



Clinical prediction rules for the prognosis of shoulder pain in general practice

Ton Kuijpers^{a,b,c,*}, Daniëlle A.W.M. van der Windt^a, A. Joan P. Boeke^a,
Jos W.R. Twisk^d, Yvonne Vergouwe^e, Lex M. Bouter^a,
Geert J.M.G. van der Heijden^e

^a Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, The Netherlands

^b Department of Allied Health Care Research, Amsterdam School of Allied Health Care Education, The Netherlands

^c Care and Public Health Research Institute, Maastricht University, The Netherlands

^d Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

^e Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands

Received 10 January 2005; received in revised form 13 September 2005; accepted 14 November 2005

Abstract

Shoulder pain is common in primary care and has an unfavourable outcome in many patients. Information about predictors of outcome is scarce and inconsistent. The objective of this study was to develop clinical prediction rules for calculating the absolute risk of persistent shoulder symptoms for individual patients, 6 weeks and 6 months after the first consultation in general practice. A prospective cohort study with 6 months follow-up was carried out in three geographic areas in The Netherlands. In this study, 587 patients with a new episode of shoulder pain were included. The main outcome measure was persistent symptoms at 6 weeks and 6 months, perceived by the patient. Potential predictors included the results of a physical examination, sociodemographic variables, disease characteristics (duration of symptoms, pain intensity, disability and comorbidity), physical activity, physical workload and psychosocial factors. Response rates to the follow-up questionnaires were 83% at 6 weeks and 92% at 6 months. A longer duration of symptoms, gradual onset of pain and high pain severity at presentation were consistently associated with persistent symptoms at 6 weeks and 6 months. The discriminative validity of our prediction rules was satisfactory with area under the curves of 0.74 (95% CI 0.70, 0.79) at 6 weeks and 0.67 (95% CI 0.63, 0.71) at 6 months. The performance of our rules needs to be tested in other populations of patients with shoulder pain to enable valid and reliable use of the rules in everyday clinical practice.

© 2005 International Association for the Study of Pain. Published by Elsevier B.V. All rights reserved.

Keywords: Musculoskeletal disorders; Shoulder; Prognosis; Psychosocial factors; Cohort study; Clinical prediction rule

1. Introduction

Shoulder pain is common with a one-year prevalence ranging between 5% and 47% (Bot et al., 2005; Luime et al., 2004; Picavet and Schouten, 2003; Bongers, 2001; Van der Heijden, 1999). The point prevalence in

the general population in The Netherlands has recently been estimated at 17% (Van der Linden et al., 2005). The annual incidence of consultation for a new episode of shoulder pain in Dutch general practice ranges between 12 and 25/1000/year (Bot et al., 2005; Okkes et al., 1998; Van der Linden et al., 2005; Van der Windt et al., 1995).

Shoulder pain has an unfavourable outcome in many patients. Only about 50% of all new episodes of shoulder pain presented in primary care show complete recovery within six months (Croft et al., 1996; Van der Windt et al., 1996; Winters et al., 1999a,b), after one year this

* Corresponding author. Tel.: +31 20 652 1172; fax: +31 20 652 1141.

E-mail address: a.c.kuijpers@hva.nl (T. Kuijpers).

URLs: www.emgo.nl, www.hva.nl (T. Kuijpers).

proportion increases to only 60% (Van der Windt et al., 1996). Knowing more about the prognostic value of clinical, psychosocial and occupational factors in patients with shoulder pain will help to provide patients with adequate information regarding the most likely course of their symptoms. Such information may also support decisions regarding treatment and referral of patients.

In a systematic review of the literature, we summarised the available evidence from 16 studies regarding predictors of outcome of shoulder pain (Kuijpers et al., 2004). Only six studies were of relatively high quality. In a primary care population, strong evidence for predicting poor outcome was only found for ‘high pain intensity’. For any other variable, including psychosocial variables, convincing evidence for their predictive value is lacking. We performed a cohort study among patients with shoulder pain consulting their general practitioners (GPs) and followed them for 6 months. The objective of this study was to determine which combination of factors predicts the outcome of an episode of shoulder pain 6 weeks and 6 months after the first consultation in a general practice population. Our aim was to develop a clinical prediction rule for calculating the absolute risk of persistent symptoms for individual patients in general practice.

2. Methods

2.1. Recruitment

Between January 2001 and June 2003, 103 GPs recruited patients at first consultation for a new episode of shoulder pain in three geographic areas in the Netherlands (Amsterdam, Groningen and Maastricht). The primary reason for consultation had to be shoulder pain. In this study, shoulder pain was defined according to the 1999 version of the Dutch guidelines for shoulder complaints, issued by the Dutch College of General Practitioners (Winters et al., 1999a,b). In the guideline, shoulder pain is characterised as pain in the deltoid and upper arm region. GPs used this definition to select patients with shoulder pain for our study.

Patients were selected by their GP if they were 18 years or older of age and had not consulted their GP or received any form of treatment for the afflicted shoulder in the preceding 3 months. GPs were instructed to select consecutive patients. Sufficient knowledge of the Dutch language was required to complete written questionnaires. Exclusion criteria were severe physical or psychological conditions (i.e., fractures or luxation in the shoulder region; rheumatic disease; neoplasm; neurological or vascular disorders; dementia). Data collection was approved by the Medical Ethics Committee of the VU University Medical Center, Amsterdam, The Netherlands.

2.2. Management of shoulder pain

All patients received standardised treatment according to the 1999 version of the Dutch guidelines for shoulder complaints issued by the Dutch College of General Practitioners (Winters et al., 1999a,b). The guidelines recommend giving

information on the prognosis of shoulder pain, advice regarding provoking activities and stepwise treatment consisting of paracetamol, Non-Steroidal Anti-Inflammatory Drugs (NSAIDs), corticosteroid injection or referral for physiotherapy. The GP made the decision regarding the content of treatment based on duration and severity of pain and disability. The participating GPs were educated and trained to apply treatment according to this guideline.

2.3. Prognostic factors

Within 10 days after they had consulted the GP, participants gave written informed consent and completed an extensive baseline questionnaire. The questionnaire contained questions on sociodemographic variables, disease characteristics (i.e., pain intensity, disability, duration of complaints, onset and comorbidity), physical activity, physical workload and psychosocial factors. Patients were also physically examined by a trained assistant at baseline.

