

Eur Spine J (2012) 21 (Suppl 6):S773–S784
DOI 10.1007/s00586-011-1910-7

REVIEW ARTICLE

Predictive value of the Acute Low Back Pain Screening Questionnaire and the Örebro Musculoskeletal Pain Screening Questionnaire for persisting problems

Martin Sattelmayer · Tobias Lorenz ·
Christoph Röder · Roger Hilfiker

Received: 21 June 2011 / Accepted: 2 July 2011 / Published online: 26 July 2011
© Springer-Verlag 2011

Abstract

Introduction A small proportion of individuals with non-specific low back pain (NSLBP) develop persistent problems. Up to 80% of the total costs for NSLBP are owing to chronic NSLBP. Psychosocial factors have been described to be important in the transition from acute to chronic NSLBP. Guidelines recommend the use of the Acute Low Back Pain Screening Questionnaire (ALBPSQ) and the Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ) to identify individuals at risk of developing persistent problems, such as long-term absence of work, persistent restriction in function or persistent pain. These instruments can be used with a cutoff value, where patients with values above the threshold are further assessed with a more comprehensive examination.

Methods We systematically reviewed studies evaluating the accuracy of the ALBPSQ and ÖMPSQ to predict persistent problems.

Results The 13 included studies used different cutoff values for the screening questionnaires ranging from 68 to 147. The pooled sensitivity was 0.59 (0.43–0.74), while the pooled specificity was 0.77 (0.66–0.86). Heterogeneity (I^2) was 90.02% for sensitivity and 95.41% for specificity.

Conclusion Thus, we do not recommend the use of one cutoff value, but the use of a prediction model with all the individual items.

Keywords Back pain · Prognosis · Screening · Psychosocial

M. Sattelmayer (✉) · R. Hilfiker
Department of Physiotherapy,
HES-SO, University of Applied Sciences
Western Switzerland, Leukerbad, Switzerland
e-mail: martin.sattelmayer@hevs.ch

T. Lorenz
Klinik Adelheid, Unterägeri, Switzerland

C. Röder
Institute for Evaluative Research in Medicine,
University of Bern, Stauffacherstrasse 78,
3014 Bern, Switzerland

R. Hilfiker
Institute Health and Social Work,
HES-SO, University of Applied Sciences Western
Switzerland, Valais, Sion, Switzerland

R. Hilfiker
Institut de recherche en réadaptation et Clinique romande de
réadaptation (Suvacare), Sion, Switzerland

Introduction

Patients with musculoskeletal problems generate enormous economic loss to society [1] due to both absenteeism and reduced productivity (presenteeism) [2]. The majority of these costs are caused by a small proportion of patients with musculoskeletal problems developing persisting restriction in participation (e.g., work and other role fulfillments). Multimodal assessment and rehabilitation, which is recommended for patients at high risk for developing persistent problems, is expensive [3]. Therefore, these expensive assessments and interventions should only be allocated to those with a substantial risk. Early detection of these patients could improve the process of treatment allocation, optimize the cost-benefit ratio, and reduce the burden of disease for society, as well as the individual patient. Several guidelines for the management of back pain or other musculoskeletal pain conditions recommend the assessment of red flags at the first visit, and yellow flags

