Title

Screening patients with acute low back pain

Permalink

https://escholarship.org/uc/item/5zh680zn

Journal

European Journal of Pain, 19(3)

ISSN

1090-3801

Authors

Mehling, WE
Avins, AL
Acree, MC
et al.

Publication Date

2015-03-01

DOI

10.1002/ejp.615

Peer reviewed

Can a back pain screening tool help classify patients with *acute* pain into risk levels for chronic pain?

W.E. Mehling[1,2], A.L. Avins[3,4], M.C. Acree[2], T.S. Carey[5], F.M. Hecht[2,3]

1 Department of Family Medicine, University of California San Francisco, USA
2 Osher Center for Integrative Medicine, University of California San Francisco, USA
3 Department of Medicine, University of California San Francisco, USA
4 Division of Research, Kaiser Permanente Northern California, Oakland, USA
5 Sheps Center for Health Service Research, University of North Carolina, Chapel Hill, USA

Correspondence
Wolf E. Mehling
E-mail: mehlingw@ocim.ucsf.edu

Abstract

Background: The 9-item STarT-Back screening tool was developed in primary care patients with low back pain (LBP) to identify those at greatest risk for chronic pain and requiring targeted treatment. We conducted a secondary data analysis study to examine the performance of comparable questionnaire items in a sample of primary care patients with well-defined *acute* LBP.

Methods: In a prospective cohort study, 605 primary care patients with LBP of less than 30 days answered a questionnaire with 6 items identical and 3 items analogous to the 9-item STarT-Back. Participants were followed up at 6 months and 2 years. STarT-Back rules were applied to classify participant's risk of chronic LBP, and the performance of the screening items in predicting outcomes was assessed using likelihood ratios.

Results: The proportion of patients with chronic pain at follow-up was considerably lower (6 months: 22%; 2 years: 25%) than in the STarT-Back validation cohort (40%) of patients with pain of any duration. The probability of developing chronic pain given a high-risk designation by items similar to the STarT-Back increased the pre-test probability to 31% and 35%. Likelihood ratios were close to 1.

Conclusions: A risk classification schema using the recommended cut-off scores with items similar to the STarT-Back in a primary care population with strictly defined *acute* LBP had limited ability to identify persons who progressed to chronic pain. The results suggest caution when applying the STarT-Back in patients with *acute* LBP and a need to consider a modification of its cut-offs.

1. Introduction

Most patients presenting with an episode of acute low back pain (LBP) in primary care will recover in 6–8 weeks with or without medical intervention (Pengel et al., 2003). Identifying which patients are at risk of developing chronic pain, which accounts for most LBP-related health expenses, is important for targeting more intensive treatment to those in whom it is most likely to be needed. Clinicians would like to know whether early targeted interventions may improve secondary prevention of chronic pain. Ideally, these interventions would be initiated as early as possible and before patients suffer from pain for several months, because patients who have not recovered by 2–3 months often become chronic LBP patients (Klenerman et al., 1995; Itz et al., 2013) and may already have a poorer prognosis for disability once they have pain for over 2 weeks (Kovacs et al., 2005).

One recently published screening tool was specifically developed for primary care patients, the STarT-Back Screening Tool, which was derived and carefully validated within the UK National Health System (Hay et al., 2008; Hill et al., 2008, 2010a,b, 2011). An item set closely approximating the STarT-Back performed similarly well when assessed by secondary data analysis using data from a US cohort (Von Korff et al., 2014). The STarT-Back is brief, with only 9 items, can be answered and scored in less than 5 min, and classifies patients into three levels of risk for LBP chronicity with different treatment indications. The items were systematically developed; each is scored dichotomously and summed for a total score (range 0–9) and a 5-item sub-score (range 0–5) for key psychological risk factors for chronic pain. Cut-offs for classification into risk groups were derived by receiver operating characteristic (ROC) curve analysis for reference standards at baseline. The authors reported positive and negative likelihood ratios for the cut-off between high and medium/low risk of 5.5 and 0.7, and between high/medium and low risk of 2.3 and 0.3, respectively, for the UK validation sample (Hill et al., 2008).

Patients enrolled in the STarT-Back samples were seen for a first doctor visit for LBP of *any* duration and not for strictly defined acute LBP; 62% of the STarT-Back validation sample had chronic LBP of >3 months, and would be expected to have a high risk of continued pain (Klenerman et al., 1995; Pengel et al., 2003), and 83% had pain for more than 1 month. Similarly, primary care patients in the US cohort had LBP of *any* duration, 41% had pain of less than 30 days and 18% of more than 3 months, with a mean of 66 days of pain in the past 6 months (Von Korff et al., 2014).

