Subgrouping low back pain: A comparison of the STarT Back Tool with the Örebro Musculoskeletal Pain Screening Questionnaire

Jonathan C. Hill *, Kate M. Dunn, Chris J. Main, Elaine M. Hay

Arthritis Research Campaign National Primary Care Centre, Primary Care Sciences, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom

A R T I C L E   I N F O

Article history:
Received 3 July 2008
Received in revised form 11 December 2008
Accepted 13 January 2009
Available online 18 February 2009

Keywords:
Low back pain
Classification
Primary care
Early identification
Psychological factors

A B S T R A C T

Introduction: Clinicians require brief, practical tools to help identify low back pain (LBP) subgroups requiring early, targeted secondary prevention. The STarT Back Tool (SBT) was recently validated to subgroup LBP patients into early treatment pathways.

Aim: To test the SBT′s concurrent validity against an existing, popular LBP subgrouping tool, the Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ), and to compare the clinical characteristics of subgroups identified by each tool.

Methods: Two hundred and forty-four consecutive ‘non-specific’ LBP consulters at 8 UK GP practices aged 18–59 years were invited to complete a questionnaire. Measures included the ÖMPSQ and SBT, disability, fear, catastrophising, pain intensity, episode duration and demographics. Instruments were compared using Spearman′s correlations, tests for subgroup agreement and discriminant analysis of subgroup characteristics according to reference standards.

Results: Completed SBT (9-items) and ÖMPSQ (24-items) data was available for 130/244 patients (53%). The correlation of SBT and ÖMPSQ scores was ‘excellent (rs = 0.80). Subgroup characteristics were similar across the low, medium and high subgroups, but, the proportions allocated to ‘low’, ‘medium’ and ‘high’ risk groups were different, with fewer patients in the SBT′s high risk group. Both instruments similarly discriminated for reference standards such as disability, catastrophising, fear, comorbid pain and time off work. The ÖMPSQ was better at discriminating pain intensity, while the SBT was better for discriminating bothersomeness of back pain and referred leg pain.

Conclusions: The SBT baseline psychometrics performed similarly to the ÖMPSQ, but the SBT is shorter and easier to score and is an appropriate alternative for identifying high risk LBP patients in primary care.

© 2009 European Federation of International Association for the Study of Pain Chapters. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Primary care evidence-based guidelines for non-specific back pain highlight the importance of identifying indicators of poor prognosis in order that treatment can be targeted appropriately (Chou et al., 2007; van Tulder et al., 2006). Investigators increasingly advocate that better identification of potentially modifiable prognostic indicators may lead to more effective early secondary prevention of back pain in primary care (Boersma and Linton, 2005; Morley and Vlaeyen, 2005; Jellema et al., 2006; Koes et al., 2006). Several back pain clinical tools exist to aid clinicians in identifying patients either ‘at risk’ of chronicity or to improve targeting of treatment (Childs et al., 2004; Hicks et al., 2005; Dionne et al., 2005; Truchon and Cote, 2005; Duitjts et al., 2006; Neubauer et al., 2006; Denison et al., 2007).

The STarT Back Tool (SBT) is a recently validated tool developed to identify subgroups of patients to guide the provision of early secondary prevention in primary care (Hill et al., 2008). The conceptual purposes of the SBT (Hill et al., 2008) were to identify patients with potentially treatable modifiable prognostic indicators using a brief, user-friendly tool and to validate cut-off scores for subgrouping patients into 1 of 3 a priori initial treatment options in primary care (low, medium and high risk groups). The development and validation of the STarT Back Tool (SBT) was recently reported (Hill et al., 2008), but a head to head comparison with a ‘reference standard’ instrument was not evaluated. The Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ: Linton and Hallden, 1998) was considered the most appropriate ‘reference standard’ against which to compare the SBT, as it is one of the most widely used tools in clinical practice to similarly differentiate primary care back pain patients, and it has a common conceptual purpose with the SBT of identifying high risk patients requiring targeted treatment. As well as being a popular instrument, the ÖMPSQ has been externally validated in numerous international
Clinicians wanting to follow evidence-based practice guidelines to assess prognostic indicators for chronicity (Chou et al., 2007), require brief and practical tools to help them identify subgroups of patients that may require early, targeted secondary prevention pathways. The publication of the SBT (Hill et al., 2008) provides clinicians with a choice, in addition to an existing ‘reference standard’ instrument, the ÖMPSQ. However, in order to decide which clinical tool to use in practice, clinicians need to know how these instruments compare, including the baseline clinical characteristics of the subgroups identified by each tool.