if a patient does not improve after 2–8 weeks after the onset of a new back pain episode (see Koes et al. [4] for an overview). Red flags are warning signs for serious diseases, such as cancer, fracture, inflammation, or progressive impairment of neurological function. Yellow flags indicate psychosocial factors suspected to be involved in the transition to chronic pain and restrictions in activity and participation [5, 6]. If no red flags are identified, the assessment of psychosocial factors should help us to (1) target the interventions and (2) support the decision of whether a multimodal, and thus expensive, rehabilitation should be envisaged, or whether monomodal treatment is sufficient [3]. One instrument that was built to integrate different biopsychosocial aspects into one questionnaire and one summary score is the Acute Low Back Pain Screening Questionnaire (ALBPSQ) [7]. This questionnaire is recommended by guidelines from New Zealand and Australia to identify individuals at risk of developing chronic non-specific low back pain (NSLBP) [8, 9]. The ALBPSQ was first published by Linton and Hallden [7], comprising different items from existing questionnaires. The Örebro Musculoskeletal Pain Screening Questionnaire (ÖMPSQ) is an adaptation of the ALBPSQ: one question is formulated differently, such that all musculoskeletal pain is addressed, instead of only back pain [10]. The ALBPSQ and ÖMPSQ each have 25 items, covering days off work, anxiety and tension, depression, pain, activities of daily living related to pain, coping, job satisfaction, fear-avoidance beliefs, and patient's expectations to recover. These items provide information about possible psychosocial problems, helping the interdisciplinary teams to discuss such problems with the patient and consider tailored interventions. A total score is calculated from 21 items and can range from 2 to 210 points. Higher values indicate more psychosocial problems. A cutoff value of 105 has been proposed for indicating those "at risk" for developing persistent problems [10, 11].

There are mainly two approaches for prediction [12]. In the predictive model approach, the risk of a given outcome is estimated by the means of appropriate regression models (e.g., logistic or cox regression). This method provides an estimated risk (e.g., a risk of 0.8 for persisting pain at 6 months means that among 100 similar patients, 80 would still have persisting pain at 6 months). However, a decision on treatment (e.g., usual care vs. expensive multimodal treatments) has to be made on a yes/no basis. The decision is relatively easy for patients with a very low or a very high risk, but it is more difficult for those with an intermediate risk. Therefore, some clinicians prefer the second approach, where the predictive value of risk factors are presented as true positive fraction (sensitivity) or true negative fraction (specificity). Both the capabilities (detection of patients who will and those who will not develop the given outcome)

can be summarized with a receiver operating characteristic curve (ROC curve).

Only few prognostic tests have both a high sensitivity and high specificity [13]; for most, there is a tradeoff between the two statistics. In the case of the ALBPSQ, lowering the cutoff would decrease the proportion of false negatives, but increase the proportion of false positives. Whether a predictive instrument should have a high sensitivity or a high specificity depends on its use: a high sensitivity (and thus a low specificity) will lead to a high proportion of patients referred to a multimodal, and thus expensive, rehabilitation, and due to the low specificity, not all of them need this expensive intervention (i.e., might get better without a multimodal intervention). On the other hand, a high specificity and thus a relatively lower sensitivity would lead to a high proportion of patients failing to receive multimodal rehabilitation. Therefore, clinicians and case managers have to carefully decide about the cutoff value. In the case of the ALBPSQ and ÖMPSQ, a high cutoff value could lead to high specificity and low sensitivity, while a low cutoff value could lead to high sensitivity and low specificity.

Most often, the ALBPSQ and ÖMPSQ will be used in a triage setting, where patients with a positive test result will then be assessed with tests that are more complex, such as a psychological assessment. Therefore, we presume that a high sensitivity would be preferable in the triage setting.

Hockings et al. [14] published a systematic review of the predictive value of the ALBPSQ and ÖMPSQ. However, no statistical pooling was performed and since then several new relevant articles have been published. With this diagnostic meta-analysis, we set out to evaluate how accurate the ALBPSQ and ÖMPSQ could predict the persistent problems, such as sick leave, pain, or decreased function in patients with NSLBP or musculoskeletal problems.

Methods

Identification of studies

We included prospective studies that used either the ALBPSQ or ÖMPSQ, and reported the predictive values for pain, work status, sick days, or function at follow-up in patients with musculoskeletal pain (back pain, shoulder pain, etc.). We searched PubMed, Embase, and Web of Science from 1998 (first publication of the ALBPSQ) to March 2010, and considered only articles in the English language. The search string is available on request. It focused on back pain, musculoskeletal pain, and the ALBPSQ or ÖMPSQ. In addition, we entered the original studies of the two questionnaires [7, 10] into the Science Citation Index Database (Web of Science, Thomson ISI, Philadelphia, PA, USA) and Google Scholar to identify citing studies.

