Development of a Prognostic Model for Patients With Shoulder Complaints in Physiotherapy

Yasmaine Karel
Arianne Verhagen
Marloes Thoomes-de Graaf
Edwin Duijn
Maaike van den Borne
Annechien Beumer
Ramon Ottenheijm
Geert-Jan Dinant
Bart Koes
Wendy Scholten-Peeters

Published in Physical Therapy
2016, 97(1)

ABSTRACT

Background

Health care providers need prognostic factors to distinguish between patients who are likely to recover compared to the ones that do not.

Objective

To describe the clinical course and identify prognostic factors of recovery, in patients with shoulder pain at 26 weeks follow-up.

Design

A prospective cohort study was carried out in the Netherlands including 389 patients consulting a physiotherapist with a new episode of shoulder pain.

Method

Patients were followed for 26 weeks. Potential predictors were selected from the literature, together with the use of diagnostic ultrasound and working alliance and evaluated in multivariable regression analysis. Multiple imputation was used to handle missing data and bootstrap methods for internal validation.

Results

Recovery rate was 60% for the total population and 65% for the working population after 26 weeks. Short duration of complaints, lower disability scores, having a paid job, better working alliance and no feelings of depression/anxiety were associated with recovery. In the working population only duration of complaints and disability remained in the final model. The area under the receiver operator curve (AUC) was 0.67 for the final model of the total population and 0.63 for the working population. After internal validation the AUC was corrected to 0.66 and 0.63.

Limitations

External validation should be done prior to the use in clinical practice.

Conclusion

Results from this study indicate that several factors can predict recovery.

INTRODUCTION

Shoulder complaints are common in western societies and belong to the top 3 of most occurring musculoskeletal complaints. [1] Prevalence rates in the Netherlands range from 6.9 to 48% in primary care. [2-4] About 13% of the patients with shoulder pain who visit the general practitioner are referred to physiotherapy. [4] In the Netherlands patients can visit the physiotherapist without a referral since 2006 and 41% of patients in physiotherapy care used direct access in 2013. [5]

Examining patients with shoulder pain is complex because history taking and physical examination have limited validity for diagnosing the patho-anatomical origin of symptoms. Knowledge about prognostic factors can help the physiotherapist by informing the patient about the expected prognosis and, when indicated, in treatment decisions or referral to other health care professionals. [6, 7] Duration of symptoms, high levels of pain and the presence of co-morbidities have been identified as predictors of poor recovery by patients consulting a General Practitioner (GP). [7-11] Because of the difficulty in diagnosing patients with shoulder pain, physiotherapists are increasing the use of diagnostic ultrasound to assist their clinical decision-making. Nevertheless, the diagnostic and prognostic consequences of using diagnostic ultrasound remains unknown. [12, 13] Furthermore, recent literature suggest patient's prognosis to be influenced by the therapeutic relationship, frequently referred to as "working alliance". [14]

Health care providers need prognostic factors to distinguish between patients who are likely to recover compared to the ones that do not, i.e. the patients which have a high risk of developing chronic shoulder pain. Prognostic factors for shoulder pain have been identified in general practice and only duration of complaints, disability score and age have been identified in a physiotherapy setting. [7, 15] Although patients visiting general practice might be similar in type and severity of complaints compared to the patients in physiotherapy practice, the moment of seeking health care and the treatment provided in both settings is different for most patients. In this study we aim to identify prognostic factors of recovery, including the use of diagnostic ultrasound and working alliance, for patients with shoulder pain in physiotherapy practice.

METHODS

Study Design

This study was a prospective cohort study with a follow-up of 26 weeks in physiotherapy practice of patients with non-specific shoulder complaints. Details of the study design were published in 2013. [16] The Medical Ethics Committee of the Erasmus Medical Center approved the study protocol (MEC-2011-414).

Study Population

From November 2011 to November 2012 physiotherapists recruited consecutive patients. Patients that consulted the physiotherapist were eligible for the study when they suffered from shoulder pain, were aged ≥ 18 years and had adequate understanding of the Dutch language. Patients were excluded if they had serious pathologies (infection, cancer or fracture), previous surgery of the shoulder in the last 12 months, or received diagnostic imaging techniques such as musculoskeletal ultrasound, magnetic resonance imaging or X-ray of the shoulder in the 3 months prior to start of the study. All patients provided written informed consent.

Procedures

During first consultation patients received study information and signed the consent form. This was sent to the researchers together with patients' name and e-mail address. Next, baseline questionnaires were sent to the e-mail address or post address when patients did not have e-mail. Follow-up questionnaires were sent 6, 12 and 26 weeks after the start of the treatment. A maximum of 2 reminders were sent when no response was received after 3 and 5 days.

Candidate predictors

Prognostic factors for recovery for patients with shoulder pain were extracted from the literature and consisted of sociodemographic variables and clinical characteristics. [7, 10, 17-19] Sociodemographic variables were age (continuous), gender, level of education (low = no education, primary school or lower vocational school, medium = lower general secondary school or middle vocational school, high = higher general secondary school, higher vocational school or university), employment status (paid job yes/no) and job description (physically heavy work, static repetitive work or work with awkward postures; yes/no).

Clinical characteristics were duration of complaints (months), previous episode of shoulder pain (yes/no), pain intensity at baseline (11-point numeric rating scale, NRS-11), and co-morbidity of arm (elbow/wrist/hand), back or neck (yes/no), sick leave due to shoulder complaint (yes/no), and increase of complaints during work (yes/no).

