the exact localisation of the anatomical site of the lesion is a prerequisite for effective treatment.<sup>5 12 13</sup> In most randomised clinical trials, the main selection criterion for patients is the diagnosis, based on history taking and physical examination.<sup>5</sup> It can be questioned whether unsatisfactory treatment outcome in some patients is because of the difficulties involved in localising the lesion. Therefore, future research should demonstrate whether additional diagnostic procedures could increase reproducibility, and thereby also improve the outcome of treatment.

A less complicated diagnostic classification system is also proposed, to reduce the complexity of diagnosing shoulder disorders.<sup>36</sup> In our study the reproducibility was assessed for a diagnostic classification that was based on the concepts of Cyriax. Other diagnostic classifications can be based on different diagnostic criteria and have a different reproducibility. Unfortunately, there are no studies that report on the reproducibility of the various clinical findings that underlie the diagnostic classification of shoulder disorders. It has, however, been shown that various diagnostic labels are applied, even when there is consensus on the clinical findings.<sup>15</sup> More insight into the reproducibility of clinical findings and careful examination of the diagnostic criteria is obviously needed before new classification systems can be adopted.

In conclusion, distinguishing between distinct shoulder disorders on the basis of history taking and physical examination seems to be rather complicated. Especially patients with high pain severity, bilateral involvement, and chronic complaints represent a diagnostic challenge. Serious doubt about the reproducibility of the diagnostic classification of shoulder disorders raises the question whether diagnosis based on history taking and physical examination is actually beneficial in the choice of treatment. Future studies should therefore determine whether additional diagnostic procedures improve diagnostic agreement. Moreover, additional research is needed to investigate the sources of diagnostic disagreement attributable to interobserver differences in clinical findings. This might be helpful in reaching further consensus on the appropriate diagnostic criteria.

AFW designed the study, was responsible for data collection and analysis, and is the first author of the manuscript. MPJ contributed to the design of the study and was also involved in data collection. RJP, who was Project Leader of the study, supervised the design and conduct of the study, together with DvS. WD supervised the data analysis. LMB initiated the study and provided overall supervision. All authors contributed to the revision of the manuscript. Funding: partially financed by the "Stichting Anna-Fonds" foundation.