Selection criteria

Two reviewers independently screened the abstracts and titles retrieved by the electronic searches. Two reviewers, using a checklist to assess the inclusion criteria, reviewed the full texts of 46 articles. Prospective studies were included if one of the two questionnaires was assessed in patients with back pain or musculoskeletal pain (of any duration) and if follow-up data were available for recovery (pain, function, work status, etc.), to construct a 2×2 table to calculate the sensitivity and specificity; i.e., true negative, false negative, true positive, and false positive values were available. Disagreement about inclusion was resolved by consensus or by a third reviewer.

Methodological quality assessment

We used the criteria proposed by Hayden [15] to assess the study quality and its risk of bias: study participation, study attrition, prognostic factor measurement and outcome measurement, and confounding measurement and account.

Data extraction

Two reviewers independently extracted the data of the studies. Inconsistencies were resolved by discussion between both the reviewers.

For the description of the study, we extracted the setting, patient recruitment, time point of assessment (relative to the course of the pain problem), time point of follow-up assessment, whether or not patients were enrolled consecutively, number of patients eligible, number of patients included at baseline, number of patients assessed at follow-up, description of pain duration and location, and definition and assessment of outcomes.

Statistical analysis

True positives, false positives, true negatives, and false negatives were calculated for different cutoff values, if possible. If data were not available to calculate these values, the authors were contacted with the question to provide values for sensitivity and specificity for at least two cutoff values, if possible for the most consistently reported cutoff values (i.e., 105).

Meta-analysis

We pooled sensitivity and specificity with a bivariate diagnostic meta-analysis (metandi: user written commands in STATA) [16, 17]. This method fits a two-level model, with independent binomial distributions for the true positives and true negatives, conditional on the sensitivity and

specificity in each study, and a bivariate normal model for the logit transforms of sensitivity and specificity between the studies [18].

We used a general linear mixed model approach for bivariate meta-analysis.

As different studies presented different thresholds (cut-off values), we plotted a summary ROC curve. A 95% confidence ellipse (bivariate model) and a 95% prediction ellipse (where we would expect that 95% of future studies will lie) within the ROC space were calculated. Each data point in the summary ROC represented a separate study.

If one study reported on more than one outcome, we chose, if possible, a work-related outcome. If one study reported on more than one follow-up time points, we selected only the longest follow-up period.

Heterogeneity

Statistical heterogeneity was assessed with the I^2 statistic of the pooled sensitivities and specificities. As a rule of thumb, 25% or less may be considered as low heterogeneity, 50% as moderate, and 75% as high heterogeneity [19].

Due to the low number of studies, a meta-regression to evaluate the potential variables explaining heterogeneity was not possible. The possible factors explaining the heterogeneity could have been: (1) pathology, (2) baseline prevalence of patients at risk (spectrum), (3) kind of outcome, and (4) time point of follow-up.

Results

The search in the electronic databases identified 46 potential eligible studies. From these, 14 studies [7, 10, 20–31] met our inclusion criteria, reporting on 16 samples. Six studies evaluated the accuracy to predict the persistent problems with the ALBPSQ [7, 22–25, 30], while eight studies and ten samples reported on the ÖMPSQ [10, 20, 21, 26–29, 31]. Table 1 shows the characteristics of the included studies. The size of the study population ranged between 89 [29] and 328 [22] participants. Seven studies focused not only on patients with low back pain [10, 20, 21, 27, 29–31], but also on those with other problems as well. Most studies assessed patients within 12 weeks after the onset of a pain episode. One study had 85% of patients with longer pain duration [31]. The outcome measurements were heterogeneous and included the following: non-return to work [20, 25, 29], decreased sick leave level [31], sick days [7, 10, 21–23, 27, 30], recovery assessed with two questions of overall recovery [24], and Roland–Morris Questionnaire [28],