The shoulder complaint was considered work related when patients with a paid job answered "yes" to one of the following three questions: (1) Do the complaints worsen or return during activities at work? (2) Have you adapted or reduced your activities at work because of your complaints? (3) Do the complaints diminish after several days off work? [20]

The Dutch Shoulder Pain and Disability Index (SPADI) consist of five items assessing pain and eight items assessing disability. The score ranges from 0 to 100% with a high score indicating more functional disability. The questionnaire has good validity and reliability. [21]

Additionally, we assessed working alliance, the use of diagnostic ultrasound (yes/no) and the anxiety/depression dimension of the EuroQOL five dimensions as possible prognostic factors. Working alliance was measured with the Flemish (Dutch) version of the Working Alliance Inventory (WAV-12) and was assessed after 6 weeks. This questionnaire has three subscales designed to assess three primary components of the working alliance: 1) how closely client and therapist agree on and are mutually engaged in the goals of treatment, 2) how closely client and therapist agree on how to reach the treatment goals and 3) the degree of mutual trust, acceptance, and confidence between client and therapist. Patients score on a 5-point scale ranging from rarely to always. This scale is validated in patients receiving psychotherapy in Belgium. [22, 23]

The EuroQOL 5 dimensions-3L (EQ-5D) was used to measure health related quality of life. Little is known about the prognostic value of psychosocial factors. Therefore we used one dimension focusing on the emotional and social functioning, questioning the patient whether he or she was anxious or depressed (not, moderate or extremely). The EQ-5D is a valid and reliable generic instrument for measuring health related quality of life. [24, 25]

Outcome measures

The primary outcome measure was the Global Perceived Effect (GPE) scale and measures whether the patient rates it's condition as improved or deteriorated since the start of the physiotherapy treatment. It uses a 7-point Likert scale scoring and ranges from 'worse than ever' to 'fully recovered'. Patients were to be considered recovered when they scored 'strongly improved' or 'completely recovered'. [24, 26]

The secondary outcome measure were: 1) pain severity and was measured with the 11 point Numeric Rating Scale (NRS) ranging from no pain (0) to intolerable pain (10) and 2) disability measured with the Shoulder Pain And Disability Index (SPADI) ranging from no disability (0) to complete disability (100).

Sample size

Based on the literature about 40% of the patients with shoulder pain will recover within 6 months. [9, 27, 38] We aimed to include 12 prognostic variables in our prognostic model. Based on the 1 in 10 rule of 10 events per variable, a total of 120 events are needed in the smallest outcome (recovered or not). [28] Adjusting for about 20% missing values, the total population should comprise a minimum of 360 subjects.

Statistical Analysis

First we performed a descriptive analysis by calculating frequencies for categorical variables and means with standard deviations (SD) for continuous variables at 6, 12 and 26 weeks. In case the data was not normally distributed median scores and the interquartile range were reported. Multiple imputation was used in case of missing data. Predictor

variables and the outcome were included in the multiple imputation and was done separately for primary and secondary outcome measures. [29-31] A total of 20 datasets were created and regressions analysis was done in all datasets. Pooled estimates were calculated according to Ruben's rule. [32] All assumptions (linearity between independent variables and log odds and multicollinearity (>0.80) for continuous variables) were checked before model building. Univariable and multivariable regression were reported for the total population and working population separately, because several work related variables (job demands and psychosocial factors at work like low decision authority and low control) are found to be related to recovery in the working population specifically. [20, 33] Unadjusted associations were checked between each candidate predictor and the outcome for significant contribution to the outcome (P>0.2). All candidate predictors derived from the literature were included in the multivariate regression analysis (full model). Multiple logistic regression analysis was used to determine which baseline variables were predictors of recovery at 26 weeks (using the GPE). Next, a backward selection procedure was used to determine which variables were kept in the model (final model). A variable was selected when the variable appeared statistically significant in 12 out of 20 imputed models. [34] A p-value of <0.05 was considered statistically significant. The reliability of the multivariable model was determined with the Hosmer-Lemeshow goodness-of-fit statistic. [35] Discriminative ability of the models was assessed using the area under the receiver-operating characteristic curve (AUC-ROC). An area under the curve (AUC), of 0.5 indicates poor discrimination above chance, 0.7 indicates fair discrimination, 0.8 indicates acceptable discrimination, whereas an AUC of 1.0 indicates perfect discrimination. [35] Optimal models were classified as those that yielded the highest AUC. Calibration of the model predictions was assessed by the amount of overlap between the predicted individual probabilities against the observed recovery. The same 12 predictors used for logistic regression modeling were used for linear regression modeling with pain as outcome to evaluate if the model would be similar for a secondary outcome measure. Only one secondary outcome (pain) was used as a secondary outcome measure in the regression model because the SPADI and NRS scores were highly correlated (a=0.87).

We performed internal validation for the primary outcome measure by bootstrapping in order to correct for overfitting. A total of 1000 new datasets were created by random drawing samples from the dataset and we assessed the AUC. [36] The performance in the bootstrap sample represents estimation of the apparent performance, and the performance in the original sample represents test performance. The difference between these is an estimate of the optimism in the apparent performance. The optimism is subtracted from the apparent performance to estimate the internally validated performance. [37] All imputed datasets were bootstrapped and the AUCs were averaged to get the apparent performance. Statistical analyses were performed by using SPSS 22.0 software. Bootstrap analyses were done with R software. [38]

RESULTS

Study population

In total 412 patients fulfilled the eligibility criteria of which 389 gave informed consent and thus entered the cohort. From the 389 patients 366 (94%) returned the baseline questionnaire. After 26 weeks 272 (70%) returned the questionnaire (figure 1). There were 11% missing values. There were no statistically significant differences in baseline characteristics in patients with or without missing data.