The overall aim of this study was to therefore provide a head to head comparison of the SBT’s concurrent validity against a best available reference standard, the ÖMPSQ. Our objectives were to test the correlation between the SBT and ÖMPSQ scores and agreement of patients allocated to ‘low’, ‘medium’ and ‘high’ risk subgroups; and to compare the abilities of the SBT and ÖMPSQ scales to discriminate patients according to validated reference standard measures.

2. Methods

2.1. Properties of the ÖMPSQ and SBT

The ÖMPSQ consists of 24 self-report items (21 items are scored), selected following a literature review to identify strong independent risk factors for work absence. The authors defined ‘poor prognosis’ as accumulated sick leave of 30 days or more at six months follow up. The ÖMPSQ’s 21 scored items use an 11-point response format, apart from item 1 (pain sites), which has five descriptive components that are double weighted. The instrument therefore provides a potential score ranging from two to 210 points. The reliability of the ÖMPSQ has been reported with a Kappa of 0.83 (Linton and Hallden, 1998) and external validity, with a number of different high risk cut-off scores, has been established in a variety of patient populations and settings (Heneweer et al., 2007; Westman et al., 2008; Hough et al., 2007; Margison and French, 2007; Nordeman et al., 2006; Jellema et al., 2007; Gabel et al., 2008; Hurley et al., 2001; Grotle et al., 2006; Linton and Boersma, 2003; Boersma and Linton, 2005; Dunstan et al., 2005).

The predictive validity of the ÖMPSQ has also been investigated in a number of studies, as summarised by Hocking et al. (2008). Since the initial development of the ÖMPSQ, other authors have used it to classify patients into low, medium and high ‘at risk’ groups (Nordeman et al., 2006). The SBT has 9-items selected as predictive of ‘poor prognosis’ following a literature review and secondary analysis to identify strong independent predictors for persistent disabling back pain. The 9-items each use a dichotomised response format (‘agree’ or ‘disagree’), apart from one bothersomeness item, which uses a Likert scale. Overall SBT scores range from 0-9 and are produced by summing all positive items; a psychosocial subscale score ranging from 0 to 5 is produced by summing bothersomeness, referred leg pain, and comorbid pain, while the ÖMPSQ discriminates on the basis of additional items for pain intensity, coping and work-related factors.

2.2. Data collection

Potential differences between the ÖMPSQ and SBT were explored using data from a cross-sectional survey of consecutive adult patients (n = 131) who consulted their GP with low back pain. The methods of patient recruitment have already been published elsewhere (Hill et al., 2008). In brief, 244 participants were invited to complete a questionnaire containing both the ÖMPSQ and the SBT from 8 general practices in North Staffordshire and Central Cheshire, UK using computerised Read Codes to identify recent back pain consulters. This study received ethical approval from the North Staffordshire Local Research Ethics Committee.

2.3. Analysis of data

The SBT and ÖMPSQ were scored according to the methods specified by the instrument developers (Hill et al., 2008; Linton and Hallden, 1998). Patients were classified into ‘low’, ‘medium’ and ‘high’ risk groups using derived cut-off scores for each instrument. For the SBT, cut-offs recommended in the original article were used (Hill et al., 2008); for the ÖMPSQ, we replicated the high risk cut-off score determined within a UK primary care population by Hurley et al. (2001) of >112; and a cut-off of 90 to separate low, from medium risk group patients, as used by Nordeman et al. (2006). To score the ÖMPSQ among patients who were not workers, (four work-related items are only relevant to those patients in work), the procedure used by Jellema et al. (2007) was followed, where the mean score of the remaining 17 items was imputed, on the condition that at least 75% of all items were completed. Correlations between the SBT and ÖMPSQ, and reference standard constructs included in both instruments: disability (Roland and Morris Disability Questionnaire, RMDQ: Roland and Morris, 1983) fear avoidance (Tampa Scale of Kinesiophobia, TSK: Kori et al., 1990) and catastrophising (Pain Catastrophising Scale, PCS: Sullivan et al., 1995) were calculated using Spearman’s rank correlations. The magnitude of the reported correlation coefficient was evaluated with a correlation of 0.1–0.3 was small, >0.3–0.5 was moderate, and greater than 0.5 was large (Cohen, 1998). In addition, box and whisker plot graphs were produced to visually present the correlation between the SBT and ÖMPSQ.