The sociodemographic variables and disease characteristics were measured using a checklist which mainly consisted of yes or no questions. Physical activity was measured with a single question (less/equally/more active than others). We measured physical workload with a self-constructed scale of 5 questions (yes/no) concerning pushing and pulling, lifting weights, working with hands above shoulder level and the use of vibrating tools on at least two days a week (total score 0–5, Cronbach’s $\alpha = 0.74$). Repetitive movements, on at least two days a week, were also measured with a single question answered with yes or no.

The psychosocial factors coping, anxiety, depression, somatisation, distress, fear-avoidance beliefs and kinesiophobia were measured with widely used standardised questionnaires. Coping was assessed with the 43-item Pain Coping and Cognition List (PCCL) (Berg et al., 2001), consisting of the subdomains catastrophising (1–6 points, Cronbach’s $\alpha = 0.63$), coping with pain (1–6 points, Cronbach’s $\alpha = 0.83$), internal (1–6 points, Cronbach’s $\alpha = 0.76$) and external loci of control (1–6 points, Cronbach’s $\alpha = 0.65$). Anxiety (0–24 points, Cronbach’s $\alpha = 0.77$), depression (0–12 points, Cronbach’s $\alpha = 0.90$), somatisation (0–32 points, Cronbach’s $\alpha = 0.82$) and distress (0–32 points, Cronbach’s $\alpha = 0.92$) were measured with the 50-item Four-Dimensional Symptom Questionnaire (4DSQ) (Terluin, 1996; Terluin et al., 2004). Fear-avoidance beliefs were assessed using the 4-item physical activity subscale of the Fear-Avoidance Beliefs Questionnaire (FABQ; 0–24, Cronbach’s $\alpha = 0.73$) (Vendrig et al., 1998; Waddell et al., 1993). Kinesiophobia, finally, was measured using two items (no. 1 and no 9.) of the Tampa Scale for Kinesiophobia (TSK; 0–12 points, Cronbach’s $\alpha = 0.82$) (Kori et al., 1990; Vlaeyen et al., 1999). The questionnaire also included a general one-item question regarding the presence (yes/no) of any psychological problems (e.g. distress, depression and anxiety).

Functions of the shoulder joint and cervicothoracic spine were tested during a physical examination. For the glenohumeral joint active and passive abduction, passive external rotation (Bergman et al., 2004a) and shoulder impingement (Neer, 1983) were tested. Two alternative functional tests, HIB (Hand-in-back) and HIN (Hand-in-neck) (Solem-Bertoft et al., 1996; Westerberg et al., 1996) measured on a 7-point

scale (score 0 = very poor range of motion, score 7 = full range of motion), were performed as well. The assistant estimated the range of motion in degrees (°).

During all mobility tests, self-reported pain was assessed on a 4-point scale (0 = no pain; 3 = severe pain). A factor analysis on the results of a physical examination in a similar population of patients with shoulder pain resulted in four factors: shoulder mobility, shoulder pain, neck mobility and neck pain (Bergman et al., 2004a).

The factor ‘shoulder mobility’ consisted of 6 mobility tests: HIB, HIN, active abduction, passive abduction, external rotation and impingement. For calculation of the sum score (0–18 points) variables were recoded into a 4-point scale, with 0 reflecting full range of motion and 3 points reflecting very poor range of motion. HIB/HIN scores were recoded as: score 7 = 0; scores 5 and 6 = 1; scores 3 and 4 = 2; scores 1 and 2 = 3. Abduction (active and passive) was recoded as 170–180° = 0; 140–170° = 1; 90–140° = 2; 0–90° = 3. External rotation was recoded as: >80° = 0; 70–80° = 1; 50–70° = 2; <50° = 3. During the impingement test, pain was measured (0 = no pain; 3 = severe pain). The factor ‘shoulder pain’ (0–18 points) consisted of the sum of the pain scores during the mobility tests.

The factor ‘neck mobility’ (0–4 points) consisted of rotation of the cervicothoracic spine in neutral, flexed and extended position, and lateral bending. These range of motion tests were scored as (1 = decreased range of motion and 0 = no decreased range of motion). The factor ‘neck pain’ (0–18 points) consisted of the sum of the pain scores during flexion and extension of the neck, rotation in a neutral, flexed and extended position, and lateral bending.

2.4. Outcome measurements

The outcome was measured by postal questionnaires at 6 weeks, 3 and 6 months. We restricted the length of the follow-up period to 6 months. Previous studies in primary care have shown that most recovery occurs in this period (Croft et al., 1996; Van der Windt et al., 1996; Winters et al., 1999a,b). This means that little information can be gained after 6 months follow-up. Our primary outcome measure ‘Patient perceived recovery’ was measured on an 8-point scale. Patients who did not report full recovery or very much improvement were denoted as having “persistent symptoms” (Bergman et al., 2004b; Van der Windt et al., 1998). Patients who did not reply at 6 weeks were re-contacted at 3 and 6 months. Secondary outcome measures were shoulder disability, measured with the 16-item shoulder disability questionnaire (SDQ; 0–100) (Van der Heijden et al., 2000), pain (0–10 numeric rating scale) (Van der Windt et al., 1998) and severity of the main complaint (0–10 numeric rating scale) (Beurskens et al., 1999). We studied the relationship between our primary and secondary outcome measures to determine if patients with persistent symptoms after 6 weeks and 6 months showed higher levels of pain and disability.

2.5. Analysis

Missing values of patient characteristics were imputed (approximately 1% of all required values at both 6 weeks and 6 months). Imputation was based on the correlation between each variable with missing values with the other patient characteristics. Univariable logistic regression analyses

were performed for all potential prognostic indicators with our primary outcome measure, i.e., persistent symptoms, at either short term (6 weeks) or long term (6 months). The linearity of the associations of continuous variables with outcome was studied. Factors were categorised if they did not show a linear association with the outcome. Variables that had a statistically significant association with the outcome (p -value ≤ 0.20) were selected as candidate predictors for the multivariable analysis. Not more than one independent variable per 10 events was included in the multivariable analysis (Harrell et al., 1996; Altman, 1991). We presented the univariable ORs along with the 95% confidence intervals as well as with p -values to enable the reader to choose alternative statistical significance levels for the selection of variables for the multivariable analysis.

