Title: Self-Efficacy and Risk of Persistent Shoulder Pain: Results of a Classification and Regression Tree (CaRT) Analysis

Rachel Chester, Mizanur Khondoker, Lee Shepstone, Jeremy Lewis, Christina Jerosch-Herold

School of Health Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK.
Rachel Chester
Lecturer in Physiotherapy

Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK.
Mizanur Khondoker
Senior Lecturer in Medical Statistics

Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK.
Lee Shepstone
Professor of Medical Statistics

Department of Allied Health Professions, School of Health and Social Work, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK.
Jeremy Lewis
Professor of Musculoskeletal Research

School of Health Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK.
Christina Jerosch-Herold
Professor of Rehabilitation Research

Correspondence to: R Chester r.chester@uea.ac.uk 0044 (0)1603 593571

Word count: 3112 excluding text box, tables, figures and their legends. 3252 including text box.

Abstract

Objectives: To (i) identify predictors of outcome for the physiotherapy management of shoulder pain and (ii) enable clinicians to subgroup people into risk groups for persistent shoulder pain and disability.

Methods: 1030 people aged ≥18 years, referred to physiotherapy for the management of musculoskeletal shoulder pain were recruited. 810 provided data at 6 months for 4 outcomes: Shoulder Pain and Disability Index (SPADI) (total score, pain sub-scale, disability sub-scale) and Quick Disability of the Arm, Shoulder and Hand (QuickDASH). 34 potential prognostic factors were used in this analysis.

Results: Four classification trees (prognostic pathways or decision trees) were created, one for each outcome. The most important predictor was baseline pain and/or disability: higher or lower baseline levels were associated with higher or lower levels at follow up for all outcomes. One additional baseline factor split participants into four subgroups. For the SPADI trees, high pain self-efficacy reduced the likelihood of continued pain and disability. Notably, participants with low baseline pain but concomitant low pain self-efficacy had similar outcomes to patients with high baseline pain and high pain self-efficacy. Cut points for defining high and low pain self-efficacy differed according to baseline pain and disability. In the QuickDASH tree, the association between moderate baseline pain and disability with outcome was influenced by patient expectation: participants who expected to recover because of physiotherapy did better than those who expected no benefit.

Conclusions: Patient expectation and pain self-efficacy are associated with clinical outcome. These clinical elements should be included at the first assessment and a low pain self-efficacy response considered as a target for treatment intervention.

1

What are the new findings?

- High levels of pain and disability at baseline are associated with high levels of pain and disability at 6 month follow up. However, this 'predicted' poor outcome is modified to a predicted better outcome if the patient has high pain self-efficacy and a greater expectation of treatment.
- (Pain self-efficacy is the extent or strength of the patient's belief in their ability to complete tasks and reach a desired outcome despite their shoulder pain).
- Low levels of baseline pain and disability are associated with low levels of follow up pain and disability. This predicted better outcome is modified to a predicted poor outcome if the patient has low pain self-efficacy.

How might it impact on clinical practice in the future?

- We recommend that pain self-efficacy and patient expectation of outcome as a result of physiotherapy treatment should be formally assessed and discussed at the first physiotherapy appointment.

2

Introduction

Persistent musculoskeletal shoulder pain is common and frequently associated with substantial disability. Over a period of one month, between 16% and 31% of the general population in the United Kingdom (UK) will have suffered from musculoskeletal shoulder pain lasting at least 24 hours.[1-3] Experiencing shoulder pain concerns many and accounts for up to three percent of visits to General Practitioners (GP) annually.[4,5] and up to 48% of people with shoulder pain visit their GP more than once over a three year period due to ongoing symptoms.[5,6] The most effective treatment is not yet known; clinical trials comparing surgical and non-surgical management, including exercises prescribed by physiotherapists report equivocal effects.[7-9] Between 8% and 11% of patients visiting their GP with shoulder pain are referred to see a physiotherapist at initial consultation,[5,10.11] rising to 18% over a three year period.[5]

Response to physiotherapy is variable. In a multicentre cohort study of 1030 patients with musculoskeletal shoulder pain attending physiotherapy, of mean duration 14 months (standard deviation 28 months), 69% of patients reported complete recovery or being much improved by 6 months follow-up; 17% reported only slight improvement and 14% reported no change or a worsening of symptoms.[12] A multivariable general linear model (GLM) was used to identify prognostic factors associated with patient rated pain and disability.[13] Several factors were consistently associated with a better outcome at 6 months. One limitation of the GLM approach is the difficulty of practical use, particularly in a clinical setting. All predictor variables within the model are used simultaneously requiring lengthy calculations (particularly when there are many predictor variables).