If the STarT-Back were sufficiently useful for early screening, it could potentially improve primary care pain management within the first month of pain onset

as well as it was reported for patients with LBP of *any* duration in the United Kingdom (Hill et al., 2011). To assess the performance of the STarT-Back in a US sample with more strictly defined *acute* LBP, for whom issues of prognosis and identification of remediable predictors of chronicity are particularly relevant, a new prospective cohort study needs to be undertaken. Until results from such a costly project are available, we thought it worthwhile to take advantage of available data from a prospective patient cohort with narrowly defined *acute* LBP that included analogous questionnaire items at baseline, and to conduct an exploratory secondary analysis.

The dataset available for this analysis is from a previously published 2-year cohort study, the Prognosis of Pain (POP) study (Mehling et al., 2012). The study was undertaken to determine bio-psycho-social risk factors for chronic pain (Fig. 1). Six of the nine STarT-Back items are identical, and the remaining three items are highly similar to items used in the POP study, allowing us to roughly estimate the performance of a similar item set in patients with acute LBP. This secondary data analysis is exploratory and cannot obviate the need for a definitive study of the original STarT-Back item set.

2. Methods

The sample comprises 605 members of Kaiser Permanente, Northern California, the largest regional integrated health plan, who were interviewed by phone and followed over 2 years; details of the sample have been previously published (Mehling et al., 2012). Briefly, the sample represented the socio-economic and ethnic diversity of primary care patients in the San Francisco Bay Area. Acute LBP was defined as back pain between rib cage and buttocks of less than 1 month, no LBP in the previous year and no prior spine surgery. The 1-month criterion was chosen, as patients often did not see their primary care provider within the first 2 weeks after pain onset. Time frames for classifying LBP as acute versus non-acute vary in the literature between 2 and 6 weeks (van Tulder et al., 1997; Kovacs et al., 2005); a Delphi approach to find consensus among researchers yielded a time frame of 4 weeks, similar to common clinical guidelines (Griffith *et al.*, 2007; Chou, 2014).

Electronic medical records were used to identify patients seen for LBP on the day after their clinic visit, who were sent an invitation by mail to join the study. Participants were interviewed over the phone at baseline and 6 months. For the 2-year follow-up, POP study participants, when contacted, were given a choice between a phone interview and an Internet-based survey using SurveyGizmo (2010). The study was approved by the Committee on Human Research of the University of California, San Francisco. Surveys were conducted between February 2008 and November 2010.
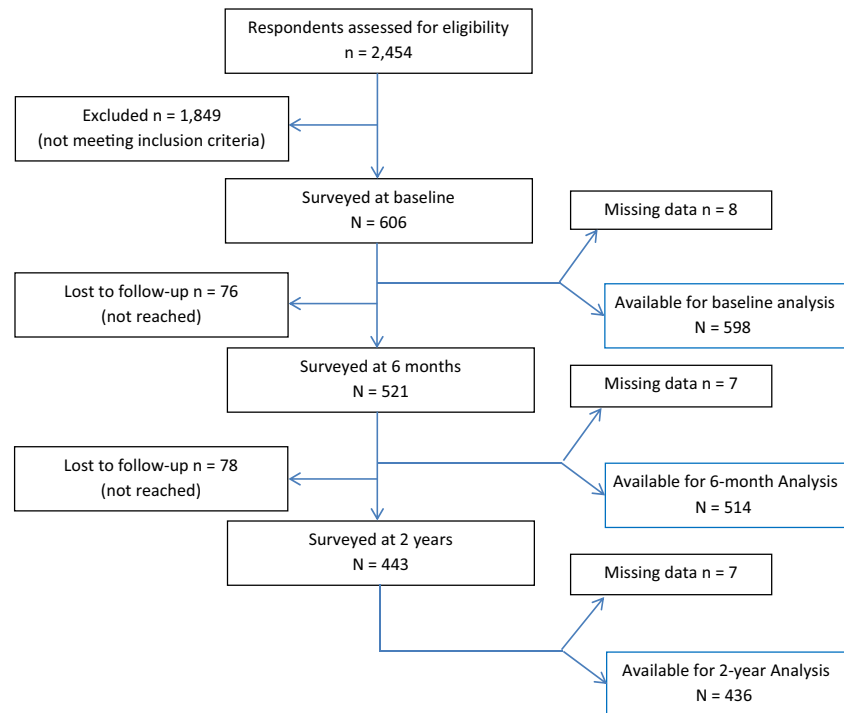
Figure 1 Flow diagram.