Separate prediction models were developed for persistent symptoms at short term and long term. A second selection step was performed in the multivariable model, that contained all candidate predictors with stepwise backward selection. Variables with the lowest predictive value were deleted from the model until further elimination of a variable resulted in a statistically significant lower model fit estimated with the log-likelihood ratio test ($p < 0.20$).

Bootstrapping techniques were used to study the internal validity of the final prediction model, i.e., to adjust the estimated regression coefficients for overfitting and the model performance for overoptimism (Efron and Tibshirani, 1993; Harrell et al., 1996). The model’s performance obtained after bootstrapping can be considered as the performance that can be expected in similar future patients. Random bootstrap samples were drawn with replacement (100 replications) from the full data set. The multivariable selection of variables was repeated within each bootstrap sample. All analyses were performed using S-plus 6.1 (Insightful Corp., Seattle, WA, USA).

2.6. Evaluation of the model

The reliability of the multivariable model was determined by use of the Hosmer–Lemeshow goodness-of-fit statistic (Hosmer and Lemeshow, 1989). Calibration of the model predictions was assessed by plotting the predicted individual probabilities against the observed individual probabilities for persistent symptoms. For this, patients were grouped into deciles according to their predicted probability for persistent symptoms according to the model. The prevalence of the endpoint within each decile represents the observed individual probability. The area under the receiver-operating characteristic curve (ROC) was used to assess the discriminative ability of the model. The ROC-curve plots the true-positive rate (sensitivity) against the false-positive rate (1 – specificity) at any given cut-off value. The curve illustrates the ability of the model to discriminate between patients with and without persistent symptoms at subsequent cut-off points along the range of the predicted probabilities. An area under the curve (AUC) of 0.5 indicates no discrimination above chance, whereas an AUC of 1.0 indicates perfect discrimination.

2.7. From a prediction model to an individual patient’s risk

We developed a clinical prediction rule (Wasson et al., 1985; Wasson and Sox, 1996; Laupacis et al., 1997) for outcome at 6 weeks and 6 months, to provide an estimate for

individual patients of their absolute risk of persistent symptoms. The probability (P) of persistent symptoms was predicted by $P = 1/[1 + \exp - (a_0 + b_1x_1 + \dots + b_jx_j)]$. The status of a patient for any dummy or binary variable included in the prediction rule can be either 0 or 1, while for a (semi) continuous variable it takes the actual observed value.

2.8. Score charts

To facilitate the calculation of an individual patient's risk, we developed score charts. We divided the regression coefficients by the lowest coefficient and rounded them off to the nearest integer to form the scores for the predictors. The sum of the scores corresponded to a risk of poor outcome.

3. Results

3.1. Study population and follow-up

At baseline 587 patients were questioned and physically examined. Table 1 lists the baseline characteristics of the participants. At 6 weeks 487 (83%) and at 6 months 538 (92%), patients returned the postal questionnaire. The drop-outs at 6 weeks and 6 months were younger than the responders (mean difference being 4 years and 6 years, respectively). The drop-outs at 6 months showed more often an acute onset (49% versus 36%) and less repetitive movements in their work (26% versus 36%) at baseline in comparison with the responders.

At 6 weeks 70% ($n = 340$) and at 6 months, 46% ($n = 249$) patients reported persistent symptoms. Of these 249 patients, only 22 reported that symptoms had recurred after initial recovery at 6 weeks. Table 2 shows that patients with persistent symptoms reported also more pain, more shoulder disability and higher severity of the main complaint.

3.2. Management of shoulder pain

At baseline most patients ($n = 423$, 72%) received a wait-and-see policy, paracetamol, or NSAIDs. Furthermore, 68 patients (12%) received an injection with corticosteroid, 58 patients (10%) were referred for physiotherapy and 28 patients (6%) received other therapies.

3.3. Prognostic factors

Table 1 also presents the univariable association of potential predictors with outcome at 6-weeks and 6-months follow-up. Given the fact that median baseline scores on distress, anxiety and somatisation were very low, scores on these psychological factors were dichotomised. Variables, which showed a univariable association ($p \leq 0.20$), were selected for the backward stepwise selection analysis. Table 3 presents the vari-

ables included in the prediction models for persistent symptoms at 6 weeks and 6 months after backward stepwise selection ($p \leq 0.20$). A longer duration of symptoms at baseline, gradual onset of shoulder complaints and higher pain intensity were associated with a poorer prognosis at both 6 weeks and 6 months. Furthermore, concomitant psychological complaints, repetitive movements and increasing neck pain scores at physical examination were associated with persistent symptoms at 6 weeks. A poor prognosis at 6 months was additionally predicted by concomitant back pain and increasing shoulder pain scores at physical examination.

3.4. Evaluation of the models

The reliability of the models was adequate, according to the Hosmer–Lemeshow statistic, with a p -value of 0.51 for the model at 6 weeks and 0.16 at 6 months. Fig. 1 shows the calibration of the predictions. The predicted and observed probabilities are rather close to the 45° line, demonstrating good calibration of the predictions by the two models. The AUCs for the models at 6 weeks and 6 months were 0.74 (95% CI 0.70–0.79) and 0.67 (95% CI 0.63–0.71). The predicted risks of persistent symptoms are widely distributed (Fig. 2).

3.5. Score charts

Fig. 3 shows the score charts for calculating the risk of persistent symptoms at the short and long term. For instance, a patient with shoulder complaints for 3 weeks at baseline with a gradual onset of symptoms, and a shoulder pain score of 1 point, has a prognostic score of 8 points for the short term and 12 points in the long term, which implies 40–50% risk of persistent symptoms at 6 weeks and 20–30% at 6 months.

4. Discussion

This is the first prospective cohort study on shoulder pain, in which a score chart is developed that may be used by GPs to calculate the risk of persistent symptoms for individual patients. Duration of complaints, gradual onset and pain intensity were strong predictors for both short- and long-term prognosis.