Baseline characteristics of the study population were described in table 1 together with missing data. The population consisted of 170 men (45%), the mean age was 49.9 (SD=13.2), 261 (71%) had a paid job and the median duration of their complaints was 12 weeks (IQR=6-26). The working population did not significantly differ from the total population except concerning disability (SPADI). All patients received physiotherapy treatment.

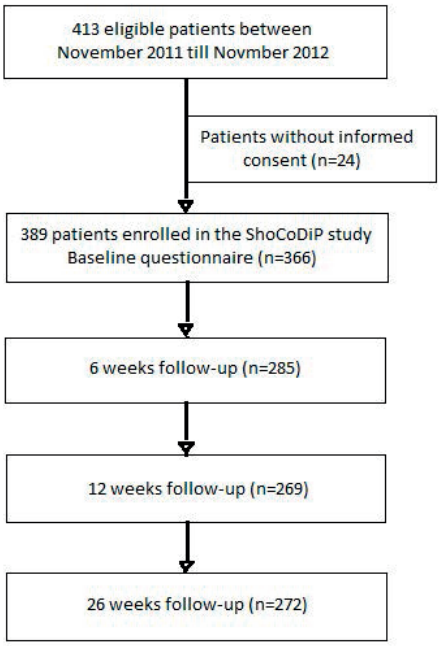

Figure 1. Flow Diagram

Clinical course

After 6 weeks follow-up 118 (41%) patients were recovered; 152 (57%) after 12 weeks and 164 (60%) after 26 weeks. Recovery rates in the working population were slightly higher; 91 patients recovered after 6 weeks (46%), 110 (60%) after 12 weeks and 119 (65%) after 26 weeks.

Median (IQR) SPADI score decreased from 49.5 (29-65) at baseline to 16.9 (3.9-43.0) at 26 weeks (Figure 2) and the NRS median score (IQR) decreased (Figure 3) from 6 (4-7) to 2 (1-5). For the working population, the disability score decreased from 44.9 (27-61) at baseline to 12.7 (3-35) at 26 weeks and pain score decreased from 6 (4-7) to 2 (0-5)

Table 1 Baseline characteristics

Baseline characteristics	Total population (n=389)	Working population (n=261)	Available data (%)
	Sociodemographic		
Age (years) mean (SD)	49.9 (13.2)	45 (10.7)	374 (96)
Male, n (%)	170 (45)	121 (46)	376 (97)
Educational level, n (%)			
Low	40 (11)	16 (6)	366 (94)
Medium	199 (54)	142 (56)	
High	127 (35)	98 (38)	
Paid work, n (%)	261 (71)	-	368 (95)
Full time, n (%)	-	136 (53)	257 (98)
Job description, n (%)			
Physically heavy work	-	64 (25)	258 (99)
Static repetitive work	-	88 (34)	
Work in awkward postures	-	11 (37)	
Work related complaints, n (%)	-	167 (69)	238 (91)
Sick leave, n (%)	-	40 (16)	257 (98)
	Clinical characteristics		
Duration in weeks, med (IQR)	12 (6-26)	12 (5-26)	371 (95)
Recurrent episode, n (%)	158 (43)	111 (44)	364 (94)
Dominant side affected, n (%)	224 (61)	159 (62)	369 (95)
Comorbidity, n (%)	236 (65)	156 (60)	364 (94)
Pain score NRS, med (IQR)	6.0 (4-7)	6.0 (4-7)	373 (96)
SPADI, med (IQR)	49.5 (29-65)	44.9 (27-61)	367 (94)
	Psycho-social characteristic		
Fear/depression EQ5D, n (%)			
not anxious/depressed	300 (83)	209 (83)	360 (93)
moderately	59 (16)	42 (16)	
anxious/depressed			
extremely	1 (0)	0 (0)	
anxious/depressed			
	Other		
Diagnostic US performed, n (%)	122 (31)	67 (26)	389 (100)
Working alliance, mean (SD)	45.3 (9.1)	46.7 (9.6)	87 (22)

N number, SD standard deviation, IQR Interquartile range, med median, NRS Numeric Rating Scale, SPADI Shoulder Pain and Disability Index, EQ-5D EuroQOL 5 Dimensions, US Ultrasound

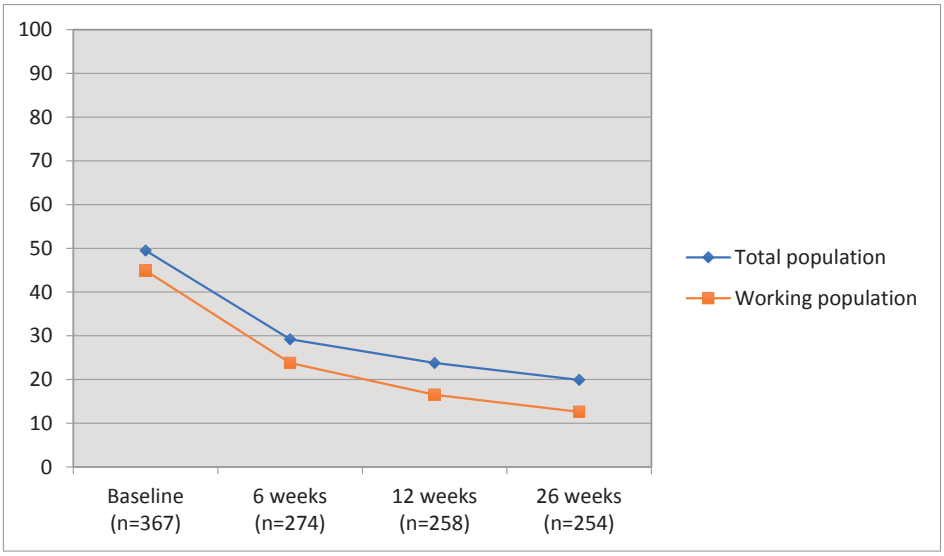

Figure 2. Median scores of disability (SPADI) at baseline, 6, 12 and 26 weeks follow-up.