Observed agreement of the two tool’s allocation to ‘low’, ‘medium’ and ‘high’ risk subgroups was examined by calculating absolute agreement, and agreement beyond chance statistically evaluated using a weighted Cohen’s kappa test. Kappa values were classified for reference as follows: less than 0.00 showed poor agreement, 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and greater than 0.80, near perfect agreement (Landis and Koch, 1977). The McNemar Bowker test was used to determine whether disagreement observed was evenly balanced or significantly skewed towards the lower or higher group. The characteristics of patients selected for ‘low’, ‘medium’ and ‘high’ subgroups using the SBT and ÖMPSQ were compared using median scores for age, gender, episode duration, bothersomeness, pain intensity, days off work, fear (TSK), catastrophising (PCS), ÖMPSQ scores and SBT scores (mean values were

1 In box and whisker plots the central line is the median, the box provides the lower and upper quartiles, the whiskers are the 2.5% and 97.5% values, and extreme values are noted with a ‘*’ (Chambers et al., 1983).
used for age). The Mann Whitney test was used to test for statistical differences between continuous data and the Chi-squared test used for categorical data among the characteristics for discordant individuals (patients for whom group allocation differed).

Finally, the ability of the two instruments’ total (and subscale) scores to discriminate patients was compared by plotting receiver operator characteristic (ROC) curves for various different reference standard ‘cases’ (using the ‘case’ definitions provided in Table 3) and calculating AUCs (95% CIs), with statistical differences tested using the Wilcoxon statistic (Hanley and McNeil, 1982). Total scores were entered as the ‘test’ variables and reference standard ‘cases’ as the ‘state’ variable.

3. Results

Complete data on the SBT and ÖMPSQ was available for 130/244 (53%) patients. The sample was 60% female, with a mean (±SD) age of 44 (±10.0) and disability score (RMDQ) of 8.6 (±6.6). Back pain

![Box plot graph of the ÖMPSQ total scores against (a) the SBT total score and (b) the SBT psychosocial subscale.](image)

**Fig. 1.** Box plot graph of the ÖMPSQ total scores against (a) the SBT total score and (b) the SBT psychosocial subscale.
episode durations were varied with 25% < 1 month, 27% 1–6 months, and 48% >6 months, and 43% reported being ‘very’ or ‘extremely’ bothered by their back pain. The median scores (range) for SBT and ÖMPSQ were 4.5 (0–9) and 98.8 (26.3–189.0), respectively. The Spearman’s rank correlation coefficients for the SBT total scores and psychosocial subscale scores with the ÖMPSQ scores were 0.802 and 0.769 respectively. The magnitude of the reported correlation coefficients was therefore, ‘large’ and was similar to the correlation between disability (RMDQ) and the subgroup tools (SBT = 0.813; ÖMPSQ = 0.830). Figures for fear (TSK) (SBT psychosocial subscale = 0.659; ÖMPSQ = 0.683) and catastrophising (PCS) (SBT = 0.671; ÖMPSQ = 0.656) were lower but also classified as ‘large’. Fig. 1 presents box plot graphs of the ÖMPSQ total score against the SBT (a) total score and (b) psychosocial subscale and demonstrate that increasing ÖMPSQ scores, correlated with higher SBT scores across the full range of both instrument scales.

The proportions of patients allocated to ‘low’, ‘medium’ and ‘high’ risk groups by the SBT and ÖMPSQ were 40% of 40%, 35% of 22%, and 25% of 38%, respectively. The ‘low’ risk proportion was therefore the same, but the SBT allocated considerably fewer patients to the ‘high’ risk group, but significantly more patients to the ‘medium’ risk group than the ÖMPSQ (McNemar Bowker test \( p = 0.022 \)).

Observed agreement between the two instruments for allocation to ‘low’, ‘medium’ and ‘high’ risk groups, is presented in Table 1, with absolute agreement in 62% of patients. Weighted Cohen’s kappa for agreement in allocation to the three subgroups beyond chance was ‘moderate’, 0.57 (95% CI 0.47–0.68, \( p = 0.000 \)). Agreement about allocation to the ‘low’ risk group alone (‘low’ versus ‘medium’ or ‘high’) was ‘substantial’ at 0.63 (95% CIs 0.50–0.77, \( p = 0.000 \)).