4.1. Prognostic factors

In a systematic review (Kuijpers et al., 2004) of the literature we found only strong evidence for 'high pain intensity' as a predictor of poor outcome. In our study, high pain intensity was also found to be a strong predictor of persistent symptoms at short-term (6 weeks) and long-term (6 months) follow-up. The results of our analyses showed somewhat different sets of predictors for short- and long-term results, but both analyses

Table 1
Baseline characteristics of patients with shoulder pain ($n = 587$) and univariable associations with persistent symptoms at 6 weeks and 6 months

Variable	n (%)	6 weeks			6 months		
		OR	95% CI	p^a	OR	95% CI	p^a
<i>Demographic</i>							
Age (years); mean (SD)	51 (14)	1.0	1.0, 1.0	0.32	1.0 ^c	1.0, 1.0	0.29
Gender: male	292 (50)	1.3	0.9, 2.0	0.14	0.9	0.6, 1.3	0.65
Education				0.04			0.12
Low ^b	210 (36)						
Middle	234 (40)	1.0	0.6, 1.6		0.7	0.5, 1.0	
High	135 (23)	0.6	0.3, 0.9		0.7	0.4, 1.0	
<i>Disease characteristics</i>							
Duration of complaints				<0.001			<0.001
0–6 weeks ^b	205 (35)						
7–12 weeks	139 (24)	2.3	1.4, 3.9		1.8	1.2, 2.9	
>3 months	242 (41)	5.4	3.3, 8.9		3.5	2.3, 5.2	
Gradual onset (versus acute)	363 (62)	2.9	1.9, 4.3	<0.001	2.2	1.5, 3.0	<0.001
Precipitating cause							
Unexpected movement	33 (6)	0.8	0.4, 1.8	0.59	1.3	0.6, 2.8	0.43
Strain/overuse: unusual activities	96 (16)	3.6	1.1, 12.2	0.21	0.6	0.4, 1.0	0.04
Strain/overuse: usual activities	138 (24)	1.8	1.1, 3.1	0.02	1.3	0.8, 1.9	0.25
Injury	33 (6)	3.6	1.1, 12.2	0.03	2.4	1.1, 5.5	0.03
Sport injury	29 (5)	1.2	0.5, 3.2	0.67	0.9	0.4, 2.0	0.82
Unknown	239 (41)	0.7	0.4, 1.0	0.03	0.9	0.7, 1.3	0.67
Shoulder complaints in the past	348 (62)	1.3	0.9, 2.0	0.16	1.3	0.9, 1.9	0.10
Neck complaints in the past	296 (51)	1.9	1.3, 2.8	0.00	1.4	1.0, 2.0	0.04
Dominant side involved	362 (62)	1.3	0.9, 1.9	0.24	1.2	0.9, 1.7	0.29
Comorbid psychological complaints	55 (9)	3.3	1.3, 8.7	0.01	1.5	0.8, 2.7	0.19
Concomitant musculoskeletal complaints							
Neck/high back	209 (36)	1.7	1.1, 2.6	0.01	1.6	1.1, 2.2	0.01
Low back pain	139 (24)	1.5	0.9, 2.3	0.13	2.2	1.5, 3.3	<0.001
Upper extremity	174 (30)	2.0	1.2, 3.8	<0.001	1.7	1.2, 2.4	0.01
Lower extremity	177 (30)	1.2	0.8, 1.9	0.44	1.7	1.2, 2.5	<0.001
Shoulder pain (0–10); mean (SD)	4.8 (2.3)	1.3	1.1, 1.4	<0.001	1.2	1.1, 1.3	<0.001
Shoulder disability (0–100); mean (SD)	59.9 (24.2)	1.0	1.0, 1.0	<0.001	1.0	1.0, 1.0	<0.001
<i>Physical examination</i>							
ROM shoulder (0–18); mean (SD)	6.8 (4.3)	1.1	1.0, 1.2	0.01	1.7 ^c	1.0, 2.2	0.22
Pain shoulder with movement (0–18); median (IQR)	4 (2–4)	1.1	1.1, 1.7	<0.001	1.1	1.1, 1.2	<0.001
ROM neck (0–4); median (IQR)	0 (0–0)	1.1	1.0, 1.3	0.10	0.9 ^c	0.6, 1.4	0.53
Pain neck with movement (0–18); median (IQR)	0 (0–0)	1.2	1.1, 1.3	<0.001	1.1	1.0, 1.2	0.01
<i>Physical factors</i>							
Dynamic physical workload (0–5); median (IQR)	1 (1–2)	1.2	1.0, 1.4	0.02	1.0 ^c	0.6, 1.5	0.11
Repetitive movements	384 (65)	2.1	1.4, 3.1	<0.001	1.2	0.8, 1.7	0.33
Physical activity in comparison to others				0.03			0.20
More active ^b	126 (39)						
Equally active	245 (42)	0.6	0.4, 0.9		0.9	0.6, 1.3	
Less active	110 (19)	1.2	0.7, 2.2		1.4	0.9, 2.3	
<i>Psychosocial factors</i>							
Coping (mean, SD)							
Catastrophising (1–6)	2.2 (0.8)	1.4	1.1, 1.8	0.02	1.4 ^c	0.7, 2.6	0.42
Coping with pain (1–6)	3.1 (1.0)	1.0 ^c	0.2, 4.2	0.96	2.2 ^c	0.6, 9.0	0.21
Internal locus of control (1–6)	3.3 (0.9)	0.8 ^c	0.2, 3.9	0.45	1.4 ^c	0.4, 5.3	0.62
External locus of control (1–6)	3.2 (0.9)	0.6 ^c	0.1, 3.1	0.73	1.3 ^c	0.7, 1.3	0.32
4DSQ (median, IQR)							
Distress (0–32)	0 (0–0)	2.2 ^d	0.7, 6.6	0.15	2.6 ^d	1.2, 5.8	0.02
Depression (0–12)	0 (0–0)	3.0 ^c	0.4, 25.2	0.52	3.0 ^c	0.6, 15.4	0.36
Anxiety (0–24)	0 (0–0)	1.8 ^d	0.4, 8.0	0.46	1.2 ^c	0.2, 5.9	0.85
Somatisation (0–32)	0 (0–2)	5.2 ^d	1.2, 22.4	0.03	2.5 ^d	1.1, 5.4	0.02
Fear-avoidance (0–24); mean (SD)	14.1 (5.6)	1.0	1.0, 1.1	0.74	1.0	1.0, 1.0	0.71
Kinesiophobia (0–12); median (IQR)	2 (0–2)	0.9 ^c	0.4, 2.0	0.99	1.5 ^c	0.7, 3.0	0.26

SD = standard deviation; IQR = inter-quartile range; ROM = range of motion; 4DSQ = four-dimensional symptom questionnaire.