Classification and Regression Tree (CART) Analysis is an alternative method of providing prognostic guidance. CART analysis considers the predictive value of prognostic factors sequentially, i.e. in a hierarchy of importance. CART typically results in a simple and readily interpretable decision 'tree' (or "what if" flow diagram). This can be graphically represented easily and requires no numeric calculations.[14] This can help guide clinicians to prioritise their initial prognostic assessment to those factors which are most influential. When modifiable, prognostic factors may become targets for interventions and inform shared decision making between clinicians and patients. The objective of these analyses was to provide clinicians with a guide to the most influential factors that predict outcome for people undergoing management for non-surgical musculoskeletal shoulder pain.

Methods

Data were available on 1030 people with shoulder pain recruited from primary and secondary care to a multicentre longitudinal cohort study in the East of England, UK, between November 2011 and October 2013. People aged 18 years or over were eligible to participate if they were referred to physiotherapy for the management of musculoskeletal shoulder pain and complained of shoulder or arm pain reproduced on movement of the shoulder. Those presenting with shoulder fractures, traumatic shoulder dislocations, systemic source of shoulder symptoms, cervical radiculopathy or had undergone shoulder surgery were excluded. Referral and treatment pathways were unaffected by participation in the study. The study protocol has been published.[15]

Outcome variables

Two validated patient rated outcome measures were collected at baseline and via postal questionnaire at six month follow up: the Shoulder Pain and Disability Index (SPADI),[16,17] and Quick Disability of the Arm, Shoulder and Hand (Quick-DASH).[18] The SPADI is a joint specific questionnaire designed to measure two domains: shoulder pain and disability. Thirteen items, five comprising a pain subscale and eight a disability subscale, are scored from zero to ten where zero represents no pain or disability and ten represents the worst pain imaginable or so difficult it requires help. For this analysis, each domain carried equal weighting in the overall score. The QuickDASH is an upper limb region specific questionnaire that includes items related to symptoms, daily activities, sleep, social and work function. Eleven items are scored from one to five where one represents no difficulty and five represents unable. Each item carries equal weighting in the final score. Scores are converted to a scale of zero to 100, where zero represents no pain or disability and 100 represents maximum pain and disability.

Baseline predictor variables:

Data for potential prognostic factors were collected prior to and during the participant's first physiotherapy appointment using bespoke questionnaires and clinical record forms. These variables included demographics, patient expectations and beliefs, lifestyle, general health, work, shoulder history and presentation, and clinical examination findings.[15] All factors statistically associated with outcome ($p \leq 0.05$) in at least one of the multivariable linear models from our previous analysis[13] were included in the CART analysis. In addition, baseline factors measured, but not found to be statistically significant, were entered if reported as a significant prognostic factor for outcome in reviews of other musculoskeletal studies.[19,20] A description of the variables included in the CART analysis are detailed in Supplementary file 1.

Patient involvement

Patient and public representatives were involved in the design of the study, in particular, details associated with the timing and procedures for recruiting and follow up of participants, and the design

and layout of questionnaires for data collection. A lay version of results, designed with patient and public representatives, were disseminated to all study participants who at their final data collection replied that they would like a copy. Patients were not involved in the actual recruitment or conduct of the study.

Statistical analysis

We used regression trees algorithms, a sub-class of Classification And Regression Tree (CART),[21] for continuous outcome variables, in our analyses. CART uses a recursive partitioning of the study sample to produce sub-groups as homogenous as possible with respect to the outcome of interest. This partitioning is based upon a binary split of each predictor variable. It is a more flexible approach for uncovering complex variable relationships than traditional linear modelling as it does not rely on any functional relationship between the outcome and predictor variables, nor does it require any distributional assumption regarding the outcome variable. CART is also less sensitive to outlying data, well suited for a large number of predictor variables and therefore offers a suitable alternative for building prediction models where the relationships among variables are unspecified and existing parametric statistical methods are not suitable to guide the model building.[22] The prediction accuracy of CART is comparable with parametric regression models and it can be more accurate when the relationship between the outcome and predictor variables is non-linear.[21] Furthermore, the partitioning in CART can be represented graphically as an easily interpretable decision tree[14] that may then be used to inform clinical practice.

We constructed four regression trees for each of the four outcome variables, i.e. SPADI overall score, SPADI pain subscore, SPADI disability subscore and QuckDASH score respectively. We used the R (R Core Team, 2015) package rpart.[23] For each of the 6-month outcome variables the respective baseline score was included within the list of predictor variables. Including the respective baseline scores, the number of predictor variables entered in different models ranged between 32 and 34. The pain and disability sub-scales were not included in the total SPADI tree analysis therefore using only 32 variables (see Supplementary file 1 for a complete list and definitions of variables).

The procedure for building a regression tree in rpart is performed as follows:[24]

Building a tree: First, the predictor variable is found which best splits the sample into two sub-groups. The 'best' is defined as the split that maximises the between groups sum-of-squares (or, equivalently, minimises the within-group error sum-of-squares). This process is applied separately to each sub-group recursively until the subgroups either reach a minimum size (set to 7 in our analysis, the default in rpart) or until no improvement can be made in the model fit.