### Table 1

| ÖMPSQ subgroup | SBT subgroup (n = 130) |
|-----------------|------------------------|
|                 | Low        | Medium | High     |
| Low             | 41         | 11     | 0        |
| Medium          | 10         | 13     | 6        |
| High            | 2          | 20     | 27       |

The clinical characteristics of the subgroups (‘low’, ‘medium’ and ‘high’) derived using the SBT and ÖMPSQ are presented in Table 2. The characteristics of the ‘low’ risk groups were almost identical (e.g. both had an RMDQ median of 2 and a pain NRS of 1.7). The ‘medium’ risk groups differed only slightly with higher pain and disability scores among those allocated by the SBT and the ‘high’ risk groups were also very similar although the ÖMPSQ ‘high’ risk group had slightly more females and had longer episode durations. However, these differences between the clinical characteristics in the ‘medium’ and ‘high’ risk groups produced by both tools were not statistically significant (\( p > 0.05 \)).

The ability of the two instruments’ total scores (and SBT subscale scores) to discriminate patients defined as ‘cases’ on reference standards is given in Fig. 2 and Table 3. There were no significant differences in the discriminative abilities of the SBT and ÖMPSQ scales for ‘cases’ of disability, catastrophising, fear, comorbid pain, time off work or episode duration reference standards. The SBT was significantly better for discriminating bothersome and referred leg pain ‘cases’ while the ÖMPSQ scores was better at discriminating ‘cases’ of pain intensity.

### 4. Discussion

This study has provided a comparative analysis of the measurement properties of the SBT and the ÖMPSQ for patients with low back pain, consulting in primary care. The ÖMPSQ (21 items) has been available for nearly a decade and is a popular clinical tool with numerous external validation studies in comparison to the STarT Back Tool (9-items), which has been more recently validated and has potential advantages in terms of its length and scoring simplicity. Both share broadly similar purposes, although subtle conceptual differences have become apparent through conducting this investigation, which should be highlighted for clinicians seeking to choose which tool is most appropriate for their context. Firstly, the ÖMPSQ was conceived as a prognostic screening tool, although in practice it has adopted a subgrouping function with established cut-offs across a variety of settings to allocate patients to low and high risk groups. In contrast, the SBT was primarily conceived as a subgrouping tool, with cut-offs based on baseline factors and not on future outcome, and for specifically defined low, medium and high risk treatment subgroups.

---

**Table 2** Medians [inter-quartile ranges] of clinical characteristics by low, medium and high subgroups derived using SBT and ÖMPSQ cut-offs.

| Reference standard clinical characteristics | Subgroup (n = 130) |
|--------------------------------------------|------------------|
|                                            | Low  | Medium | High   |
|                                            | SBT  | ÖMPSQ  | SBT    | ÖMPSQ  | SBT  | ÖMPSQ  |
| Age, median [IQR]                          | 43   | [33–52] | 44   | [34–56] | 45   | [39–52] | 48   | [39–56] | 49   | [39–54] | 50   | [37–54] |
| Female, number (%)                         | 60   | 58     | 57   | 52     | 54   | 63     |
| Episode duration category, median [IQR]    | 2    | [1–4]  | 2    | [1–4]  | 3    | [2–5]  | 3    | [2–4]  | 4    | [4–5]  | 5    | [3–5]  |
| SBT                                         | 2    | [2–3]  | 2    | [2–3]  | 4    | [3–4]  | 4    | [3–4]  | 4    | [4–5]  | 4    | [3–5]  |
| Pain intensity (NRS), median [IQR]         | 1.7  | [0.7–3.3] | 1.7  | [0.7–3.0] | 4.3  | [3.0–5.7] | 3.3  | [3.0–5.3] | 6.7  | [4.8–8.2] | 6.3  | [4.9–7.7] |
| Time off work in past 6/12 category, median [IQR] | 2    | [1–2]  | 2    | [1–2]  | 3    | [1–4]  | 3    | [1–4]  | 4    | [4–5]  | 4    | [3–4]  |
| Fear (TSK), median [IQR]                   | 38   | [34–42] | 37   | [33–39] | 41   | [37–44] | 41   | [40–44] | 47   | [43–52] | 45   | [42–51] |
| Catastrophising (PCS), median [IQR]        | 9    | [5–17]  | 9    | [5–13]  | 19   | [11–23] | 19   | [14–27] | 31   | [21–38] | 28   | [20–38] |
| Disability (RMDQ), median [IQR]            | 2    | [0–5]  | 2    | [0–5]  | 11   | [6–13]  | 11   | [7–13]  | 16   | [13–19] | 14   | [11–17] |
| ÖMPSQ total, median [IQR]                  | 69   | [47–87] | 65   | [47–79] | 108  | [90–127] | 101  | [94–105] | 142  | [124–175] | 141  | [126–158] |
| SBT total, median [IQR]                    | 2    | [0–3]  | 2    | [0–3]  | 5    | [4–6]  | 5    | [3–6]  | 8    | [7–8]  | 6    | [6–8]  |
| SBT Psych subscale, median [IQR]           | 0    | [0–1]  | 2    | [2–3]  | 4    | [4–5]  | 4    | [3–4]  |      |        |      |        |