^a Variables with a univariable p -value ≤ 0.20 were selected for the multivariable analysis of persistent symptoms at 6 weeks and 6 months.

^b Reference category.

^c In case of non-linear associations, continuous variables were divided into categories. The table presents the Odds Ratio (OR) for the highest versus lowest category.

^d Variable was dichotomised.

Table 2
Secondary outcome measures for patients with and without persistent symptoms at 6 weeks and 6 months

Outcome measures; mean (SD)	6 weeks		6 months	
	Persistent symptoms		Persistent symptoms	
	Yes	No	Yes	No
Pain (0–10)	4.3 (2.1)	0.5 (0.9)	4.1 (2.3)	0.4 (1.1)
Shoulder disability (SDQ) (0–100)	53.0 (25.5)	10.4 (16.6)	52.2 (26.7)	5.9 (14.5)
Severity of main complaint (0–10)	4.8 (2.6)	0.8 (1.5)	5.0 (2.8)	0.6 (1.3)

demonstrated that duration and severity of symptoms (disease characteristics) were more important in predicting outcome than physical or psychosocial factors. It has previously been suggested that psychosocial factors such as dysfunctional pain cognition or mistaken beliefs about pain and inappropriate pain behaviour are likely to predict a poor outcome of painful musculoskeletal conditions (Van der Heijden, 1999). The association between psychosocial factors and musculoskeletal pain has been established in patients with chronic pain syndromes. The scores on all psychosocial variables measured in our population were low. Although significant univariable associations with persistent symptoms at 6 weeks were found in this study for several psychosocial factors (distress, somatisation, catastrophising), in a multivariable model these factors had little to add to a simple yes or no question about the presence of psychological complaints. For the applicability of the prediction rules in primary care this is an advantage,

as easy-to-measure predictors are preferred to more time-consuming ones.

4.2. Management of shoulder pain

We did not include treatment in the model, as we assumed that confounding by indication could influence our findings. Patients with more severe symptoms and thus, probably a poorer outcome are more likely to receive more extensive treatment (Miettinen, 1983). Only 68 patients (12%) received an injection and 58 (10%) were referred to a physiotherapist, which is a low proportion compared to that of an earlier study in The Netherlands (Van der Windt et al., 1996). The Dutch practice guidelines on shoulder complaints, which recommend a wait-and-see policy during the first 2–4 weeks, may have led to a change in practice over the past 5 years. As most patients received wait-and-see policy or medication, we had a relatively homogeneous group regarding treatment at baseline. Adding treatment variables to our models, indeed, did not improve their predictive value, nor strongly influence the association of other predictors with outcome (data not shown).

4.3. Model fit and discrimination

The calibration plots (Fig. 1) show that some predicted probability deciles were slightly too high and some slightly too low. But, in general, both models are rather well calibrated. The AUCs of the models (0.74 for 6 weeks and 0.67 for 6 months) implied satisfactory discrimination between patients with persistent shoulder symptoms and patients without persistent symptoms.

4.4. Analysis

To facilitate comparison between the univariable and multivariable regression analysis, we presented uni- and multivariable ORs in Tables 1 and 3. In case of high event rates (30–50% risk of persistent symptoms), ORs are an overestimation in comparison to the underlying relative risks (RR) and should not be interpreted as such. In our study, we provide, using the prediction rule, the patient and the GP with absolute risks instead of relative risks or odds ratios, because these are easier to understand.

Table 3
Multivariable model with predictors of persistent shoulder symptoms (yes/no) at 6 weeks and 6 months after stepwise backward selection

Predictor	Scale	OR	95% CI
<i>6 weeks (n = 486)</i>			
Duration of complaints			
0–6 weeks ^a			
7–12 weeks	(Yes/no)	1.9	1.1–3.3
>3 months	(Yes/no)	2.6	1.5–4.4
Gradual onset	(Yes/no)	1.8	1.1–2.9
Concomitant psychological complaints	(Yes/no)	2.3	0.9–6.4
Repetitive movements	(Yes/no)	2.0	1.2–3.1
Shoulder pain	(0–10)	1.1	1.0–1.2
Neck pain score at physical examination	(0–18)	1.1	1.0–2.7
<i>6 months (n = 538)</i>			
Duration of complaints			
0–6 weeks ^a			
7–12 weeks	(Yes/no)	1.4	0.9–2.3
>3 months	(Yes/no)	1.9	1.2–3.0
Gradual onset	(Yes/no)	1.4	1.0–1.8
Concomitant low back pain	(Yes/no)	1.6	1.1–2.5
Shoulder pain	(0–10)	1.1	1.0–1.2
Shoulder pain score at physical examination	(0–18)	1.0	1.0–1.1

^a Reference category.