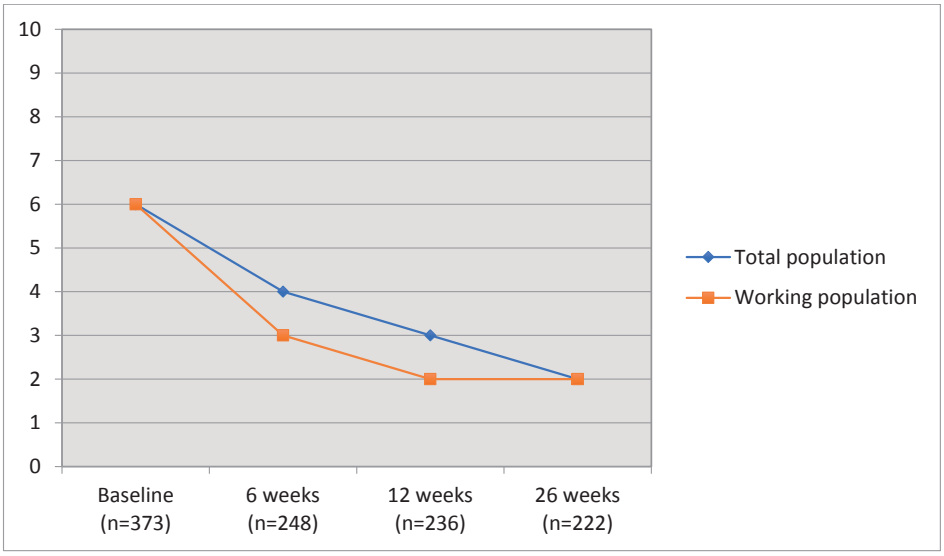

Figure 3. Median scores of pain severity (NRS-11) at baseline, 6, 12 and 26 weeks follow-up.

Table 2. Univariable & multivariable associations with recovery at 26 weeks.

	Total population (n=389) OR [95% CI] Beta		Working population (n=261) OR [95% CI] Beta	
Prognostic factors	Univariable	Multivariable	Univariable	Multivariable
Sociodemographic variables				
Age (years)	0.98[0.96-1.00]*† -0.017	0.99 [0.96-1.02] † -0.008	0.99 [0.97-1.02] † -0.006	1.01 [0.98-1.05] † 0.009
Female	0.9 [0.6-1.6] -0.058	1.1 [0.6-2.0] 0.307	0.9 [0.5-1.7] -0.072	2.0 [0.7-5.3] 0.690
Educational level				
Low	1.0	1.0	1.0	1.0
	0.7 [0.3-1.8]	0.4 [0.2-1.1]	0.6 [0.1-2.6]	0.5 [0.1-2.2]
Medium	-0.348	0.486	-0.451	-0.696
	0.9 [0.4-2.2]	0.5 [0.2-1.2]	0.8 [0.2-3.5]	0.7 [0.1-3.1]
High	-0.078	0.499	-0.101	-0.391
Clinical characteristics				
Duration in weeks	0.99 [0.99-1.00]** † -0.006	0.99 [0.99-0.99]** † -0.006	0.99 [0.99-1.00]** † -0.005	0.99 [0.99-1.00]** † -0.007
Recurrent episode (no)	1.7 [1.0-2.7]** 0.506	1.4 [0.8-2.5] 0.329	1.8 [0.9-3.4]** 0.562	1.5 [0.8-3.1] 0.435
Comorbidity (no)	1.3 [0.7-2.4] 0.270	1.0 [0.5-2.1] 0.012	1.1 [0.6-2.1] 0.111	0.9 [0.4-2.0] -0.084
Pain score NRS	0.9 [0.8-1.0]** -0.133	1.0 [0.8-1.2] 0.010	0.9 [0.8-1.0]* -0.120	1.0 [0.8-1.3] -0.004
Disability score, SPADI	0.98 [0.97-1.00]** † -0.017	0.99 [0.97-1.00] † -0.014	0.98 [0.97-1.00]** † -0.018	0.98 [0.96-1.01] † -0.017
Work related characteristics				
Paid work (no)	0.5 [0.3-0.9]** -0.667	0.6 [0.3-1.2] -0.583		
Full time (no)			0.6 [0.3-1.2]* -0.472	0.5 [0.2-1.2] -0.799
Job description				
Physically heavy work			0.8 [0.3-1.7]	0.9 [0.4-2.3]
			-0.276	-0.091
Static repetitive work			1.1 [0.5-2.4]	1.4 [0.6-3.4]
			0.142	0.352
Work in awkward postures			1.0 [0.2-4.4]	2.0 [0.3-12.1]
			0.094	0.710
Other			1.0	1.0

Table 2. Univariable & multivariable associations with recovery at 26 weeks. (continued)