---

TSK, Tampa Scale of Kinesiophobia; RMDQ, Roland-Morris Disability Questionnaire; NRS, Numerical Rating Scale; PCS, Pain Catastrophising Scale; ÖMPSQ, Orebro Musculoskeletal Pain Screening Questionnaire; SBT, STarT Back Tool.

Episode duration categories: 1 = <1 month; 2 = 1–3-months; 3 = 4–6-months; 4 = 7 months to 3 years; 5 = >3 years.

Bothersomeness categories: 1 = not at all; 2 = slightly; 3 = moderately; 4 = very much; 5 = extremely.

Time off work categories: 1 = no time off, 2 = <1 week, 3 = 1–4-weeks, 4 = 1–3-months, 5 = >3-months.

None of the above clinical characteristic differences between groups were statistically significant using the Mann Whitney test for continuous data and the Chi-squared test for categorical data among discordant individuals (patients for whom group allocation differed).
4.1. Differences in subgroup allocation

This study found that the completion rate for both instruments was high \((n = 130/131\) for both) and that their three subgroups’ clinical characteristics were broadly similar. Agreement over the two instrument’s allocation to a ‘low’ risk subgroup was ‘substantial’, suggesting that both tools have a similar ability to differentiate patients based on their baseline characteristics. Although the characteristics of patients in the three subgroups allocated by each tool were not significantly different, significant differences were observed in the proportions of patients allocated to the ‘high’ and ‘medium’ risk groups, with a smaller proportion of patients allocated to the high risk group by the SBT than the OMPSQ. One factor in the choice of instrument used by clinicians may therefore be the availability of services for the management of a high risk group of patients.

An analysis of both instrument scales’ abilities to discriminate a number of relevant reference standards was investigated to determine whether similarities in the subgroup clinical characteristics were due to tool cut-off points used, or due to similar discriminative abilities of the instrument scales. The results demonstrated that the two instrument scales similarly discriminated

\[\text{Fig. 2. Receiver operating characteristic (ROC) curves of OMPSQ and SBT scores against reference standards for (a) disability and pain, (b) catastrophising and fear, (c) referred leg pain and bothersomeness, and (d) time off work and episode duration.}\]
Table 3
AUCs (95% CIs) of SBT (total and subscale scores) and ÖMPSQ scores for discriminating ‘cases’ on reference standards.

| Clinical characteristics defined using ‘cases’ on reference standards | AUCs (95% CIs) | p Value |
|---------------------------------------------------------------|----------------|---------|
| Bothered by back pain ≥ ‘very’ | .924 (.880, .968) |        |
| Disability RMDQ > 7 | .937 (.896, .978) |        |
| Catastrophising PCS ≥ 20 | .787 (.708, .867) |        |
| Fear TSK > 41 | .787 (.708, .866) |        |
| Pain intensity NRS ≥ 3.7 | .825 (.754, .896) |        |
| Referred leg pain ‘yes’ | .841 (.774, .908) |        |
| Comorbid pain ≥ 3 sites | .681 (.586, .780) |        |
| Time off work ≥ 30 days | .885 (.818, .952) |        |
| Episode duration ≥ 3-months | .706 (.613, .799) |        |

RMDQ, Roland and Morris Disability Questionnaire; PCS, Pain Catastrophising Scale; TSK, Tampa Scale of Kinesiophobia; NRS, Numerical Response Scale; SBT, STarT Back Tool; ÖMPSQ, Örebro Musculoskeletal Pain Screening Questionnaire.