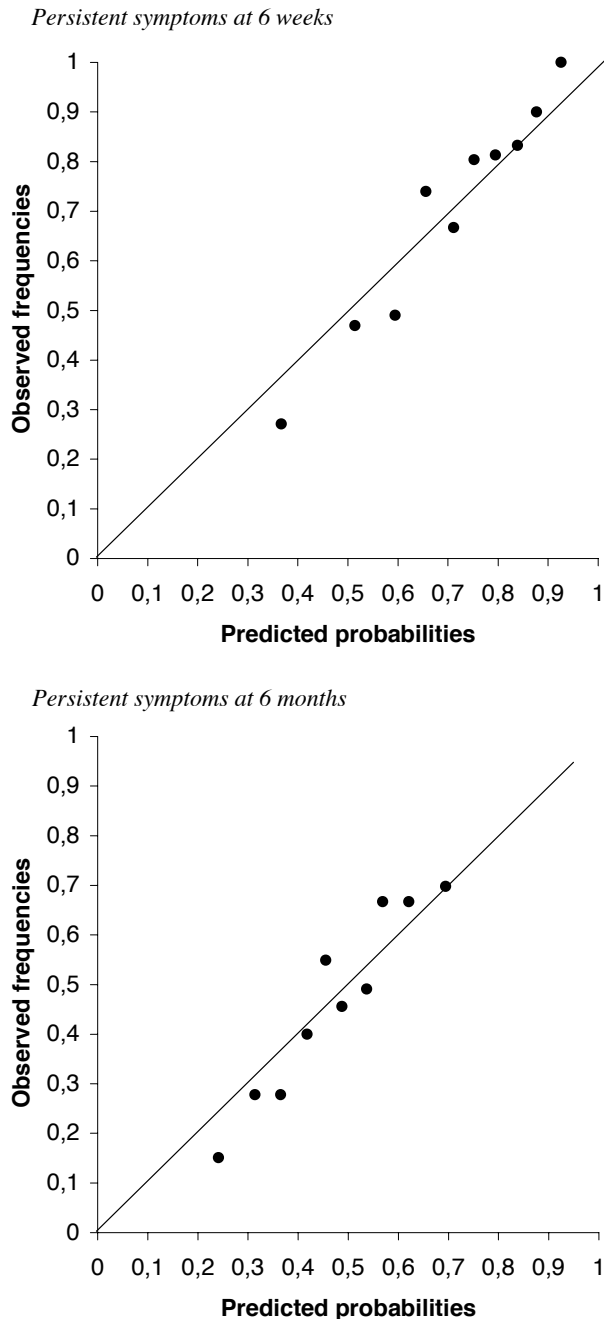


Fig. 1. Calibration plots showing the observed frequencies versus the predicted probabilities for persistent symptoms at 6 weeks and 6 months.

4.5. Internal and external validity of collected data

The response to the questionnaires was high (between 83% and 92%) in this large cohort study. Given the low drop-out rate and only slight differences at baseline between drop-outs and responders, we assume that the results can be generalised to all shoulder patients in our study. The GPs were instructed to recruit consecutive patients. We do not have reliable information to gain insight into the percentage of patients who were

eligible at first consultation of their GP and actually participated in the study. In the 10-day period between first consultation and baseline assessment, most patients only received advice or medication. Nevertheless, in this brief time period some recovery may have occurred. One may argue that a better prediction of non-recovery can be made after a short ‘wait-and-see’ period than at the time of consultation for shoulder pain. This may possibly have led to some over-estimation of the predictive validity of our prediction rules. This is one of the reasons why we want to stress the importance of validating the prediction rules in a daily practice situation, for which they have been developed.

Laupacis et al. (1997) have stressed the importance of inter-observer reproducibility of the variables in a prediction rule. In our study, this may be particularly relevant for the elements of physical examination that were included in our prediction rules (shoulder pain and neck pain on examination). Previous studies have reported conflicting results on the inter-observer variability of examination of the shoulder joint (Bamji et al., 1996; De Winter et al., 1999; Liesdek et al., 1997; Pellecchia et al., 1996), but most studies seem to indicate that classification of shoulder pain into medical diagnostic categories can only be achieved with moderate agreement. In our study, we did not attempt to classify shoulder pain into diagnostic categories. The results of physical examination were transformed to four factors (shoulder mobility, shoulder pain, neck mobility and neck pain) which explained 50.4% of the variation (Bergman et al., 2004a). Pain on examination as reported by the patient was scored on 4-point scales, and subsequently transformed to dichotomous scores to reduce inter-observer variation. Further testing of the prediction rules in clinical practice should demonstrate to what extent the predictive validity is affected by inter-observer variability.

The recovery rates of 30% after 6 weeks and 54% after 6 months are similar to those found in other studies carried out in primary care populations (Croft et al., 1996; Van der Windt et al., 1996; Winters et al., 1999a,b), which may strengthen generalisability of our findings to other primary care patients with shoulder disorders. However, before considering implementation of our score charts in clinical practice the generalisability (‘external validity’) of the models needs to be tested in other populations of patients with shoulder disorders (Justice et al., 1999). First, the generalisability to another primary care population can be tested. If satisfactory, the generalisability to a community sample, occupational setting, or secondary care population may be tested.

4.6. Clinical usefulness

Perhaps most importantly, the clinical usefulness of the developed prediction rules should be established: can the prediction rules be helpful to the clinician when