As no consensus exists about the definition of chronic pain (Mehling et al., 2011), we applied three outcome standards in our secondary data analysis: (1) a score of 7 or higher on the Roland–Morris Disability Questionnaire (RMDQ), a 24-point 'legacy measure' (Deyo et al., 2014) for functional disability due to LBP (Roland and Morris, 1983); a score of ≥7 was used as key outcome criterion in the original STarT-Back validation (Hill et al., 2008); (2) a Grade 2 or higher chronic pain level according to the Chronic Pain Risk by Von Korff, a complex measure that yields a 4-grade-level chronic pain score as a function of pain intensity and pain-related disability; it has been used in numerous studies worldwide, validated in the UK and compared with the STarT-Back (Von Korff et al., 1992, 2014; Dunn et al., 2008); (3) a recently proposed and systematically developed third measure combining a lack of perceived recovery (less than 'much improved') on a widely used single-item 6-point Likert Perceived Recovery Scale (Beurskens et al., 1996) and current pain intensity of 3 or more on 0–10 Pain Numeric Rating Scale (Krebs et al., 2007; de Vet et al., 2010; Mehling et al., 2011).

The following six POP study items were identical with STarT-Back items (for this paper we are using item numbering from the 2008 publication, slightly different from the website (Hill et al., 2008; Hill, 2011): bothersomeness of pain (#1), presence of pain radiating below the knee (#2), additional pain in neck or shoulders (#3), getting dressed more slowly (#5), walking only short distances due to LBP (#6), and whether the pain is felt to be terrible and never going to get better (#8). The three remaining STarT-Back items (numbers 4, 7 and 9) were analogous to items with corre-

sponding face validity taken from established questionnaires for the same psychological constructs of fear avoidance (#4), catastrophizing-rumination (#7), and depression (#9) presented in Table 1. From here on, we refer to these 9 items in the POP study as Nine Analogous Predictive Screening Items (9APSI). In the POP study, items 4, 7 and 9 were continuous variables (range 0–10; anchored 'never' and 'always'), which we dichotomized according to the STarT-Back website's information (the website includes a key for comparing the 'yes/no' screening tool version with a STarT-Back Clinical Measurement Tool version that uses a 0–10 scale for the same items (Hill, 2011). The POP time frame was the duration from onset (<4 weeks; mean 17 ± 8 days; median 14 days) or the past week, comparable to the past 2 weeks assessment in the STarT-Back.

We applied the same risk-level criteria developed for the STarT-Back: items were scored as dichotomous and summed. Five items served as a psychosocial subscale, and patients were allocated to the high-risk group if their *5-item subscores* were ≥4. The remaining patients were allocated to the low-risk group if the *overall 9-item scores* were <4, and to the medium-risk group if: (1) the *overall scores* were ≥4 and (2) the psychosocial *subscores* were <4. Using the three risk-level criteria of the STarT-Back, we tested how well the 9APSI classification predicted chronic pain (according to the three outcome definitions above) in the POP sample at two follow-up time points by calculating sensitivity, specificity, and positive and negative predictive value (PPV, NPV) in the POP sample using Stata statistical software (Stata12, 2013). Positive and negative likelihood ratios are strong indicators of the performance of a diagnostic instrument. The 'positive

Table 1 Comparing items applied in STarT-Back studies and in POP study with response format. Item numbers as in original publication.[b] Dichotomization according to website instruction for STarT-Back Clinical Measurement Tool.[c]