Pruning the tree: The resultant model is typically too complex and likely to over-fit the data. The second stage of the procedure consists of using cross-validation to trim back the full tree. We used 10-

fold cross-validation to evaluate model fit at a series of model complexities and chose the optimal (pruned) tree by inspecting the plot of the cross-validated error against model complexity. Statistical analyses were completed by a statistician without prior knowledge of the clinical area or results of earlier analyses. Based on the cross validated predicted residual error sum of squares (PRESS) statistic, these trees provided the best predicting models. Retention of more baseline variables within the trees did not improve their predictive capacity. See supplementary file 2 for the plot for cross-validated errors versus model complexity for SPADI (total score) at 6 months.

Estimating a model and evaluating prediction accuracy on the same dataset generally over-estimates model performance. It is, therefore, recommended to evaluate the model performance on an independent dataset (i.e. a dataset that was not used to estimate the model). In the absence of an independent test dataset for model evaluation, cross-validation approach is an alternative way to create independent datasets for model assessment by holding apart a small portion of the sample for model evaluation. More specifically, 10-fold cross-validation involves randomly splitting the data into 10 parts of similar size, holding aside one part ($1/10^{th}$ of the whole sample) for testing and using the rest ($9/10^{th}$) for model estimation. The process is repeated 10 times, meaning that each of the 10 folds is used as independent test set for model evaluation. Overall model performance is typically assessed by calculating prediction errors on each of the 10 folds and averaging across all of these results.

Cross-validation is a widely used and acceptable way of validating model accuracy/performance and we used this approach to select the optimal CART model, i.e., to select the variables that are most predictive of the respective outcome variables (and also to remove those not contribution enough to the prediction model). The results of this validation process ensures that our selected CART models are optimal, despite not having a separate independent validation dataset.

Results

One thousand and fifty-five participants were assessed by physiotherapists and subsequently recruited and consented into the study. One thousand and thirty participants were found to be eligible for the study and provided adequate baseline data. There were no potential prognostic factors at baseline for which more than 2% of data were missing. Eight hundred and eleven participants (79%) provided outcome data at 6 month follow up. One participant was excluded due to incomplete outcome data. See flow diagram in figure 1.

Figure 1: STROBE flow diagram. Participant recruitment and follow-up

All participants providing complete outcome data at six months were included in the CART analysis for the SPADI and QuickDASH (n=810). There were differences between those participants who provided complete data at six months and those who did not. Completers were older by a mean of 10 years, had greater pain self-efficacy by a mean of almost 4 out of a possible 60 points, were almost twice as likely to exercise, and had a two-fold lower likelihood to report anxiety or depression. A summary description of baseline characteristics for all participants' (n=1030) for each of the 34 variables entered into the CART analysis are provided in supplementary file 3.

Figures 2-5 represent the resulting pruned regression trees for the total SPADI, SPADI pain subscale, SPADI disability sub scale, and QuickDASH at six months follow up. Three variables were identified as important predictors of six-month outcomes: 1) baseline pain or disability levels, 2) pain self-efficacy and 3) patient expectation of "change as a result of physiotherapy treatment". All three variables were collected prior to the participant's first physiotherapy attendance. Pain self-efficacy is the extent or strength of the patient's belief in their ability to complete tasks and reach a desired outcome despite their shoulder pain.[25] Pain self-efficacy was measured using the pain self-efficacy questionnaire (PSEQ)[26] which comprises of 10 items rated 0 to 6, zero representing minimum pain self-efficacy and 6 representing maximum pain-self efficacy. The total score is out of 60, a higher score representing higher pain self-efficacy. Patient expectation of change was collected in response to the following question "How much *do you expect* your shoulder problem to change as a result of physiotherapy treatment" and was measured on a 7 point Likert scale ranging from "completely recover" to "worse than ever".[15]

The first 'node' (at the top of the trees) represents the sample (i.e. all 810 participants). This then divides into two, based on cut-off values for baseline pain or disability (SPADI, SPADI subscale score or QuickDASH). The baseline score was therefore considered the most important variable in predicting the respective six-month outcome. In addition to baseline pain or disability, each pruned regression tree retained only one other variable of the 34 variables considered: baseline pain self-efficacy or patient expectation. Either pain self-efficacy or patient expectation led to classification of participants into four subgroups. The number of participants in these subgroups ranged from 48 to 487.