Table 1 Characteristics of the included studies

References	Version of Questionnaire (language, country)	N at baseline, (n at follow-up, % at follow-up)	Characteristics	Diagnosis, inclusion criteria	Time point of assessment	Clinical outcome (follow-up)
Dunstan et al. [20]	ÖMPSQ (English, Australia)	196 (NA)	34% women Age 18–65 years 22% laborers 36% back injuries 42% women Mean age 39 ± 9 years Pain site: lower back (100%); lower back and leg (55%); lower back and neck (21%); back, neck and shoulder (9%) Mean duration of pain: 5.3 weeks 44% women Age: 18–75 years	Musculoskeletal strain or sprain	Subacute phase (4–12 weeks post injury)	Non-return to work at 6 months
Gable et al. [21]	ÖMPSQ (English, Australia)	110 (110, 100%)	Workers' compensation patients' (aged 18–58 years) identified by a medical practitioner with musculoskeletal symptoms in the lower back, including pain	At admission	At admission	Long-term absenteeism (>28 days) at 6 months
Grimmer-Somers et al. [22]	ALBPSQ (Auckland region, New Zealand)	328 (328, 100%)	Patients with acute low back pain that did not resolve after two or more general practitioner (GP) visits. The predominant diagnosis was lumbar sprain (73.3%), followed by lumbar spine pain (7.8%)	Second GP visit	Second GP visit	>90 sick days at 1 year
Grotle et al. [23]	ALBPSQ (Norwegian, Norway)	123 (112, 91%)	54% women Mean age: 38 years (SD 10.1) 46% had less than 12 years of education 38% were smoking 89% were in an ordinary working situation 24% had 2 or more neurological signs on clinical examination	LBP with or without radiation and lasting 3 weeks, age 18–60 years, and who contacted a healthcare provider because of low back pain for the first time	At admission	>30 sick days at 1 year
Heneweer et al. [24]	ALBPSQ (Dutch, The Netherlands)	66 (56, 85%)	Women: 39% Mean age: 40.8 years (recovered), 34.1 years (not recovered) Duration of current complaint: <4 weeks: 65% (recovered), 36% (not recovered) 4–6 weeks: 29% (recovered), 24% (not recovered) 7–12 weeks: 7% (recovered), 40% (not recovered)	New referrals consulting physical therapist for the first time with a first or a new episode of low back pain	At admission	Recovery at 12 weeks (recovery assessed with two questions on overall improvement: recovery yes/no, work absenteeism yes/no)

Table 1 continued

References	Version of Questionnaire (language, country)	N at baseline, (n at follow-up, % at follow-up)	Characteristics	Diagnosis, inclusion criteria	Time point of assessment	Clinical outcome (follow-up)
Hurley et al. [25]	ALBPSQ (English, Northern Ireland)	118 (90, 76%)	60% women Median age: 41.5 years (17–77 years) A wide range of physiotherapy was used 76 persons were employed 56% <3 months pain duration in this episode 45% had pain in the previous year 41% had previous physiotherapy 70% had pain not radiating below the knee	Referred by a medical practitioner for physiotherapy treatment of low back pain (without red flags)	Within 6 weeks of onset of low back pain (70%)	Non-return to work at 1 year
Jellema et al. [26]	ÖMPSQ (Dutch, The Netherlands)	314 (298, 95%)	48% women Mean age: 42.7 years (SD 11.6) Dutch nationality: 98% Employed: 82% Sick leave because of low back pain (among working population): 38% Episodes in last year: 1 or 2 episodes: 60% 3 or more: 19% Exacerbation: 21% Pain radiation below knee: 14%	New episode of low back pain, duration <12 weeks; or exacerbation of mild symptoms Age 19–65 years	At admission	Unfavorable course at 1 year (patients rated their recovery on a 7-point Likert scale: very much improved; much improved; slightly improved, no change, slightly worse, much worse. Unfavorable: score of slightly improved or worse
Linton et al. [7]	ALBPSQ (Swedish, Sweden)	147 (137, 97%)	65% women Mean age: 43.6 years (22–63 years) 58% back pain 53% neck pain 55% shoulder pain All patients received treatment	Acute or subacute pain (back or neck) <4 months off work in last year because of the pain	At admission	>30 sick days at 6 months
Linton et al. [10]	ÖMPSQ (Swedish, Sweden)	122 (107, 88%)	48% women Mean age: (22–66 years) 93% born in Sweden 56% low back pain 44% neck pain 44% shoulder pain	Acute or subacute pain (duration: <3 months) in the back or neck; <6 months of accumulated sick leave during the past year	At admission	>30 sick days and recovery of functional ability at 6 months