	STarT-Back	POP
Item 1 Pain intensity	'Overall, how bothersome has your back pain been in the last 2 weeks?' Response format: 'Not at all', 'Slightly', 'Moderately', 'Very much', 'Extremely'; dichotomized between 'Moderately' and 'Very much'	'On a scale from 0–10, how would you rate the average pain you have had during the past week?' Response format: Scale 0–10 dichotomized between 6 and 7
Item 2 Sciatica	'My back pain has spread down my leg(s) at some time in the last 2 weeks.' Response format: 'agree' or 'disagree'	'Did your low back pain ever go below the knee' (time frame since onset ≤4 weeks, median 14 days) Response format: 'yes' or 'no'
Item 3 Widespread pain	'I have had pain in the shoulder or neck at some time in the last 2 weeks.' Response format: 'agree' or 'disagree'	'Do you have pain in the ☐ neck, ☐ shoulder?' Response format: separate check boxes ☐
Item 4[a] Fear avoidance	'It's really not safe for a person with a condition like mine to be physically active.' Response format: 'agree' or 'disagree'	'In your view, do physical activities make your pain worse?'[d] Response format: Scale from 0 (*not at all*) to 10 (*absolutely*) dichotomized between scores of 6 and 7
Item 5 Pain-related disability	'In the last 2 weeks, I have dressed more slowly than usual because of my back pain.' Response format: 'agree' or 'disagree'	'In the past week because of your back pain have you get dressed more slowly than usual?' Response Format: 'yes' or 'no'
Item 6 Pain-related disability	'In the past 2 weeks, I have only walked short distances because of my back pain.' Response format: 'agree' or 'disagree'	'In the past week because of your back pain have you only walked short distances?' Response Format: 'yes' or 'no'
Item 7[a] Catastrophizing rumination	'Worrying thoughts have been going through my mind *a lot of the time* in the last weeks.' Response format: 'agree' or 'disagree'	'You couldn't seem to keep the pain out of your mind.'[e] Response format: Scale from 0 (*never*) to 10 (*always*) dichotomized between scores of 2 and 3
Item 8 Catastrophizing magnification	'I feel that my back pain is terrible and that it's never going to get any better.' Response format: 'agree' or 'disagree'	'When you feel pain, how much do you do feel: it is terrible and that it is never going to get any better?' Response format: Scale from 0 (*never*) to 10 (*always*) dichotomized between scores of 5 and 6
Item 9[a] Depressed mood	'In general in the last 2 weeks, I have **not** enjoyed all the things I used to enjoy' Response format: 'agree' or 'disagree'	'How much do you still enjoy doing things you liked before the pain started?'[f] Response format: Scale from 0 (*never*) to 10 (*always*) reversed and dichotomized between scores of 6 and 7

POP, Prognosis of Pain study sample.
[a]POP items referring to same parameter as STarT-Back in different language.
[b]Hill et al., 2008.
[c]Hill, 2011.
[d]Linton and Hallden, 1998.
[e]Sullivan et al., 1995.
[f]Neubauer et al., 2006.

likelihood ratio' (LR+) indicates how much the probability of a correct diagnosis is increased if the test is positive, while the 'negative likelihood ratio' (LR–) indicates how much that probability is decreased with a negative test result. Values above 5 or below 0.2 are generally seen as supporting a strong test, whereas values close to 1 indicate poor test performance.

3. Results

Due to the large number of patients with back pain of longer duration, only 25% of those who responded to the recruitment letter ($N = 606$) were eligible and interviewed at baseline. The follow-up samples included $N = 521$ (86%) at 6 months and $N = 443$ (73%) at 2 years. In our sample, average age was 50.5 (±12.6) years, 56% were female, 65% Caucasian-American, 61% had a college degree, 59% were employed full-time (for further details, see Mehling et al., 2012). In the follow-up, about 22% had developed chronic pain (defined as RMDQ score ≥7) at 6 months and 25% at 2 years (Mehling et al., 2012).

3.1 Risk classification at baseline and chronic pain at follow-up

Using the three risk-level criteria of the STarT-Back with the 9APSI, 32.3% of POP patients were charac-

Table 2 Proportion of STarT-Back and POP samples classified as low, medium and high risk of poor outcome according to STarT-Back[a] criteria at baseline.

Risk	STarT-Back derivation sample (N = 131) (%)	STarT-Back validation sample (N = 500) (%)	POP baseline (N = 598) (%)
Low	40	47	32.3
Medium	35	38	45.8
High	25	15	21.9

[a]Hill et al., 2008.