Figure 2: Regression tree for total SPADI score

Explanatory legend: Cut off points for the SPADI and PSEQ have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of total SPADI scores at 6 month follow up. The median SPADI score at 6 month follow up, (represented by the horizontal line

dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

Figure 3: Regression tree for SPADI Pain Subscale score

Explanatory legend: Explanatory legend: Cut off points for the SPADI Pain Subscale scores and PSEQ have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of total SPADI Pain Subscale scores at 6 month follow up. The median SPADI Pain Subscale score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

Figure 4: Regression tree for SPADI Disability Subscale Score

Explanatory legend: Explanatory legend: Explanatory legend: Cut off points for the SPADI Disability Subscale scores and PSEQ have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of total SPADI Disability Subscale scores at 6 month follow up. The median SPADI Disability Subscale score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

The cut point for baseline SPADI scores (total, pain and disability sub-scores) at the first node of each tree ranged from 62 to 75. When sub-dividing patients with lower baseline SPADI or baseline SPADI pain sub scores into two groups using baseline pain self-efficacy scores, the cut off for the PSEQ was 40 and 41 respectively. When sub-dividing patients with higher baseline SPADI pain or disability scores into two groups using pain self-efficacy scores, the cut off point for the PSEQ was consistently 48.

Figure 5: Regression Tree for QuickDASH

Explanatory legend: Cut off points for the QuickDASH scores have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of QuickDASH scores at 6 month follow up. The median QuickDASH score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

Table 1 and figures 2-4 show that the size of any subgroup with low pain self-efficacy ranged from 16% (n=127) to 20% (n=161) of participants in the cohort. The SPADI pain tree includes two subgroups with low pain self-efficacy and this constitutes as 36% (n=288) of the cohort. The discrimination between median outcome scores associated with different pain self-efficacy scores differs between trees and baseline pain and/or disability. For example, the median difference in subgroups is most marked for participants with high baseline SPADI pain subscores (≥75) and least for participants with lower baseline total SPADI scores (<68). Twenty three percent (n=239) of the cohort had a baseline of 41-59 on the QuickDASH, their outcomes were differentiated by their expectation of "change as a result of physiotherapy treatment": the median difference between subgroups at outcome being 23/100 on the QuickDASH.

Table 1: Median (interquartile range) outcome (SPADI total, SPADI pain subscale, SPADI disability subscale and QuickDASH) for each subgroup for each tree.

SPADI tree at 6 months

Baseline	Number (%)	Median	IQR
<68 SPADI,≥40 PSEQ	487 (60)	9	3 to 23
<68 SPADI, <40 PSEQ	140 (17)	25	10 to 49
68 to 81 SPADI	135 (17)	36	13 to 60
≥82 SPADI	48 (6)	66	27 to 80

SPADI Pain tree at 6 months

Baseline	Number (%)	Median	IQR
<75 SPADI Pain ,≥41 PSEQ	474 (58)	12	4 to 27
<75 SPADI Pain, <41 PSEQ	161 (20)	30	12 to 56
≥75 SPADI Pain, ≥48 PSEQ	48 (6)	20	12 to 56
≥75 SPADI Pain, <48 PSEQ	127 (16)	56	26 to 77

SPADI disability tree at 6 months

Baseline	Number (%)	Median	IQR
<42 SPADI Disability	404 (50)	5	1 to 13
42 to 61 SPADI Disability	203 (25)	15	5 to 39
≥62 SPADI Disability, ≥48 PSEQ	48 (6)	13	7 to 36
≥62 SPADI Disability, <48 PSEQ	155 (19)	44	18 to 69

QuickDASH tree at 6 months

Baseline	Number (%)	Median	IQR
<41 QuickDASH	474 (59)	9	2 to 18
41 to 59 QuickDASH, Pt expectation: CR or much improved	180 (22)	18	7 to 36

41 to 59 QuickDASH, Pt expectation: SI, same or worse	59 (7)	41	25 to 52
≥60 QuickDASH	97 (12)	45	27 to 61

Validation

External validation of the results was not possible as we were unable to identify an external dataset containing the same or similar variables. We have, however, conducted an informal internal validation of the results by partitioning the QuickDASH outcome data based on the classifications of the SPADI regression trees and comparing the distribution of QuickDASH outcome within each sub-group with that of the SPADI outcomes. Comparison of QuickDASH distributions corresponding to the total SPADI, SPADI pain and SPADI disability trees are displayed in Supplementary files 4, 5 and 6 respectively. The similarity of the pattern distributions of the SPADI outcome on the left and QuickDASH outcome on the right demonstrate the replicability of the SPADI tree.

Discussion

The objective of these analyses was to identify important predictors of outcome for patients presenting with non-surgically managed musculoskeletal shoulder pain. We identified that only three of 34 baseline variables considered in the classification trees were predictive of outcome. These were i) baseline pain or disability measured by the SPADI or QuickDASH, ii) pain self-efficacy measured by the PSEQ,[26] and iii) patient's expectation of "change as a result of physiotherapy treatment", measured on a 7-point Likert scale.[15] As expected, there was a positive association between pain and disability at baseline and at six month follow up, i.e. those with higher scores at baseline tended to have higher scores at follow-up. However, in all three SPADI classification trees higher pain self-efficacy influenced this relationship: for patients with high baseline pain or disability (cut off points 75 and 62 respectively), higher pain self-efficacy (PSEQ≥48) reduced the likelihood of continued high levels of pain and disability at six-month follow up. Between 16 and 19% of participants were at risk of continued high levels of pain and disability (measured by SPADI pain and disability subscores) at 6 months due to i) high baseline pain and disability and ii) low pain self-efficacy. For patients with moderate levels of baseline pain and disability measured with the QuickDASH (41 to 59), the association was influenced by patient expectation: participants who expected to completely recover or much improve as a result of physiotherapy did better than patients who expected to only slightly improve, stay the same or worsen. Participants at risk of continued high levels of pain and disability at six-month follow up due to a lower expectation of recovery constituted 7% of our cohort at six month follow up.