Table 1 continued

References	Version of Questionnaire (language, country)	N at baseline, (n at follow-up, % at follow-up)	Characteristics	Diagnosis, inclusion criteria	Time point of assessment	Clinical outcome (follow-up)
Linton et al. [27]	ÖMPSQ (Swedish, Sweden)	Sample 1 (occupational sample): 324 (unclear) Sample 2 (primary care sample) 185 (183, 99%)	Occupational sample: 67% women Age 18–60 years back pain Primary care sample: 83% females Mean age: 48 years back or neck pain	Occupational sample: suffering from musculoskeletal pain; no determined red flags; and <3 months of sick leave Primary care sample: employed; between 20 and 60 years of age; report of nonspecific back or neck pain; <4 months of sick leave during the past year for spinal pain; and no physical therapy during the past year	At admission	Sick leave during the follow-up at 1 year
Maher et al. [28]	ÖMPSQ (Norwegian, Norway; English, Australia and New Zealand)	Norwegian cohort: 97 (7, 9%) Australasian cohort: 136 (0, 0%)	Norwegian cohort: Age: 38.7 years Female: 56% Smoker: 33% Currently undertakes regular exercise: 59% Duration of current episode of low back pain: <1 week: 50% 1–2 weeks: 21% 2–3 weeks: 29% Australasian cohort: Age: 43.3 years Smoker: 43% Currently undertakes regular exercise: 57% Duration of current episode of low back pain: 6–8 weeks: 45% 9–11 weeks: 38% 12 weeks: 17%	Norwegian cohort: Acute pain in the lower back (duration: <3 weeks), patients previously treated for low back pain were excluded Australasian cohort: Subacute nonspecific low back pain (duration from 6 weeks to 3 months)	At admission	Disability; >4 on the RMQ at 1 year

Table 1 continued

References	Version of Questionnaire (language, country)	N at baseline, (n at follow-up, % at follow-up)	Characteristics	Diagnosis, inclusion criteria	Time point of assessment	Clinical outcome (follow-up)
Margison et al. [29]	ÖMPSQ (English and French, Canada)	English: 200 for model derivation. 122 validation sample French: 89 combined validation sample: 211	Women: 39% Mean age: 41.5 years (SD 9.8) Site of primary injury Back: 51% Multiple body areas: 25% Other body areas: 12% Mean weeks from accident to beginning work conditioning: 16.7 (SD 13.8) Pain sites: Neck: 23% Shoulders: 32% Upper back: 21% Lower back: 56% Arm, wrist, and hand: 17% Leg, ankle, and foot: 32% Other pain sites: 14% Mean Pain duration (weeks): 6.6 (SD 1.5) Mean pain intensity in the past 3 months: 7.1 (SD 2.1)	Case manager referred patients when duration on benefit exceeded injury-specific disability duration guidelines (e.g., week 7 for soft-tissue injuries) and there was no plan to begin a Return to Work program within 2 weeks Treatment received after baseline assessment: 6 weeks work conditioning	At admission	Non-return to work at 6 weeks
Vos et al. [23, 30]	ALBPSQ (Dutch, The Netherlands)	187 (180, 96%)	Women: 64% Employed: 79% On sick leave: 36% Smoking: 33% Underwent previous treatment for neck pain: 40% Duration of acute neck pain shorter than 2 weeks: 42% Pain radiating to: shoulders: 56% Arm(s): 37% Back: 5% Between shoulder blades: 41%	Patients with a new or recurrent episode (after a period free of pain for at least 3 months) of acute neck pain (0–6 weeks after onset) and 18 years or older were included	At admission	>7 days sick leave at 1 year