terized as low risk, 45.8% as medium risk, and 21.9% as high risk at baseline (Table 2). Individual item scores as well as STarT-Back total scores and psycho-social subscores were not different between responders and non-responders at follow-up (all *p*-values >0.10). Overall, 22% and 25% of patients met criteria for chronic LBP at 6 months and 2 years, respectively, using the RMDQ score criterion, the original reference for the STarT-Back study high-risk outcome (Table 3), compared with 40% in the STarT-Back validation sample. Using the von Korff criteria with the 9APSI, 19% had Grade 2 or higher chronic pain at 2 years, and using the combined perceived recovery and current pain outcome, 13% and 18% had chronic pain at 6 months and 2 years, respectively. Kappa statistics for pairwise comparisons of the three outcomes provided 81–86% agreement and kappa values varying from 0.44 for (1 with 3) to 0.53 for (2 with 3) indicating moderate agreement among the three outcomes. Compared with the STarT-Back validation sample, the proportion of chronic LBP patients in the POP study classified into the high-risk category at baseline was similarly low as that of the low-risk category, according to three outcome definitions at follow-up (Table 3). In the POP sample, multivariate regression models for the 9APSI binary variables with both 6-month and 2-year outcomes showed consistent significance for the catastrophizing/magnification item.

3.2 Screening accuracy by ROC curves

Areas under ROC curves (AUC) for the overall 9APSI score ranged from 0.54 to 0.63 for the three outcome parameters when using unaltered raw scores. When we used regression-based methods to predict back pain outcomes from the 9APSI item set (binary scores multiplied by beta coefficients), rather than the standard scoring method, AUC values rose slightly to 0.67 (highest with Roland–Morris outcome at 6 months and combined outcome or von Korff grading at 2 years) or 0.66 after bootstrapping (1000 replications), respectively.

3.3 Screening performance by likelihood ratios

Applying the 9APSI to the POP sample with exclusively *acute* LBP patients yielded low positive predictive values as well as positive and negative likelihood ratios between 0.48 and 1.63, which are much closer to 1 than the values in the STarT-Back validation sample of patients with LBP of *any* duration (Table 4). In the POP sample, the average patient with acute LBP had a pretest probability of developing chronic pain of 22% at 6 months and of 25% at 2 years (using identical RMDQ criteria; Table 3). An average patient who was classified by the 9APSI as belonging to the high-risk group increased this probability from 22% to 31% at 6 months and from 25% to 35% (using the RMDQ outcome definition; see Table 4) at 2 years.

4. Discussion

The STarT-Back clearly performed better in the UK validation sample (Hill et al., 2011) and the US sample

Table 3 Proportion of participants in each STarT-Back and Nine Analogous Predictive Screening Items (9APSI) risk category at baseline with chronic low back pain at follow-up.

	Proportion of chronic pain patients at follow-up by baseline risk category							
	STarT-Back 6-month validation sample	POP at 6 months (using 9APSI)			POP at 2 years (using 9APSI)			
Risk category at baseline	RMDQ score ≥7	N	RMDQ score ≥7	Perceived non-recovery and ≥3 of pain	N	RMDQ score ≥7	Grade ≥2 Von Korff	Perceived non-recovery and ≥3 of pain
Total % (n of N)	39.7% (196 of 494)	514	22.4% (115)	13.4% (69)	417/436	25.2% (105/417)	19.0% (83/436)	18.3% (80/436)
Low % (n)	16.7% (39 of 234)	66	15.2% (10)	12.1% (8)	59/61	15.3% (9)	11.5% (7)	8.2% (5)
Medium % (n)	53.2% (99 of 186)	336	26.8% (90)	14.6% (49)	265/279	30.9% (82)	23.7% (66)	21.5% (60)
High % (n)	78.4% (58 of 74)	112	13.4% (15)	10.7% (12)	93/96	15.1% (14)	10.4% (10)	15.6% (15)

POP, Prognosis of Pain study sample; RMDQ, Roland–Morris Disability Questionnaire.

Table 4 Sensitivity, specificity, PPV, NPV, LR+ and LR− for STarT-Back and 9APSI criteria to identify high- and low-risk patients in the STarT-Back validation and the POP samples using available definitions for a poor outcome.

	SBST[a]		9APSI-POP									
Poor outcome definition	RMDQ ≥7 at 6 months		RMDQ ≥7 at 6 months		Perceived non-recovery and ≥3 pain at 6 months		RMDQ ≥7 at 2 years		Grade 2 Von Korff at 2 years		Perceived non-recovery and ≥3 pain at 2 years	
Risk group	H versus L/M	H/M versus L	H versus L/M	H/M versus L	H versus L/M	H/M versus L	H versus L/M	H/M versus L	H versus L/M	H/M versus L	H versus L/M	H/M versus L
Sensitivity	29.6%*	80.1%	32.1%	81.7%	27.5%	76.8%	36.6%	77.1%	21.2%	80.7%	25.9%	73.8%
Specificity	94.6%	65.4%	80.1%	37.8%	82.3%	35.1%	77.6%	39.4%	81.5%	38.5%	83.5%	36.8%
PPV	78.4%	60.4%	29.6%	27.6%	25.9%	16.4%	28.6%	30.0%	21.7%	23.6%	28.6%	20.7%
NPV	67.1%	83.3%	82.0%	87.8%	83.5%	90.7	83.3%	83.7%	81.0%	89.5%	82.3%	86.2%
LR+	5.48	2.32	1.61	1.31	1.55	1.18	1.63	1.27	1.15	1.31	1.57	1.17
LH−	0.74	0.30	0.85	0.48	0.88	0.66	0.82	0.58	0.97	0.50	0.89	0.71