Pain self-efficacy also influenced outcome for patients with low levels of baseline pain and disability: for patients with low baseline SPADI and SPADI pain scores (<68 and <75 respectively), low pain self-efficacy (PSEQ<40 and <41 respectively) increased the likelihood of persistent pain. Perhaps surprisingly, patients reporting low baseline pain but low pain self-efficacy (n=161, 20% of cohort) had a similar or worse outcome on the SPADI pain subscale to patients with high baseline pain but high pain self-efficacy (n=48, 6% of cohort).

Our regression tree analyses provide a useful and simple clinical guide, highlighting the influence of patient beliefs and expectations of treatment on outcome, irrespective of baseline pain and disability. Whilst these finding are consistent with those from the GLM analysis,[13] the CART analysis selects variables based on prediction power rather than statistical significance or p values. Variables are included in order of importance; the most predictive variable is included first, the analysis then searches for the second most important variable among the rest, and so on. The prediction error curve estimated using cross-validation gives a clear indication at what point in the selection process the additional predictors are not contributing enough to the prediction model. The prediction based variable selection combined with cross-validation for assessing model performance ensures that only the relevant and most predictive variables are included in the optimal model.

CART analysis has advantages over traditional regression modelling in that it does not require a specified distribution of outcome data or a large sample size[27]. In terms of predictive power CART analysis is comparable to traditional modelling.[21] However CART does have limitations. Defining subgroups based on data driven cut-points for continuous measures (i.e. the PSEQ) is subject to sampling variability, but the CART methodology does not provide a measure of uncertainty (e.g., standard errors or confidence intervals) associated with the cut-off points. A different cut off point may be selected in a different sample, but it was not our intention to provide a ready to deploy clinical tool with definitive cut-off points at this stage. We rather aimed to demonstrate that an easily interpretable prediction tool with potential for clinical applications can be developed which can be further examined in bigger and external cohorts to derive more generalisable cut-offs. However, use of cross-validation approach for model selection should make the derived models sufficiently robust at least for the population represented by the study cohort. Also, being a multicentre study with broad eligibility criteria increases the generalisability of the results to the wide range of patients and presentations of shoulder pain commonly seen by physiotherapists within primary and secondary care. This is further supported by similar patterns of the distributions of the QuickDASH outcome based on classification of participants using the SPADI trees in our informal internal validation.

With regards to non-surgically managed shoulder pain, this study is one of only two known using a CART analysis to investigate the hierarchy of predictive factors associated with outcome. Vergouw et

al[28] compared the results of CART and logistic linear regression for 587 patients with musculoskeletal shoulder pain attending General Practice in the Netherlands, however, they did not include patient expectation of change as a result of physiotherapy and pain self-efficacy. A positive association between patient expectation and outcome has been consistently reported for a range of health conditions,[29-31] although ours is the first to investigate patient expectation of outcome in non-surgically managed shoulder pain. The association between pain self-efficacy and chronic non-cancer pain has also been consistently reported for a range of health conditions.[32] Ours is one of only two studies to investigate self-efficacy in non-surgically management for shoulder pain. A randomised controlled trial of 102 participants,[33] did not find an association between baseline pain self-efficacy and the outcome of supervised exercise or radial extracorporeal shockwave therapy.

Based on our findings that pain self-efficacy and patient expectation are important predictors of outcome we recommend that they be formally assessed in all patients with musculoskeletal pain. There is currently no standardised method of measuring patient expectation and we therefore recommend using a patient rated Likert scale that includes a worsening as well as improvement of shoulder pain.[29] There are several validated measurement tools for pain self-efficacy and for the busy clinician we recommend using shortened patient rated versions such as the PSEQ-2[34] comprising two items. Standardised questionnaires like the PSEQ-2 and a single question on expectation of outcome provide an opportunity to openly discuss patient beliefs and expectations which healthcare practitioners may find challenging otherwise.[35, 36] Such patient-clinician dialogues around the potential impact of expectations and beliefs further supports shared decision-making. Our results suggest that cut points will vary according to baseline pain and disability and therefore the use of specific cut-points for stratification is not justified. Further research is also needed to validate our point estimates in an external cohort.