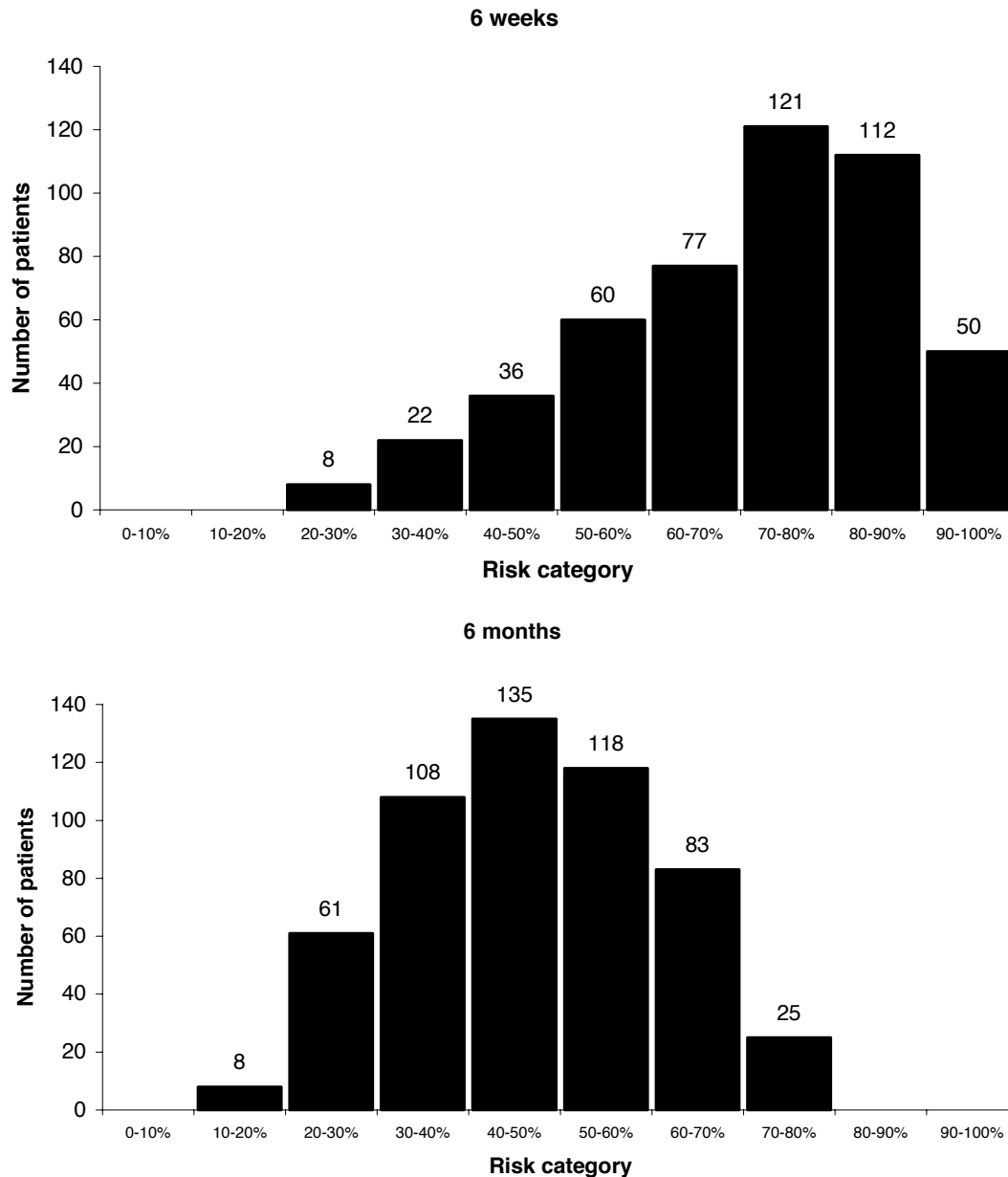


Fig. 2. Number of patients in risk categories for persistent symptoms of the score charts for 6 weeks ($n = 486$) and 6 months ($n = 538$).

making decisions in the management of patients with shoulder pain, for example, whether or not to consider additional diagnostic testing, start a certain treatment or refer the patient to secondary care (Vergouwe et al., 2002). Fig. 2 shows that a relatively small proportion of patients is shifted into the lower risk categories at 6 weeks and a somewhat higher proportion at 6 months. So, a small number of patients can be reassured by their GP. Patients in the high risk categories possibly benefit from earlier and more extensive treatments. An important objective for future research is to study from which interventions patients in the high risk categories benefit most.

5. Conclusion

In conclusion, longer duration of symptoms, a gradual onset of symptoms and high pain intensity at baseline were consistently associated with a poor outcome. The prediction rule and score chart may be used by GPs to calculate the absolute risk of persistent symptoms in individual patients with shoulder pain. The performance of our models still needs to be tested in other populations of patients with shoulder pain to enable valid and reliable use of the score charts in clinical practice.

Score chart for prediction of persistent shoulder symptoms at 6 weeks

Duration of complaints			Total score	Risk
<6 weeks	0	...	≤2	20% - 30%
6-12 weeks	7	...	3 - 7	30% - 40%
>3 months	11	...	8 - 11	40% - 50%
Gradual onset	7	...	12 - 16	50% - 60%
Psychological complaints	10	...	17 - 21	60% - 70%
Repetitive movements	8	...	22 - 27	70% - 80%
Shoulder pain (0-10)	score	...	28 - 36	80% - 90%
Neck pain score at physical examination (0-18)	score	...	≥37	90% - 100%
Total score		... +		

The predicted probability of persistent symptoms at 6 weeks was determined by $P=1/[1+ \exp - (-1.19 + 0.64 \times \text{duration of complaints } 6-12 \text{ weeks} + 0.95 \times \text{duration of complaints } >3 \text{ months} + 0.59 \times \text{gradual onset} + 0.85 \times \text{concomitant psychological complaints} + 0.68 \times \text{repetitive movements} + 0.13 \times \text{shoulder pain} + 0.09 \times \text{neck pain score at physical examination})]$.

Score chart for prediction of persistent shoulder symptoms at 6 months

Duration of complaints			Total score	Risk
<6 weeks	0	...	≤1	10% - 20%
6-12 weeks	9	...	2 - 16	20% - 30%
>3 months	17	...	17 - 28	30% - 40%
Gradual onset	10	...	29 - 39	40% - 50%
Concomitant low back pain	13	...	40 - 49	50% - 60%
Shoulder pain (0-10)	score x2	...	50 - 61	60% - 70%
Shoulder pain score at physical examination (0-18)	score	...	≥62	70% - 100%
Total score		... +		

The predicted probability of persistent symptoms at 6 months was determined by $P=1/[1+ \exp - (-1.48 + 0.34 \times \text{duration of complaints } 6-12 \text{ weeks} + 0.64 \times \text{duration of complaints } >3 \text{ months} + 0.37 \times \text{gradual onset} + 0.50 \times \text{concomitant low back pain} + 0.08 \times \text{shoulder pain} + 0.04 \times \text{shoulder pain score at physical examination})]$.

Instruction

If a predictor is scored positively, the given weight needs to be filled in. Subsequently the scores are added to calculate the 'Total score'. From the table next to the score chart the risk (%) of persistent symptoms for an individual patient can be determined.

Fig. 3. Prognostic score charts for prediction of persistent symptoms at 6 weeks and 6 months.

Acknowledgment

This study was supported by a grant (No. 904-65-09) from the Netherlands Organisation for Scientific Research (NWO), The Hague, The Netherlands.

References

Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.
 Bamji AN, Erhardt CC, Price TR, Williams PL. The painful shoulder: can consultants agree? Br J Rheumatol 1996;35:1172-4.
 Berg SGM, Vlaeyen JWS, Ter Kuil MM, Spinhoven P, Van Breukelen G, Kole-Snijders AMJ. Instruments for measuring chronic pain, part 2. Pain coping and cognition list (Meetinstrumenten chronische pijn, deel 2. Pijn Coping en Cognitie Lijst). Maastricht: Pijn Kennis Centrum; 2001 [in Dutch].