9APSI, Nine Analogous Predictive Screening Items; H, high risk; L, low risk according to STarT-Back criteria; LH−, negative likelihood ratio; LR+, positive likelihood ratio; M, medium risk; NPV, negative predictive value; POP, Prognosis of Pain study sample; PPV, positive predictive value; RMDQ, Roland–Morris Disability Questionnaire.

*The value of 39.6% in Table 4 by Hill et al. (2008) may be a typo, as we recalculated all values using the numbers provided in the text on page 638: 58 of 196 are 29.6% participants with RMDQ score ≥7.

[a]Six individuals were not accounted for reducing the sample size with complete data to 494.

(Von Korff et al., 2014) of patients with LBP of *any* duration than the 9APSI in the POP sample with strictly defined *acute* LBP. In particular, in the POP sample, the specificity of the high-risk category for predicting chronic LBP with the 9APSI was substantially lower than in the UK validation sample with the STarT-Back tool. This resulted in a low positive predictive value of developing chronic LBP using the 9APSI high-risk designation, and an overall weak test performance indicated by the likelihood ratios closer to 1.

The STarT-Back items address parameters that were similar to those which strongest predicted LBP at follow-up in the POP study, such as pain intensity, sciatica, widespread pain, functional disability, catastrophizing and ability to enjoy life (Mehling et al., unpublished data). Additional parameters for the POP sample were (1) confidence in self-management of stress, (2) ignoring pain, and (3) level of education, but not (4) fear avoidance beliefs. To our knowledge, parameters 1 and 2 have not been studied before in this context, whereas a low education level is known to be predictive for chronic pain (Dionne et al., 1995) and fear avoidance was found to be mildly predictive previously (Grotle et al., 2006).

Several issues may explain the apparent differences between the performances of the instrument in the two samples. First, the 9APSI is not identical to the STarT-Back. Second, the STarT-Back instrument was not developed for, nor validated in, patients with *acute* LBP defined as being of less than 30-day duration, for which the 9APSI used in the current analysis does not show a strong predictive performance. Patients

enrolled in the STarT-Back samples were seen for a first doctor visit for LBP of *any* duration, which also was similar to the population in studies that used the von Korff outcome (Dunn et al., 2008; Turner et al., 2013) which we applied here as well. In contrast, the POP study enrolled only patients with carefully defined acute LBP. The STarT-Back may better classify the typical patients seen for a first doctor's visit for LBP of *any* duration in primary care, while the POP study is representative of patients seen in primary care for acute LBP. We exclusively studied patients with strictly defined acute LBP when clinicians need to prescribe early targeted interventions to prevent the development of chronic pain. Third, using the median split of the disability score derived from a majority of chronic pain patients at follow-up (RMDQ ≥ 7/24) as reference for high-risk patients necessarily leads to higher cut-offs for the STarT-Back or 9APSI than would be expected in a sample of *acute* patients only, in which approximately 80% will have recovered by follow-up. Only about a quarter of the POP study participants had RMDQ scores of 7 or higher at 6 months and 2 years, when the median RMDQ scores were 1 (mean 3.6 ± 4.8) and 2 (mean 4.4 ± 5.4), respectively. This suggests that, for classifying patients with *acute* LBP into risk groups, the STarT-Back may at least require a modification of cut-points.