- 1 Croft P, Pope D, Silman A. The clinical course of shoulder pain; prospective cohort study in primary care. *BMJ* 1996;313:601-2.
- 2 Windt DAWM van der, Koes BW, Boeke AJP, Devillé W, Jong BA de, Bouter LM. Shoulder disorders in general practice: prognostic indicators of outcome. *Br J Gen Pract* 1996;46:519-23.
- 3 Vecchio PC, Kavanagh RT, Hazleman BL, King RH. Community Survey of shoulder disorders in the elderly to assess the natural history and effects of treatment. *Ann Rheum Dis* 1995;54:152-4.
- 4 Chard MD, Satelle LM, Hazleman BL. The long-term outcome of rotator cuff tendinitis: a review study. *Br J Rheumatol* 1988;27:385-9.
- 5 Green S, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. *BMJ* 1998;316:354-60.
- 6 Neviasser RJ. Painful conditions affecting the shoulder. *Clin Orthop* 1983;173:63-9.
- 7 Neer CS. Impingement lesions. *Clin Orthop* 1983;173:70-7.
- 8 Uthoff HK, Sarkar K. An algorithm for shoulder pain caused by soft-tissue disorders. *Clin Orthop* 1990;254:121-7.
- 9 Uthoff HK, Sarkar K. Classification and definition of tendinopathies. *Clin Sports Med* 1991;10:707-20.
- 10 Hedtmann A, Fett H. The so-called periartropathia humeroscapularis. Classification and analysis of 1266 cases. *Z Orthop* 1989;127:643-9.
- 11 Bakker JF, Jongh L de, Jonquiere M, Mens J, Oosterhuis WW, Poppelaars A, et al. *Standaard schouderklachten*. Huisarts en Wetenschap 1990;33:196-202.
- 12 Cyriax J. *Textbook of orthopaedic medicine*. 7th ed. London: Bailliere Tindall, 1981:190-239.
- 13 Cyriax J. The shoulder. *Br J Hosp Med* 1975;19:185-92.
- 14 Liesdek C, Windt DAWM van der, Koes BW, Bouter LM. Soft-tissue disorders of the shoulder: a study of the inter-observer agreement between general practitioners and physiotherapists and an overview of physiotherapeutic treatment. *Physiotherapy* 1997; 83:12-17.
- 15 Bamji AN, Erhardt CC, Price TR, Williams PL. The painful shoulder: can consultants agree? *Br J Rheumatol* 1996; 35:1172-4.
- 16 Pellicchia GL, Paulino J, Connell J. Intertester reliability of the Cyriax evaluation in assessing patients with shoulder pain. *J Orthop Sports Phys Ther* 1996;23:34-8.
- 17 Windt DAWM van der, Heijden GJMG van der, Winter AF de, Koes BW, Devillé W, Bouter LM. The responsiveness of the Shoulder Disability Questionnaire. *Ann Rheum Dis* 1998;57:82-7.
- 18 Heijden GJMG van der, Leffers P, Bouter LM. Shoulder Disability Questionnaire: design and responsiveness of a functional status measure. *J Clin Epidemiol* (in press).
- 19 Spielberger CD. *Manual for state trait anxiety inventory*. Form Y. Palo Alto, CA: Consulting Psychologists Press, 1983.
- 20 Spielberger CD, Jacobs G, Russell S, Crane RS. Assessment of anger: State Trait Anger Scale. In: Butcher JN, Spielberger CD, eds. *Advances in personality assessment*. Vol 2. Hillsdale, NJ: Erlbaum, 1983:159-87.
- 21 Zung WWK. A self-rating depression scale. *Arch Gen Psychiatry* 1965;12:63-70.
- 22 Scheier MF, Carver CS. Optimism, coping and health: assessment and implications of generalized outcome expectancies. *Health Psychol* 1985;4:219-47.
- 23 Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;304:1491-4.
- 24 Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37-46.
- 25 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley, 1989.
- 26 Miller EM, Hui SL. Validation techniques for logistic regression models. *Stat Med* 1991;10:1213-26.
- 27 Balich SM, Sheley RC, Brown TR, Sauser DD, Quinn SF. MR imaging of the rotator cuff tendon: interobserver agreement and analysis of interpretative errors. *Radiology* 1997;204:191-4.
- 28 Sobel JS, Winters JC. Shoulder complaints in general practice. In: *Shoulder disorders in general practice. Diagnosis and treatment*. [PhD thesis]. Amsterdam: Meditext, 1996:15-23.
- 29 Croft P. The epidemiology of pain: the more you have, the more you get. *Ann Rheum Dis* 1996;55:859-60.
- 30 Larson HM, O'Connor FG, Nirschl RP. Shoulder pain: the role of diagnostic injections. *Am Fam Physician* 1996;53:1637-43.
- 31 Grechenig W, Clement HG, Schatz B, Grechenig M. Stellenwert der Sonographie am Stütz- und Bewegungsapparat—unter besonderer Berücksichtigung von strahlenbelastung und kostenreduktion. *Biomed Tech (Berl)* 1997;42:132-7.

- 32 Geraets JJXR, Stefelmans RJPAM. Accuracy and reliability of shoulder sonography, a meta-analysis [in Dutch]. Maastricht: Unigraphic, 1998.
- 33 Tsao LY, Mirowitz SA. MR imaging of the shoulder; imaging techniques, diagnostic pitfalls, and normal variants. *Magn Reson Imaging Clin N Am* 1997;5:683-704.
- 34 Fritz RC, Stoller DW. MR imaging of the rotator cuff. *Magn Reson Imaging Clin N Am* 1997;5:735-54.
- 35 Deutsch AL, Klein MA, Mink JH, Mandelbaum BR. MR imaging of miscellaneous disorders of the shoulder. *Magn Reson Imaging Clin N Am* 1997;5:881-95.
- 36 Winters JC, Groenier KH, Sobel JS, Arendzen HH, Meyboom-de Jongh B. Classification of shoulder complaints in general practice by means of cluster analysis. *Arch Phys Med Rehabil* 1997;78:1369-74.