It is plausible that patient expectation and pain self-efficacy are mediating factors.[29] Adherence to non-surgical management is reportedly low.[37] The therapeutic effect of a home exercise and/or self-management programme cannot be realised if not enacted by the patient.[38] One of the suggested mechanisms by which higher patient expectation is associated with outcome is through an increased motivation to engage and adhere to an intervention that participants believe will have a beneficial outcome.[39]

Although not previously reported for those experiencing shoulder pain, high self-efficacy has been shown to be significantly associated with greater exercise adherence [40-42] as well as other health behaviours such as physical activity[43-45] and taking medications as prescribed.[46] A consistent and statistically significant association between all three factors; changes in self-efficacy, adherence and

outcome, has yet to be demonstrated. Further studies are needed to explore if moderating self-efficacy affects outcome.

Further development and testing of, educational interventions targeting healthcare practitioners with strategies to increase patients' pain self-efficacy and expectations of treatment is needed. A number of promising interventions exist for increasing patients' self-efficacy and include positive feedback on performance, observation of mastery in others, graded activity, identifying realistic goals for which the patient is likely to succeed and selecting tasks and activities relevant to the patient.[32,43] Variability in reported effectiveness suggests that the purpose, content and delivery may need to be tailored to each patient, requiring a person centred approach.

Conclusion

This is the first known study to subgroup people with shoulder pain of musculoskeletal origin attending physiotherapy into risk groups for persistent pain and disability based on a range of baseline personal, clinical, activity, and participatory variables. This multicentre study provides evidence that for a given baseline measure of shoulder pain and disability, pain self-efficacy and patient expectation of change as a result of physiotherapy, are the most influential predictors of patient rated outcome at six month follow up. Additionally, this is the first study to demonstrate that for people with shoulder pain higher pain self-efficacy reduced the likelihood of continued high levels of pain and disability at six-month follow up, for those with high baseline pain or disability. The likelihood of persistent pain increased in the subgroup that were categorised as having low levels of baseline pain and disability and concomitant low pain self-efficacy. Of importance those identified as having low baseline pain and low self-efficacy had similar or worse outcome on the SPADI pain subscale to those with high baseline pain and high pain self-efficacy.

Although our findings are applicable to people referred to physiotherapy for the management of shoulder pain of any duration and in primary and secondary care, they are likely to be applicable beyond this group.

Based on our findings we suggest that pain self-efficacy and patient expectation should be formally assessed and discussed at the first physiotherapy appointment. Further research should investigate whether these factors can be targeted and modified by therapeutic interventions and improve patient outcomes.

Declaration of competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that (1) RC had support from a National Institute of Health Clinical Doctoral Research Fellowship for the submitted work; (2) the authors have no relationships with companies that might

have an interest in the submitted work in the previous 3 years; (3) the authors' spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (4) the authors have no non-financial interests that may be relevant to the submitted work."

Contributors: RC conceived the study, RC, LS, JL and CJH authors contributed to the design of the study. RC directed the study and managed data acquisition. RC had full access to all of the data in the study and takes responsibility for the integrity of the data. MK undertook and LS provided expert advice on the statistical analyses. RC and MK drafted the manuscript. LS, JL and CJH provided additional important intellectual and substantial scientific input throughout the study and on all redrafts of the report.

Ethics approval: This was obtained from the National Research Ethics Service, East of England - Norfolk, UK in July 2011 (reference 11/EE/0212).

Funding: RC was funded by the National Institute for Health Research (NIHR CAT CDRF 10–008). CJH was funded by the National Institute for Health Research (NIHR- SRF-2012-05-119). The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report or decision to submit the article for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR, NHS or the Department of Health.

Transparency declaration: The lead author (the manuscript's guarantor) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Data sharing: No additional data available.

Patient involvement: Patient and public representatives were involved in the design of the study, in particular, details associated with the timing and procedures for recruiting and follow up of participants, and the design and layout of questionnaires for data collection. A lay version of results,

designed with patient and public representatives, were disseminated to all study participants who at their final data collection replied that they would like a copy. Patients were not involved in the actual recruitment or conduct of the study.

Acknowledgements: The authors would like to thank Christine Christopher and members of the study steering group, including representatives from the Public & Patient Involvement in Research (PPIRes) project, for their comments and suggestions during the study. We would also like to thank the Principal Investigators and physiotherapists who contributed and collected data for the study and to the patients who generously gave their time and participated in the study.

References

1. Parsons S, Ingram M, Clarke-Conwell A, Symmons DPM. A heavy burden. The occurrence and impact of musculoskeletal conditions in the United Kingdom today. The University of Manchester 2011. Accessible online: http://www.medicine.manchester.ac.uk/musculoskeletal/aboutus/publications/heavyburden.pdf

2. Pope DP, Croft PR, Pritchard CM, Silman AJ. Prevalence of shoulder pain in the community: the influence of case definition. *Ann Rheum Dis* 1997;56:308-12.

3. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis* 1998;57:649-655.

4. Bot SDM, van der Waal JM, Terwee CB, et al. Incidence and prevalence of complaints of the neck and upper extremity in general practice. *Ann Rheum Dis* 2005;64:118-123.

5. Linsell L, Dawson J, Zondervan K, Randall RT, Fitzpatrick R, Carr A et al. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patterns of diagnosis and referral. *Rheumatology (Oxford)* 2006;45:215–21.

6. van der Windt DAWN, Koes BW, De Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis* 1995;54:959-64.

7. Dorrestijn O, Stevens M, Winters JC, van der Meer K, Diercks RL. Conservative or surgical treatment for subacromial impingement syndrome? A systematic review. *J Shoulder Elbow Surg* 2009;18:652-660.

8. Ketola S, Lehtinen JT, Arnala I. Arthroscopic decompression not recommended in the treatment of rotator cuff tendinopathy a final review of a randomised controlled trial at a minimum follow-up of ten years. *Bone Joint J* 2017;99-B:799–805.

9. Saltychev M, Ville Äärimaa V, Virolainen P, Laimi K. Conservative treatment or surgery for shoulder impingement: systematic review and meta-analysis. *Disabil Rehabil* 2015;37:1-8.

10. Kooijman MK, Swinkels ICS, Dijk C, Bakker D, Veenhof C. Patients with shoulder syndromes in general and physiotherapy practice: an observational study. *BMC Musculoskelet Disord* 2013;14:1-13.

11. Reilingh ML, Kuijpers T, Tanja-Harfterkamp AM, van der Windt DAWN. Course and prognosis of shoulder symptoms in general practice. *Rheumatology (Oxford)* 2008;47:724-730.

12. Chester R, Jerosch-Herold C, Lewis J, Shepstone L. SPADI and QUICKDASH are Similarly Responsive in Patients Undergoing Physical Therapy for Shoulder Pain. *J Orthop Sports Phys Ther* 2017;47:538-547.

13. Chester R, Jerosch-Herold C, Lewis J, Shepstone L. Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: a multicentre longitudinal cohort study. *Br J Sports Med* 2016;0:1–8.

14. Loh WY. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2011;1:14-23.

15. Chester R, Shepstone L, Lewis JS, Jerosch-Herold C. Predicting response to physical therapy treatment for musculoskeletal shoulder pain: protocol for a longitudinal cohort study. *BMC MusculoskeletDisord* 2013;14:1-8.

16. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. *BMC Musculoskelet Disord* 2006;7(12):1-11.

17. Roach KE, Budiman-Mak E, Songsiride JN, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care & Research* 1991;4:143-149.

18. Institute for Work and Health. The QuickDASH outcome measure. Information for Users. 2006 Available at: http://www.dash.iwh.on.ca/sites/dash/files/downloads (Accessed November 2016)

19. Artus M, van der Windt DA, Afolabi EF, et al. Management of shoulder pain by UK general practitioners (GPs): a national survey. *BMJ Open* 2017;7:e015711. doi:10.1136/bmjopen-2016-015711

20. Chester R, Shepstone L, Daniell H, Sweeting D, Lewis J, Jerosch-Herold C. Predicting response to physiotherapy in the treatment of musculoskeletal shoulder pain: a systematic review. BMC Musculoskeletal Disorders 2013;14(1):203. DOI:10.1186/1471-2474-14-203

21. Breiman L, Friedman J, Stone C J, Olshen RA. Classification and regression trees. CRC press. 1984

22. Taxman FS, Kitsantas P. Availability and capacity of substance abuse programs in correctional settings: A classification and regression tree analysis. *Drug Alcohol Depend* 2009;103:S43-S53.

23. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015 Accessible online: https://www.R-project.org

24. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the RPART routines. Mayo Foundation: Technical report 1997;61:452.

25. Bandura A. Self-efficacy - toward a unifying theory of behavioural change. *Psychological Review* 1977;84:191-215.

26 Nicholas MK. The pain self-efficacy questionnaire: taking pain into account. *Eur J Pain* 2007;11:153-163.

27 Phelps, M.C. and Merkle, E.C., Classification and Regression Trees as Alternatives to Regression. Proceedings of the 4th Annual GRASP Symposium, Wichita State University, 2008.

28 Vergouw D, Methodological issues of clinical prediction models for shoulder pain in general practice [dissertation] EMGO Institute for Health and Care Research: VU University 2013

29. Bialosky JE, Bishop MD, Cleland JA. Individual expectation: an overlooked, but pertinent, factor in the treatment of individuals' experience musculoskeletal pain. *Phys Ther* 2010; 90:1345-55.

30. Bishop MD, Mintken P, Bialosky JA, Cleland JE. Patient Expectations of Benefit From Interventions for Neck Pain and Resulting Influence on Outcomes. *J Orthop Sports Phys Ther* 2013;43:457-465.