Other attempts at developing screening tools for patients with a LBP have had mixed results. A large study conducted in Spain did not include psychological risk factors, assessed patients with median pain duration of 180 days, and its predictors discriminated

poorly with AUC of 0.655 for pain and 0.640 for disability at 3 months after inception (Kovacs et al., 2012). A study with 123 Norwegian patients with a LBP of <3 weeks used the Acute Low Back Pain Screening Questionnaire (Linton and Hallden, 1998), defined non-recovery as RMDQ score of >4, which classified 17% as cLBP, and found AUCs of 0.68 at 6 months and 0.72 at 1 year. Studies that compared the predictors for long-term disability between patients with aLBP versus cLBP found a high degree of similarity when aLBP included subacute pain of up to 3 months duration (Grotle et al., 2010); however, they appeared to differ when aLBP included patients with pain up to 2 or 3 weeks (Kovacs et al., 2005; Grotle et al., 2006). This indicates that variations in time frames used for defining aLBP may be responsible for different study results, and supports the sampling of pain patients with duration of 4 weeks or less for prognostic studies.

In conclusion, our data suggest that the STarT-Back screening tool should be used cautiously when using the current cut-offs in patients presenting with *acute* (less than 4 weeks) rather than chronic LBP. Rigorous testing of the original STarT-Back screening tool in future studies may need to consider modification of cut-offs when studying patients with *acute* LBP, in which an early recognition of increased risk for chronic pain is of particular interest for clinicians to allow for prescribing targeted interventions for high-risk patients. Our data emphasize the importance of clearly distinguishing samples of patients with LBP of *any* duration from those with acute LBP in future research.

Author contributions

All authors discussed the results and commented on the manuscript. W.M.: Design, acquisition, analysis and interpretation of data. A.A.: Design, acquisition and interpretation of data. M.A., T.C. and R.H.: Design, analysis and interpretation of data. All authors: Drafting the article and revising it critically for important intellectual content.

References

Beurskens, A.J., de Vet, H.C., Koke, A.J. (1996). Responsiveness of functional status in low back pain: A comparison of different instruments. *Pain* 65, 71–76.

Chou, R. (2014). In the clinic. Low back pain. *Ann Intern Med* 160, ITC6-1.

de Vet, H.C., Terluin, B., Knol, D.L., Roorda, L.D., Mokkink, L.B., Ostelo, R.W., Hendriks, E.J., Bouter, L.M., Terwee, C.B. (2010). Three ways to quantify uncertainty in individually applied 'minimally important change' values. *J Clin Epidemiol* 63, 37–45.

Deyo, R.A., Dworkin, S.F., Amtmann, D., Andersson, G., Borenstein, D., Carragee, E., Carrino, J., Chou, R., Cook, K., DeLitto, A., Goertz, C., Khalsa, P., Loeser, J., Mackey, S., Panagis, J., Rainville, J., Tosteson, T., Turk, D., Von Korff, M., Weiner, D.K. (2014). Report of the NIH task

force on research standards for chronic low back pain. *J Pain* 39, 1128–1143.

Dionne, C., Koepsell, T.D., Von Korff, M., Deyo, R.A., Barlow, W.I., Checkoway, H. (1995). Formal education and back-related disability. In search of an explanation. *Spine* 20, 2721–2730.

Dunn, K.M., Croft, P.R., Main, C.J., Von Korff, M. (2008). A prognostic approach to defining chronic pain: Replication in a UK primary care low back pain population. *Pain* 135, 48–54.

Griffith, L.E., Hogg-Johnson, S., Cole, D.C., Krause, N., Hayden, J., Burdorf, A., Leclerc, A., Coggon, D., Bongers, P., Walter, S.D., Shannon, H.S., Meta-Analysis of Pain in the Lower Back., Work Exposures Collaborative Group (2007). Low-back pain definitions in occupational studies were categorized for a meta-analysis using Delphi consensus methods. *J Clin Epidemiol* 60, 625–633.

Grotle, M., Foster, N.E., Dunn, K.M., Croft, P. (2010). Are prognostic indicators for poor outcome different for acute and chronic low back pain consulters in primary care? *Pain* 151, 790–797.

Grotle, M., Vollestad, N.K., Brox, J.I. (2006). Clinical course and impact of fear-avoidance beliefs in low back pain: Prospective cohort study of acute and chronic low back pain: II. *Spine (Phila Pa 1976)* 31, 1038–1046.

Hay, E.M., Dunn, K.M., Hill, J.C., Lewis, M., Mason, E.E., Konstantinou, K., Sowden, G., Somerville, S., Vohora, K., Whitehurst, D., Main, C.J. (2008). A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial Study Protocol. *BMC Musculoskelet Disord* 9, 58.

Hill, J.C., Dunn, K.M., Lewis, M., Mullis, R., Main, C.J., Foster, N.E., Hay, E.M. (2008). A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis Rheum* 59, 632–641.