	Total population (n=389) OR [95% CI] Beta		Working population (n=261) OR [95% CI] Beta	
Prognostic factors	Univariable	Multivariable	Univariable	Multivariable
Work related complaints (no)			0.5 [0.2-1.8] -0.538	0.4 [0.1-1.6] -0.834
Sick leave (no)			0.9 [0.3-2.4] 0.225	1.3 [0.5-3.9] 0.295
Psycho-social characteristics				
Fear/depression, EQ5D,				
No feelings of	1.9 [1.0-3.3]**	2.0 [0.9-4.0]	1.9 [0.9-4.0]*	1.8 [0.7-4.3]
anxiety/depression	0.518	0.655	0.532	0.566
Other				
Diagnostic US performed (no)	1.5 [0.9-2.4]* 0.394	1.2 [0.7-2.2] 0.174	1.4 [0.8-2.7] 0.340	1.3 [0.6-2.8] 0.264
Working alliance	1.0 [1.0-1.1] 0.010	1.0 [0.9-1.1] 0.010	1.0 [1.0-1.1] 0.010	1.0 [0.9-1.1] 0.009

OR: Odds Ratio, CI: Confidence Interval, SPADI: Shoulder Pain and Disability Index, NRS: Numeric Rating Scale, EQ-5D: EuroQOL 5 Dimensions
** P <0.10
* P <0.20
† rounded off with 2 decimals because of small CI

Predictors and model evaluation

All predictors

For all variables included in the model the variance inflation factors were < 1.5 and correlation coefficients <0.8, suggesting that linearity and multicollinearity was not a problem. In the univariable regression analysis, 8 factors were related (P<0.20) with recovery at 26 weeks (Table 2). There was only one patient who scored "very anxious/ depressed" on the depression score of the EQ-5D and therefore this answer option was combined with 'moderately depressed' and the EQ-5D was thus dichotomized in the regression analysis.

First we tested a model that included all prognostic variables (n=12) selected from the literature (Table 2). The R^2 was 0.17 and the ROC curve demonstrated a fair discriminating ability for the regression model with an AUC of 0.70 (95% CI 0.36-1.03) and correctly classified 66% of patients. The model in the working population resulted in similar results (see table 2). The R^2 for the working population was 0.19 and the AUC was 0.72 (95% CI 0.37-1.10) and the model correctly classified 69% of patients.

Backward regression analysis

Results from the backward regression resulted in a model where: a short duration of complaints, lower disability score, having a paid job, no feelings of depression/anxiety and high working alliance were related to recovery (table 3). The R^2 was 0.12 and the AUC was 0.67 (95% CI 0.34-1.0) and the model correctly classified 65% of patients.

In the working population we found identical results (table 3). The final model showed a short duration of complaints and low disability scores were related to recovery. The R^2 was 0.05 and the AUC was 0.63 (95% CI 0.25-1.00) and the model correctly classified 67% of patients.

Table 3 Final model; results from backward logistic regression

Final model after Backward Wald regression for recovery				
	Total population (n=389)		Working population (n=261)	
	OR [95% CI]	Beta	OR [95% CI]	Beta
Duration in weeks	0.99 [0.99-1.00]* †	-0.007*	0.99 [0.99-1.00]* †	-0.006*
Disability score, SPADI	0.99 [0.97-1.00]* †	-0.014*	0.98 [0.97-1.00]* †	-0.017*
Paid work (no)	0.6 [0.3-1.0]*	-0.592*		
Fear/depression, EQ5D,				
No Feelings of anxiety/depression	1.8 [0.9-3.6]	0.588		
Working Alliance	1.0 [0.9-3.6]	0.004		
Performance measures				
R^2	0.12		0.05	
AUC	0.67		0.63	
Bootstrapped AUC	0.66		0.63	
Final model after Backward Wald regression for pain				
Recurrent episode (no)	NA	0.738*	NA	0.779*
Duration in weeks	NA	0.004*	NA	0.005
Disability score, SPADI	NA	0.031*	NA	0.034*
Performance Measures				
R^2	0.13		0.15	

OR odds ratio, CI confidence interval, SPADI Shoulder Pain And Disability Index, EQ5D EuroQol 5 dimensions, AUC Area Under the Curve, R^2 R Squared

* p-value <0.05

† rounded off with 2 decimals because of small CI

Secondary outcome

Using pain as outcome resulted in a model including duration of complaints, recurrent episode and disability score in both the total (R^2=0.13) and working population (R^2=0.15).

Internal validation

Bootstrap method to assess optimism was checked in all prediction models (full and final model after backward elimination) for the primary outcome measure. Discriminative ability decreased in all models after bootstrap. The apparent performance (bootstrap corrected AUC) of the full model in the total population decreased from 0.70 to 0.67. The expected optimism for the AUC of the total population in the full model was 0.024 and 0.0409 in the working population. Optimism of the final model in the total population was 0.008 and 0.002 in the working population (table 3).

DISCUSSION

Our study showed that a short duration of complaints, not having feelings of depression or anxiety, having a paid job, a better working alliance and a low disability score were predictors of recovery after 6 months. Duration of complaints and disability were also predictors of recovery in the working population. In the prediction model for pain a recurrent episode of shoulder pain, short duration of complaints and low disability scores, were the predictors in the final model.