To test for statistical differences between AUCs, the DeLong et al. (1988) algorithm for comparing AUCs was used.

the majority of clinical characteristics including disability, catastrophising, fear, comorbid pain, time off work or episode duration. The ÖMPSQ total scale was significantly better at discriminating patients’ baseline pain intensity, while the SBT was better for discriminating baseline bothersomeness of back pain and referred leg pain.

4.2. Limitations

One substantial limitation of this study is that only cross-sectional data was available for directly comparing the performance of both instruments, preventing a comparison of the instrument’s predictive validity. However, guidelines for the management of non-specific back pain highlight the importance of identifying indicators of poor prognosis, rather than predicting individual patient’s outcomes, in order that treatment can be targeted appropriately (Chou et al., 2007; van Tulder et al., 2006). Traditionally, validation of the utility of such instrument scales has predominantly focused on establishing an instrument’s predictive validity, although the SBT subgroup cut-offs were specifically validated to optimally differentiate patients with specific baseline clinical characteristics. To date, there have been no studies that have compared the predictive abilities of both instruments within a single sample, but published data on the predictive validity of the ÖMPSQ (Hockings et al., 2008) reports AUCs for persisting disability from 0.68 to 0.83, which compares similarly to reported AUCs of 0.80 for the SBT (Hill PhD Thesis, 2008). A further limitation is the potential for non-response bias influencing generalisability as only 53% of the sample responded, which may have inflated the proportion classified by the two instruments to the medium and high risk subgroups, as people with longer term or more severe problems are more likely to respond to the questionnaire. However, non-response bias is likely to have equally affected both instruments and is unlikely to have influenced the tool comparisons made in this study.

5. Conclusion

This study has compared measurement properties of the SBT to those of the ÖMPSQ, and found them to be similar in respect to the subgroup patient characteristics and abilities of the scales to differentiate according to validated reference standard measures. The SBT is quicker for patients to complete and easier for clinicians to score, although it is able to identify a high risk subgroup with similar clinical characteristics to the ÖMPSQ. Apart from the differences in instrument length, the results of this study suggest that the main difference between the two instruments was the proportion of patients each tool allocated to the high risk group, with the SBT identifying 25% as high risk, and the ÖMPSQ 38% as high risk. This might be of relevance to clinicians working within services that have resource limitations for treating high risk patients. Recommendations for future research include; a direct comparison of the predictive validity of both instruments, and also an investigation of possible instrument item redundancy. Clinicians need to be aware of the relative strengths and weaknesses of the two clinical tools before selecting which tool is most appropriate to help decision making in their specific clinical context and setting. This head to head comparison has demonstrated that the SBT’s discriminative abilities appear to be at least equivalent, and preferable in some criteria, to the ÖMPSQ, which is the most widely used current tool in clinical practice.

Acknowledgments

This study was supported by a Programme Grant from the Arthritis Research Campaign (arc), UK (Grant Code: 13413) and by the North Staffordshire Primary Care Research Consortium. The authors would like to thank the administrative and health informatics staff at Keele University’s Primary Care Musculoskeletal Research Centre and the Keele General Practice Partnership. Thanks also to the members of the independent monitoring committee, staff and patients of the eight participating general practices and to the wider members of the research team who were involved with the study. Dr Jonathan Hill is funded as an arc Lecturer in Physiotherapy awarded by Arthritis Research Campaign (arc) UK.

References

Boersma K, Linton SJ. Screening to identify patients at risk: profiles of psychological risk factors for early intervention. Clin J Pain Special Top Ser: Cogn-Behav Treat Chronic Pain 2005;21:38–43.
Childs JD, Fritz JM, Flynn TW, Irgang JJ, Johnson KK, Majkowski GR, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Ann Intern Med 2004;141:920–8.
Chou R, Qaseem A, Snow V, Casey D, Cross Jr JT, Shekelle P, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007;147:478–91.
Cohen J. Statistical power analysis for the behavioural sciences. Hillsdale, NJ: Erlbaum Associates; 1998.
DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating curves: a nonparametric approach. Biometrics 1988;44:837–45.
Denison E, Asenföld P, Sandborgh M, Lindberg P. Musculoskeletal pain in primary health care: subgroups based on pain intensity, disability, self-efficacy, and fear-avoidance variables. J Pain 2007;8:67–74.
Dionne CE, Bourbonnais R, Fremont P, Rossignol M, Stock SR, Larocque I. A clinical return-to-work rule for patients with back pain. CMAJ 2005;172:1559–67.
Duijts SF, Kant JJ, Landeweerd JA, Swaen GM. Prediction of sickness absence: development of a screening instrument. Occup Environ Med 2006;63:564–9.
Dunstan DA, Covic T, Tyson GA, Lennie IG. Does the Örebro Musculoskeletal Pain Questionnaire predict outcomes following a work-related compensable injury? Int J Rehabil Res 2005;28:369–70.
Gabel CP, Burkett B, Neller A, Yelland M. Can long-term impairment in general practitioner whiplash patients be predicted using screening and patient-reported outcomes? Int J Rehabil Res 2008;31:79–80.