Table 1 continued

References	Version of Questionnaire (language, country)	N at baseline, (n at follow-up, % at follow-up)	Characteristics	Diagnosis, inclusion criteria	Time point of assessment	Clinical outcome (follow-up)
Westman et al. [31]	ÖMPSQ (Swedish, Sweden)	158 (149, 94%)	<p>Women: 70%</p> <p>Employed: 73%</p> <p>Unemployed: 19%</p> <p>Students: 3%</p> <p>Primary pain site: neck: 62%</p> <p>Shoulder: 70%</p> <p>Back–upper part: 31%</p> <p>Back–lower part: 85%</p> <p>Leg: 35%</p> <p>Other symptoms: 37%</p> <p>Pain duration in weeks:</p> <p>0–11: 15%</p> <p>12–23: 15%</p> <p>24–52: 15%</p> <p>≥53: 55%</p> <p>Sick leave at baseline:</p> <p>25% sick leave: 4%</p> <p>50% sick leave: 12%</p> <p>75% sick leave: 3%</p> <p>100% sick leave: 80%</p>	<p>Nonacute pain problems (1–6 months sick leave); age 18–65 years; sick listed between 28 and 180 days and/or had consulted the doctor about the same problem ≥3 times in the last 12 months</p>	At admission	Maintained or increased sick leave level at the 3 years of follow-up

Risk of bias of the included studies

For the categories of study participation, study attrition, prognostic factor measurement, and outcome measurement, we considered the risk of bias to be minimal. For the confounding measurement and account, there may be a moderate risk of bias due to uncontrolled factors, such as treatment. For the criterion “analysis,” we could calculate true positives, true negatives, false positives, and false negative for all studies, except for the study by Westman et al. [31], where we estimated the content of the four cells with the reported sensitivity, specificity, and number of patients.

Overall results

The pooled summary sensitivity was 0.59 (0.43–0.74), implying that 59% of the patients developing persistent problems would be correctly classified as “at risk” when the ALBPSQ or ÖMPSQ was applied, whereas the pooled specificity was 0.77 (0.66–0.86), implying that 77% of the patients not developing persistent problems would be

classified as “at no risk.” Heterogeneity for sensitivity and specificity was (I^2) 90.02% (95% CI 86.24–93.79) and 95.41% (95% CI 94.03–96.78), respectively (see Figs. 1, 2).

Discussion

With this systematic review and meta-analysis, we found that the prediction with the ALBPSQ or ÖMPSQ has too much heterogeneity to recommend the use of a definite cutoff value (e.g., 105). Due to this high heterogeneity, the pooled sensitivity and specificity should be regarded with care. Overall, the two questionnaires would appear to have only weak to moderate predictive value for the development of persistent problems.

This is the first attempt to statistically summarize the predictive value of two widely used questionnaires for the prediction of persistent musculoskeletal pain.