In this prognostic cohort study 60% of patients reported to be recovered after 6 months. This is slightly higher than the 21-51% reported by studies in GP practice. [9, 27, 39]

In line with previous research we found that a shorter duration of symptoms and lower disability scores were significantly associated with recovery. [7, 10, 15, 40-42]

Other prognostic models found the predictors; age, gender, [10] repetitive movement [9] and co-morbidities, [9, 20, 27, 43] which we included as possible predictor but did not remain in the final model. The reason that we did not find co-morbidity to be a predictor might be due to the difference in defining co-morbidity. Like this study, one study formulated co-morbidity as musculoskeletal (yes/no) [20] but others only measured concomitant low back pain [9] or concomitant neck pain [27]. Furthermore, we only asked for the co-morbidities around the shoulder region. Several studies have shown that other co-morbidities (like obesity, headache) also has an impact on an individual's ability to recover. [44-46]

Contrary to our findings, previous studies have not found a significant association of psychosocial factors and shoulder complaints. [7] However, in studies including patients with complaints of the arm, neck and shoulder psychosocial factors appear to have a predictive effect on patient outcome. [20] This effect has not been found in the literature specific for patients with only shoulder pain. We included only one item about depression and anxiety from the EQ-5D. This variable was dichotomized which might contribute to a loss of information. However the variable remained in the final model. One other study found catastrophizing at baseline to be a predictor of function. [44]

Working alliance remained in the final model as well.

It has been suggested that patient reported outcome measures, such as recovery and pain, are sensitive to the effect of interactions between patients and treatment providers. [47] One review has shown that a good working alliance can improve treatment outcomes. [14] Also, good working alliance scores might result in higher levels of adherence. [48] Treatment adherence is important to achieve optimal treatment outcomes and it is widely accepted that a lack of adherence to long-term therapies result in poor treatment outcomes and high costs of health care. The argument is that a good working alliance could help patients adhere to the treatment regime. [48] A good working alliance is partially determined by the communication between the patient and therapist. For that reason effective communication should be an essential skill that therapists need to master in order to improve health care.

Various other studies suggest that working alliance is associated with recovery in physical rehabilitation settings, but more research is needed to determine the strength of the possible relationship between the therapeutic alliance and recovery. [14]

Strength of this study is that we evaluated the prognostic value of two new variables, working alliance and the use of diagnostic ultrasound, upon variables that were described before. Furthermore the number of potential prognostic variables was not large, leading to more valid statistical derivations. [49, 50] There is a possibility that variables not mentioned in the literature were left out of this model but might have been significant predictors in our population.

In the model the use of diagnostic US was added as a dichotomous variable. This is because we assumed that a more specific diagnosis, as found using diagnostic US, leads to a more specific treatment and should lead to better patient outcomes. The low number of patients with an US diagnosis limited our ability to perform any additional analysis.

The percentage of missing values for the outcome was 30% after 6 months follow-up. Missing data was handled adequately with multiple imputations, although the large amount of missing data for working alliance might influence the validity of the data.

The model's performance is likely to be overestimated in the developmental dataset. Therefore we assessed the amount of optimism and corrected by using bootstrapping techniques to internally validate the model. The expected optimism after internal validation was small in all but one model. The optimism in the full model of the working population was substantial, probably due to the relatively small sample size. Similar levels of optimism have been observed earlier in smaller sample sizes. [50, 51] Furthermore the performance of the final model was not very good. Several 95% CI's around the AUC estimates crossed the 0.50 threshold indicating a high likelihood of poor discrimination.

All patients received physiotherapy treatment but it consisted of several treatment modalities resulting in heterogeneity. Besides heterogeneity in treatment, patients with

more severe complaints are more likely to receive more treatment sessions thus possibly influencing recovery status.

Future research.

Based on the relatively low AUC scores the prognostic model could be improved by possibly adding other psychosocial factors besides depression/anxiety and evaluate if the physiotherapy treatment and the number of treatment sessions could cause interaction effects. Hardly any prognostic models are routinely used in clinical practice, probably because most have not been externally validated. [52] It is crucial to quantify the performance of a prognostic model in different populations before applying it in daily practice. Since prognostic models in primary care for patients with shoulder pain seem to have similar performance estimates the next step might be to externally validate a high quality model with appropriate performance/discrimination in a new dataset. [9, 53, 54]

CONCLUSION

We developed and internally validated a model predicting recovery of patients with shoulder complaints in physiotherapy practice. Other variables should be evaluated to improve predictive capacity of the model and next the model should be externally validated before it can be used in clinical practice. In daily practice physiotherapists constantly predict the risk or probability of an individual to recover. Based on the predicted prognosis they inform individual patients about the course of the disease or the choice for further treatment. Knowledge of the predictors described in literature can be informative for the physiotherapist for their prognostic potential. When a model performs well at external validation it will probably be a useful tool, as it may enhance communication. Nevertheless its impact on patient outcomes should be assessed using a clinical trial design.

Acknowledgements

We thank Caspar Looman for statistical support

We thank all the physiotherapists and patients for participating in this study

Funding

Authors declare no conflict of interest.

This study is funded by SIA-RAAK. The ministry of education has made this funding available for the innovation and promotion of research. This study is partly funded by a program grant of the Dutch Arthritis Foundation.