Grotle M, Vøllestad N, Brox I J. Screening for yellow flags in first-time acute low back pain: reliability and validity of a Norwegian version of the Acute Low Back Pain Screening Questionnaire. Clin J Pain 2008;22:458–67.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29–36.

Heneweer H, Aufdemkampe G, Van Tulder MW, Kiers H, Stappaerts KH, Vanhees L. Psychosocial variables in patients with subacute low back pain: an inception cohort in primary care physical therapy in The Netherlands. Spine 2007;32:3586–92.

Hicks GE, Fritz JM, Delitto A, McGill SM. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. Arch Phys Med Rehabil 2005;86:1753–62.

Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008;58(5):632–41.

Hockings RL, McAuley JH, Maher CG. A systematic review of the predictive ability of the Örebro Musculoskeletal Pain Questionnaire. Spine 2008;33:E494–500.

Hough E, Stephenson R, Swift L. A comparison of manual therapy and active rehabilitation in the treatment of non specific low back pain with particular reference to a patient's Linton & Hallden psychological screening score: a pilot study. BMC Musculoskelet Disord 2007;8:106.

Hurley DA, Dusoit TE, McDonald SM, Moore AP, Baxter GD. How effective is the acute low back pain screening questionnaire for predicting 1-year follow-up in patients with low back pain? Clin J Pain 2001;17:256–63.

Jellema P, Van der Horst HE, Vlaeyen JW, Stalman WA, Bouter LM, Van der Windt DA. Predictors of outcome in patients with subacute low back pain differ across treatment groups. Spine 2006;31:1699–705.

Jellema P, Vander Windt DA, Vander Horst HE, Stalman WA, Bouter LM. Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. Brit J Gen Pract 2007;57:15–22.

Koes BW, Van Tulder MW, Thomas S. Diagnosis and treatment of low back pain. BMJ 2006;332:1430–4.

Kori SH, Miller RP, Todd DD. Kinisophobia: a new view of chronic pain behavior. Pain Manage 1990;Jan/Feb:35–43.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. Clin J Pain 2003;19:80–6.

Linton SJ, Hallden K. Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. Clin J Pain 1998;14:209–15.

Margison DA, French DJ. Predicting treatment failure in the subacute injury phase using the Örebro Musculoskeletal Pain Questionnaire: an observational prospective study in a workers’ compensation system. J Occup Environ Med 2007;49:59–67.

Morley S, Vlaeyen JWS. Epilogue to the special topic series. Clin J Pain Special Top Ser: Cogn-Behav Treat Chronic Pain 2005;21:69–72.

Neubauer E, Junge A, Pirron P, Seemann H, Schiltenwolf M, HKF-R 10 – screening for predicting chronicity in acute low back pain (LBP): a prospective clinical trial. Eur J Pain 2006;10:559–66.

Nordeman L, Nilsson B, Moller M, Gunnarsson R. Early access to physical therapy treatment for subacute low back pain in primary health care: a prospective randomized clinical trial. Clin J Pain 2006;22:505–11.

Roland M, Morris R. A study of the natural-history of back pain. 1. Development of a reliable and sensitive measure of disability in low-back-pain. Spine 1983;8:141–4.

Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assessment 1995;7:524–32.

Truchon M, Cote D. Predictive validity of the Chronic Pain Coping Inventory in subacute low back pain. Pain 2005;116:205–12.

Van Tulder M, Becker A, Bekkerling T, Breen A, deReel MT, Hutchinson A, et al. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006;15(Suppl. 2):S169–91.

Westman A, Linton SJ, Ohrvik J, Wahlen P, Leppert J. Do psychosocial factors predict disability and health at a 3-year follow-up for patients with non-acute musculoskeletal pain? A validation of the Örebro Musculoskeletal Pain Screening Questionnaire. Eur J Pain 2008;12(5):641–9.