Hill, J.C., Dunn, K.M., Main, C.J., Hay, E.M. (2010a). Subgrouping low back pain: A comparison of the STarT Back Tool with the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 14, 83–89.

Hill, J.C., Vohora, K., Dunn, K.M., Main, C.J., Hay, E.M. (2010b). Comparing the STarT back screening tool's subgroup allocation of individual patients with that of independent clinical experts. *Clin J Pain* 26, 783–787.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., Konstantinou, K., Main, C.J., Mason, E., Somerville, S., Sowden, G., Vohora, K., Hay, E.M. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. *Lancet* 378, 1560–1571.

Itz, C.J., Geurts, J.W., van Kleef, M., Nelemans, P. (2013). Clinical course of non-specific low back pain: A systematic review of prospective cohort studies set in primary care. *Eur J Pain* 17, 5–15.

Klenerman, L., Slade, P.D., Stanley, I.M., Pennie, B., Reilly, J.P., Atchison, L.E., Troup, J.D., Rose, M.J. (1995). The prediction of chronicity in patients with an acute attack of low back pain in a general practice setting. *Spine* 20, 478–484.

Kovacs, F.M., Abraira, V., Zamora, J., Fernandez, C. (2005). The transition from acute to subacute and chronic low back pain: A study based on determinants of quality of life and prediction of chronic disability. *Spine* 30, 1786–1792.

Kovacs, F.M., Seco, J., Royuela, A., Corcoll Reixach, J., Abraira, V., Spanish Back Pain Research, N. (2012). Predicting the evolution of low back pain patients in routine clinical practice: Results from a registry within the Spanish National Health Service. *Spine J* 12, 1008–1020.

Krebs, E.E., Carey, T.S., Weinberger, M. (2007). Accuracy of the pain numeric rating scale as a screening test in primary care. *J Gen Intern Med* 22, 1453–1458.

Linton, S.J., Hallden, K. (1998). Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain* 14, 209–215.

Mehling, W.E., Gopisetty, V., Acree, M., Pressman, A., Carey, T., Goldberg, H., Hecht, F.M., Avins, A.L. (2011). Acute low back pain and primary care: How to define recovery and chronification? *Spine (Phila Pa 1976)* 36, 2316–2323.

Mehling, W.E., Gopisetty, V., Bartmess, E., Acree, M., Pressman, A., Goldberg, H., Hecht, F.M., Carey, T., Avins, A.L. (2012). The prognosis of acute low back pain in primary care in the United States: A 2-year

prospective cohort study. *Spine (Phila Pa 1976)* 37, 678–684.Neubauer, E., Junge, A., Pirron, P., Seemann, H., Schiltenwolf, M. (2006). HKF-R 10 – Screening for predicting chronicity in acute low back pain (LBP): A prospective clinical trial. *Eur J Pain* 10(6), 559–566.

Pengel, L.H., Herbert, R.D., Maher, C.G., Refshauge, K.M. (2003). Acute low back pain: Systematic review of its prognosis. *BMJ* 327, 323.

Roland, M., Morris, R. (1983). A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8, 141–144.

Stata12 (2013). *Statistics/Data Analysis* (College Station, TX: Stata Corporation).

Sullivan, M.J.L., Bishop, S.R., Pivik, J. (1995). The Pain Catastrophizing Scale: Development and validation. *Psychol Assess* 7(4), 524–532.

Turner, J.A., Shortreed, S.M., Saunders, K.W., Leresche, L., Berlin, J.A., Korff, M.V. (2013). Optimizing prediction of back pain outcomes. *Pain* 154, 1391–1401.

van Tulder, M.W., Koes, B.W., Bouter, L.M. (1997). Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 22, 2128–2156.

Von Korff, M., Ormel, J., Keefe, F.J., Dworkin, S.F. (1992). Grading the severity of chronic pain. *Pain* 50, 133–149.

Von Korff, M., Shortreed, S.M., Saunders, K.W., Leresche, L., Berlin, J.A., Stang, P., Turner, J.A. (2014). Comparison of back pain prognostic risk stratification item sets. *J Pain* 15, 81–89.

Web references

Hill, J. (2011) Keele STarT Back Screening Tool website. Retrieved from http://www.keele.ac.uk/sbst/ (accessed October 2014).

SurveyGizmo (2010) Retrieved from www.surveygizmo.com (accessed October 2009); 4888 Pearl East Cir. Suite 399W, Boulder, CO 80301, USA.