REFERENCES

1. Picavet, H., & Schouten, J. (2003). Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC3-study. *Pain,* 102(1-2), 167-178.
2. Luime, J., Koes, B., Hendriksen, I., Burdorf, A., Verhagen, A., Miedema, H., & Verhaar, J. (2004). Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scandinavian Journal of Rheumatology*, 33, 73-81.
3. Feleus, A., Bierma-Zeinstra, S., Miedema, H., Bernsen, R., Verhaar, J., & Koes, B. (2008). Incidence of non-traumatic complaints of arm, neck and shoulder in general practice. *Manual Therapy*, 13(5), 426-33.
4. Kooijman, M., Swinkels, I., van Dijk, C., de Bakker, D., & Veenhof, C. (2013). Patients with shoulder syndromes in general and physiotherapy practice: an observational study. *BMC Musculoskeletal Disorders*, 14(128).
5. Barten, D.J.A., Verberne, L.D.M., Koppes, L.L.J. Zorg door de fysiotherapeut – aanmelding [Internet]. NIVEL Zorgregistraties eerste lijn; 2015 [cited 2015 august 23]. Available from: www.nivel.nl/node/3097
6. Dinant, G.J., Buntinx, F., Butler, C.C. (2007). The necessary shift from diagnostic to prognostic research. *BMC Family Practice,* 13(8), 53.
7. Kuijpers, T., Van der Windt, D., Van der Heijden, G., & Bouter, L. (2004). Systematic review of prognostic cohort studies in shoulder disorders. *Pain*, 109, 420-431.
8. Keijsers, E., Feleus, A., Miedema, H., Koes, B., & Bierma-Zeinstra, S. (2010). Psychosocial factors predicted nonrecovery in both specific and nonspecific diagnoses at arm, neck, and shoulder. *Journal of Clinical Epidemiology*, 63, 1370-79.
9. Kuijpers, T., van der Windt, D., Boeke, A., Twisk, J., Vergouwe, Y., Bouter, L., & van der Heijden, G. (2006). Clinical prediction rules for the prognosis of shoulder pain in general practice. *Pain*, 120(3), 276-285.
10. Kennedy, C., Manno, M., Hogg-Johnson, S., Haines, T., Hurley, L., McKenzie, D., & Beaton, D. (2006). Prognosis in soft tissue disorders of the shoulder: predicting both change in disability and level of disability after treatment. *J Am Phys Ther Assoc.*, 86(7), 1018-1032.
11. Bruls, V.E., Bastiaenen, C.H., de Bie, R.A. (2015). Prognostic factors of complaints of arm, neck, and/or shoulder: a systematic review of prospective cohort studies. *Pain*, 156(5), 765-99.
12. Hanchard, N., Lenza, M., Handoll, H., & Takwoingi, Y. (2013). Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. *The Cochrane Database of systematic reviews*, 30(4).
13. Wright, A., Wassinger, C., Frank, M., Michener, L., & Hegedus, E. (2013). Diagnostic accuracy of scapular physical examination tests for shoulder disorders: a systematic review. *British Journal of Sports Medicine*, 47(13), 886-92.
14. Hall, A., Ferreira, P., Maher, C., Latimer, J., & Ferreira, M. (2010). The influence of the therapist-patient relationship on treatment outcome in physical rehabilitation: a systematic review. *Physical Therapy*, 90(8), 1099-110.
15. Chester, R., Shepstone, L., Daniell, H., Sweeting, D., Lewis, J., Jerosch-Herold, C. (2013). Predicting response to physiotherapy treatment for musculoskeletal shoulder pain: a systematic review. *BMC Musculoskeletal Disorders,* 14(203).
16. Karel, Y., Scholten-Peeters, W., Thoomes-de Graaf, M., Duijn, E., Ottenheijm, R., van den Borne, M., .et.al. (2013). Current management and prognostic factors in physiotherapy practice for patients with shoulder pain: design of a prospective cohort study.

BioMedCentral Musculoskeletal Disorders, 14(62).

17. Holtermann, A., Hansen, J., Burr, H., & Sogaard, K. (2010). Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. *Scand J Work Environ Heal.*, 36(1), 34-41.
18. Zengh, X., Simpson, J., van der Windt, D., & Elliott, A. (2005). Data from a study of effectiveness suggested potential prognostic factors related to the patterns of shoulder pain. *Journal of Clinical Epidemiology*, 58(8), 823-830.
19. Ginn, K., & Cohen, M. (2003). Conservative treatment for shoulder pain: prognostic indicators of outcome. *Archives of physical medicine and. Rehabilitation*, 85(8), 1231-1235.
20. Karels, C., Bierma-Zeinstra, S., Burdorf, A., Verhagen, A., Nauta, A., & Koes, B. (2007). Social and psychological factors influenced the course of arm, neck and shoulder compaints. *Journal of Clinical Epidemiology*, 60, 839-848.
21. Thoomes-de Graaf, M., Scholten-Peeters, G., Duijn, E., Karel, Y., Koes, B., & Verhagen, A. (2014). The Dutch Shoulder Pain and Disability Index (SPADI): a reliability and validation study. *Quality of Life Research*, 24(6), 1515-9.
22. Stinckens, N., Ulburghs, A., & Claes, L. (2009). De werkalliantie als sleutelelement in het therapie gebeuren; meting met behulp van de WAV-12, de nederlandstalige verkorte versie van working alliance inventory. *Tijdschrift Klinische Psychologie*, 39, 44-60.
23. Hatcher, R., & Gillaspy, J. (2006). Development and validation of a revised short version of the working alliance inventory. *Psychotherapy Research*, 16, 12-15.
24. Kamper, S., Ostelo, R., Knol, D., Maher, C., de Vet, H., & Hancock, M. (2010). Global perceived effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of Clinical Epidemiology*, 63(7), 760-766.
25. Brooks, R. (1996). EuroQol: the current state of play. *Health Policy*, 37(1), 53-72.
26. Bekkering, G., Hendriks, H., & Van Tulder, M. (2005). Prognostic factors for low back pain patients referred for physiotherapy. *Spine*, 30(16), 1881–1886.
27. Van der Windt, D., Koes, B., Boeke, A., Deville, W., De Jong, B., & Bouter, L. (1996). Shoulder disorders in general practice: prognostic indicators of outcome. British *Journal of General Practitioners*, 46, 519-523.
28. Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373-9.
29. Moons, K., Donders, R., Stijnen, T., & Harrell, F. J. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, 59, 1092-1101.
30. Janssen, K., Vergouwe, Y., Donders, A., Harrell, F. J., Chen, Q., Grobbee, D., & Moons, K. (2009). Dealing with Missing Predictor Values When Applying Clinical Prediction Models. *Clinical Chemistry*, 55(5).
31. van Buuren, S., Boshuizen, HC., Knook, DL. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine,* 18(6), 65-92.
32. Marshall, A., Altman, D., Holder, R., & Royston, P. (2009). Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*, 9(57). doi:10.1186/1471-2288-9-57
33. Hesselman Borg, J., Westerståhl, M., Lundell, S., Madison, G., Aasa, U. (2016). Longitudinal study exploring factors associated with neck/shoulder pain at 52 years