Bergman GJD, Van der Heijden GJMG, Winters JC, Groenier KH, Meyboom-deJong B, Postema K, et al. Manipulative therapy for patients with shoulder complaints: physical examination results in a randomised controlled trial. Manipulative therapy for shoulder complaints in general practice. Groningen: Drukkerij De Regenboog; 2004a.
 Bergman GJD, Winters JC, Groenier KH, Pool JJ, Meyboom-de Jong B, Postema K, et al. Manipulative therapy in addition to usual medical care for patients with shoulder dysfunction and pain: a randomized, controlled trial. Ann Intern Med 2004b;141:432-9.
 Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, et al. A patient-specific approach for measuring functional status in low back pain. J Manipulative Physiol Ther 1999;22:144-8.
 Bongers PM. The cost of shoulder pain at work. BMJ 2001;322:64-5.
 Bot SD, Van der Waal JM, Terwee CB, Van der Windt DAWM, Schellevis FG, Bouter LM, et al. Incidence and prevalence of complaints of the neck and upper extremity in general practice. Ann Rheum Dis 2005;64:118-23.

- Croft P, Pope D, Silman A. The clinical course of shoulder pain: prospective cohort study in primary care. *BMJ* 1996;313:601–2.
- De Winter AF, Jans MP, Scholten RJPM, Devillé W, Van Schaardenburg D, Bouter LM. Diagnostic classification of shoulder disorders: inter-observer agreement and determinants of disagreement. *Ann Rheum Dis* 1999;58:272–7.
- Efron B, Tibshirani R. An introduction to the bootstrap. Monographs on statistics and applied probability. New York (NY): Chapman and Hall; 1993.
- Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
- Hosmer DW, Lemeshow S. Applied logistic regression. New York: - John Wiley & Sons; 1989. p. 135–75.
- Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–24.
- Kori SH, Miller RP, Todd DD. Kinesiophobia: a new view of chronic pain behavior. *Pain Manag* 1990;35–43.
- Kuijpers T, Van der Windt DAWM, Van der Heijden GJGM, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. *Pain* 2004;109:420–31.
- Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488–94.
- Liesdek C, Van der Windt DAWM, Koes BW, Bouter LM. Soft-tissue disorders of the shoulder: a study of inter-observer agreement between general practitioners and physiotherapists and an overview of physiotherapeutic treatment. *Physiotherapy* 1997;83:12–7.
- Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33:73–81.
- Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983;2:267–71.
- Neer CS. Impingement lesions. *Clin Orthop* 1983;70–7.
- The Dutch College of General Practitioners (NHG) Practice Guideline. Accessed at <<http://nhg.artssennet.nl/upload/104/guidelines2/E08.htm/>> on 1 September 2003.
- Okkes IM, Oskam SK, Lamberts H. Van klacht naar diagnose: Episodegegevens uit de huisartspraktijk. Bussum: Coutinho; 1998.
- Pellecchia GL, Paolino J, Connell J. Inter-tester reliability of the Cyriax evaluation in assessing patients with shoulder pain. *J Orthop Sports Phys Ther* 1996;23:34–8.
- Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. *Pain* 2003;102:167–78.
- Solem-Bertoft E, Lundh I, Westerberg CE. Pain is a major determinant of impaired performance in standardized active motor tests. A study in patients with fracture of the proximal humerus. *Scand J Rehabil Med* 1996;28:71–8.
- Terluin B. Four-Dimensional Symptom Questionnaire (4DSQ) De Vierdimensionele Klachtenlijst (4DKL). Een vragenlijst voor het meten van distress, depressie, angst en somatisatie [in Dutch]. *Huisarts wet* 1996;39(12):538–47.
- Terluin B, Van Rhenen W, Schaafeli W, De Haan M. The Four-Dimensional Symptom Questionnaire (4DSQ): measuring distress and other mental health problems in a working population. *Work Stress* 2004;18:187–207.
- Van der Heijden GJMG. Shoulder disorders: a state-of-the-art review. *Baillieres Best Pract Res Clin Rheumatol* 1999;13:287–309.
- Van der Heijden GJMG, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. *J Clin Epidemiol* 2000;53:29–38.
- Van der Linden MW, Westert GP, De Bakker DH, Schellevis, FG. Complaints and disorders in the population and general practice (Klachten en aandoeningen in de bevolking en in de huisartsenpraktijk). Utrecht/Bilthoven: NIVEL/RIVM; 2005 [in Dutch].
- Van der Windt DAWM, Koes BWK, Jong BAD, Bouter LM. Shoulder disorders in General Practice: incidence, patients characteristics, and management. *Ann Rheum Dis* 1995;54:959–64.
- Van der Windt DAWM, Koes BWK, Boeke AJ, Deville W, De Jong BA, Bouter LM. Shoulder disorders in general practice: prognostic indicators of outcome. *Br J Gen Pract* 1996;46:519–23.
- Van der Windt DAWM, Koes BWK, Deville W, Boeke AJ, De Jong BA, Bouter LM. Effectiveness of corticosteroid injections versus physiotherapy for treatment of painful stiff shoulder in primary care: randomised trial. *BMJ* 1998;317:1292–6.
- Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Validity of prognostic models: when is a model clinically useful? *Semin Urol Oncol* 2002;20:96–107.
- Vendrig A, Deutz P, Vink I. Dutch version of the Fear-Avoidance Beliefs Questionnaire (Nederlandse vertaling en bewerking van de Fear-Avoidance Beliefs Questionnaire), vol. 18. *Nederlands: Tijdschrift voor Pijn en Pijnbestrijding*; 1998. p. 11–4 [in Dutch].
- Vlaeyen JWS, Seelen HAM, Peters M, Jong Pd, Aretz E, Beisiegel E, et al. Fear of movement/(re)injury and muscular reactivity in chronic low back pain patients: An experimental investigation. *Pain* 1999;82:297–304.
- Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157–68.
- Wasson JH, Sox HC. Clinical prediction rules. Have they come of age? *JAMA* 1996;275:641–2.
- Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313:793–9.
- Westerberg CE, Solem-Bertoft E, Lundh I. The reliability of three active motor tests used in painful shoulder disorders. Presentation of a method of general applicability for the analysis of reliability in the presence of pain. *Scand J Rehabil Med* 1996;28:63–70.
- Winters JC, Jongh AC, van der Windt DA, Jonquiere M, Winter AF, van der Heijden GJ, et al. Practice guideline “Shoulder complaints” (NHG-Standaard Schouderklachten). *Huisarts en wetenschap*. 1999a;42:222–31 [in Dutch].
- Winters JC, Sobel JS, Groenier KH, Arendzen JH, Meyboom-de Jong B. The long-term course of shoulder complaints: a prospective study in general practice. *Rheumatology* 1999b;38:160–3.