31. Mondloch MV, Cole DC, Frank JW. Does how you do depend on how you think you'll do? A systematic review of the evidence for a relation between patients' recovery expectations and health outcomes. *CMAJ* 2001;165:174-179.

32. Jackson T, Wang Y, Wang Y, Fan H. Self-Efficacy and Chronic Pain Outcomes: A Meta-Analytic Review. *J Pain* 2014;15:800-814.

33. Engebretsen K, Grotle M, Bautz-Holter E, Ekeberg OM, Brox JI. Predictors of shoulder pain and disability index (SPADI) and work status after 1 year in patients with subacromial shoulder pain. *BMC Musculoskelet Disord* 2010;11:1-9.

34. Nicholas MK, McGuire BE, Asghari A. A 2-Item Short Form of the Pain Self-Efficacy Questionnaire: Development and Psychometric Evaluation of PSEQ-2. *J Pain* 2015;16:153-163.

35. Alexanders J, Anderson A, Henderson S. Musculoskeletal physiotherapists' use of psychological interventions: a systematic review of therapists' perceptions and practice. *Physiother* 2015;101:95-102.

36. Singla M, Jones M1, Edwards I, Kumar S. Physiotherapists' assessment of patients' psychosocial status: Are we standing on thin ice? A qualitative descriptive study. *Man Ther* 2015;20:328-334

37. Peek K, Sanson-Fisher R, Mackenzie L, Carey M. Interventions to aid patient adherence to physiotherapist prescribed self-management strategies: a systematic review. *Physiother* 2016; 102:127–135.

38. Hayden JA, van Tulder MW, Tomlinson G. Systematic review: Strategies for using exercise therapy to improve outcomes in chronic low back pain. *Ann Intern Med* 2005;142:776-785.

39 Flood AB, Lorence DP, McPherson K, Black NA.The role of expectations in patients' reports of post-operative outcomes and improvement following therapy. *Med Care* 1993;31:1043-56.

40. Medina-Mirapeix F, Escolar-Reina P, Gascón-Cánovas JJ, Montilla-Herrador J, Jimeno-Serrano FJ, Collins SM. Predictive factors of adherence to frequency and duration components in home exercise programs for neck and low back pain: an observational study. BMC Musculoskelet Disord 2009,10:155

41. Dumoulin C, Alewijnse D, Bo K, et al. Pelvic-floor-muscle training adherence: tools, measurements andstrategies – 2011 ICS State-of-the-Science Seminar Research Paper II of IV. *Neurourol Urodyn* 2015;34:615–21.

42. Hay-Smith EJC, McClurg D, Frawley H, Dean SG. Exercise adherence: integrating theory, evidence and behaviour change techniques. Physiotherapy 2016;102:7–9.

43. Ashford S, Edmunds J, French DP. What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *Br J Health Psychol* 2010;15:265–288.

44. Bachmann C, Oesch P, Bachmann S. Recommendations for Improving Adherence to Home-Based Exercise: A Systematic Review. *Phys Med Rehab Kuror* 2018;28: 20-31.

45. Tierney S, Mamas M, Woods S, Rutter MK, Gibson M, Neyses L, Deaton C. What strategies are effective for exercise adherence in heart failure? A systematic review of controlled studies. *Heart Fail Rev* 2012;17:107-115.

46. Náfrádi L, Nakamoto K, Schulz PJ. Is patient empowerment the key to promote adherence? A systematic review of the relationship between self-efficacy, health locus of control and medication adherence. *PLoS ONE* 2017;12(10): e0186458. https://doi.org/10.1371/journal.pone.0186458

Legends:

Figure 1: STROBE flow diagram. Participant recruitment and follow-up

Figure 2: Regression tree for total SPADI score

Explanatory legend: Cut off points for the SPADI and PSEQ have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of total SPADI scores at 6 month follow up. The median SPADI score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

Figure 3: Regression tree for SPADI Pain Subscale score

Explanatory legend: Explanatory legend: Cut off points for the SPADI Pain Subscale scores and PSEQ have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of total SPADI Pain Subscale scores at 6 month follow up. The median SPADI Pain Subscale score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

Figure 4: Regression tree for SPADI Disability Subscale Score

Explanatory legend: Explanatory legend: Explanatory legend: Cut off points for the SPADI Disability Subscale scores and PSEQ have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of total SPADI Disability Subscale scores at 6 month follow up. The median SPADI Disability Subscale score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.

Figure 5: Regression Tree for QuickDASH

Explanatory legend: Cut off points for the QuickDASH scores have been rounded up or down to whole numbers. The 4 boxplots at the bottom of the figure illustrate the distribution of QuickDASH scores at 6 month follow up. The median QuickDASH score at 6 month follow up, (represented by the horizontal line dissecting the box), is lowest (better outcome) in the subgroup represented by the box furthest left and highest (poorer outcome) in the subgroup represented by the box furthest right.