of age. *Journal of Pain Research*, 24(9), 303-10

34. Heymans, M., van Buuren, S., Knol, D., van Mechelen, W., & de Vet, H. (2007). Variable selection under multiple imputation using bootstrap in a prognostic study. *BMC Medical Research Methodology*, *7*(33). doi:10.1186/1471-2288-7-33
35. Hosmer, D., Lemeshow, S., & Sturdivant, R. *Applied Logistic Regression*. New Jersey: John Wiley & Sons; 2013.
36. Steyerberg, E., Harrell Jr, F., Borsboom, J., Eijkemans, M., Vergouwe, Y., & Habbema, J. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, *54*(8), *774–781*.
37. Efron, B., & Tibshirani, R. *An introduction to the bootstrap*. Monographs on statistics and applied probability. New York: Chapman & Hall; 1993.
38. R Core Team. *R: A language and environment for statistical computing*; 2012. Available from R Foundation for Statistical Computing: http://www.R-project.org/
39. Croft, P., Pope, D., & Silman, A. (1996). The clinical course of shoulder pain: prospective cohort study in primary care. Primary Care Rheumatology Society Shoulder Study Group. *British Medical Journal*, 313, 601-602.
40. Mintken, P., Cleland, J., Carpenter, K., Bieniek, M., Keirns, M., & Whitman, J. (2010). Some factors predict successful short-term outcomes in individuals with shoulder pain receiving cervicothoracic manipulation: a single-arm trial. *Physical Therapy*, 90(1), 26-42.
41. Engebretsen, K., Grotle, M., Bautz-Holter, E., Ekeberg, O., & Brox, J. (2010). Predictors of shoulder pain and disability index (SPADI) and work status after 1 year in patients with subacromial shoulder pain. *BMC Musculoskeletal Disorders*, 11(218).
42. Deutscher, D., Horn, S., Dickstein, R., Hart, D., Smout, R., Gutvirtz, M., & Ariel, I. (2009). Associations between treatment processes, patient characteristics, and outcomes in outpatient physical therapy practice. *Clinical Rheumatology*, 90(8), 1349-63.
43. Thomas, E., van der Windt, D.A., Hay, E.M.,Smidt, N., Dziedzic, K., Bouter, L.M., Croft, P.L. (2005). Two pragmatic trials for shoulder disorders in primary care: generalisability, course and prognostic indicators. *Annals of Rheumatic Diseases*, 64(7), 1056-61.
44. Bingefors, K., Isacson, D. (2004). Epidemiology, co-morbidity, and impact on health-related quality of life of self-reported headache and musculoskeletal pain – a gender persepective. *European Journal of Pain*, 8(5), 435-50
45. Coronado, R.A., Alappattu, M.J., Hart, D.L., George, S.Z. (2011). Total number and severity of comorbidities do not differ based on anatomical region of musculoskeletal pain. *Journal of Orthopaedis Sports and Physical Therapy*, 41(7), 477-85
46. George, S.Z., Beneciuk, J.M., Bialosky, J.E., Lentz, T.A., Zeppeiri, G. jr., Pei, Q., Qu, S.S. (2015). Development of a Review-of-Systems Screening Tool for Orthopaedic Physical Therapists: Results from the Optimal Screening for Prediction of Referral and Outcome (OSPRO) Cohort. *Journal of Orthopaedic Sports and Physical Therapy*, 45(7), 512-26
47. De Bruijn, C., de Bie, R., Geraets, J., Goossens, M., van den Heuvel, W., van den Heijden, G., et.al. (2007). Effect of an education and activation programme on functional limitations and patient-perceived recovery in acute and sub-acute shoulder complaints – a randomized clinical trial. *BMC Musculoskeletal Disorders*, 8(112).
48. Abdel-Tawaba, N., Roter, D. (2002). The relevance of client-centered communication to family planning settings in developing

countries: lessons from the Egyptian experience. *Soc Sci Med.*, 54, 1357–1368

49. Schönberger, M., Humle, F., Zeeman, P., Teasdale, TW. (2006). Working alliance and patient compliance in brain injury rehabilitation and their relation to psychosocial outcome. *Neuropsychol Rehabil*, 16, 298–314.
50. Peduzzi, P., Concato, J., Kemper, E., Holford, T., & Feinstein, A. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49, 1373-1379.
51. Steyerberg, E., Bleeker, S., Moll, H., Grobbee, D., & Moons, K. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, 56, 441-447.
52. Steyerberg, E., Harrell Jr, F., Borsboom, J., Eijkemans, M., Vergouwe, Y., & Habbema, J. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54(8), 774–781.
53. Collins, GS., Reitsma, JB., Altman, DG., Moons, KGM. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Medicine*, 13(1).
54. Vergouw, D., Heymans, M.W., de Vet, H.C.W., van der Windt, D.A.W.M., van der Horst, H.E. (2011). Prediction of persistent shoulder pain in general practice: Comparing clinical consensus from a Delphi procedure with a statistical scoring system. *BMC Family Practice*, 12(63).