There are several limitations inherent to our meta-analysis: we pooled the different outcomes, being aware of different problems with this approach because: (1) the effect sizes vary between the different outcomes and the

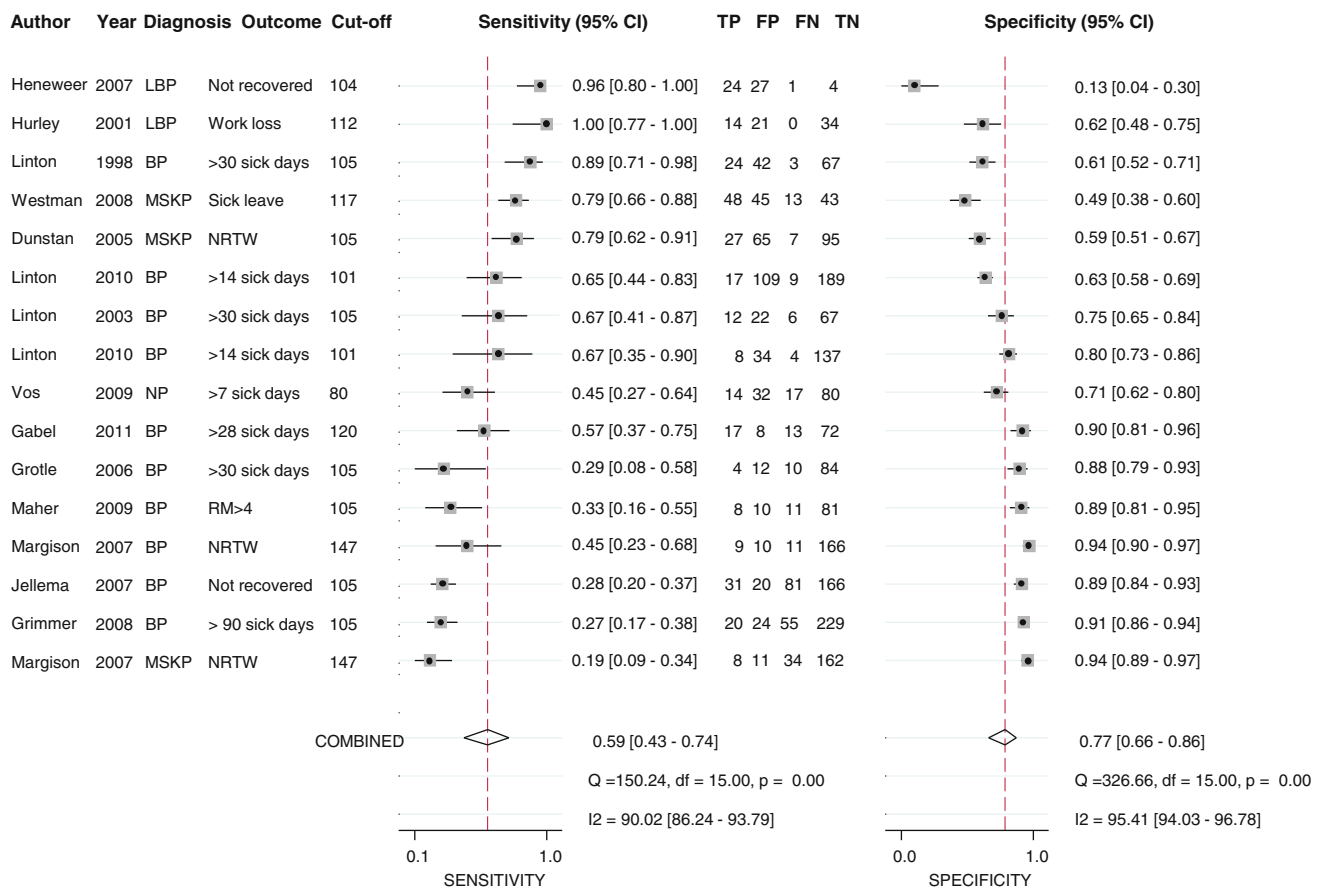


Fig. 1 Forest plot: ALBPSQ or ÖMPSQ with different outcomes. TP true positive, FP false positive, FN false negative, TN true negative, LBP low back pain, BP back pain, MSKP musculoskeletal pain, NP neck pain, NRTW non-return to work, RM Roland–Morris Questionnaire

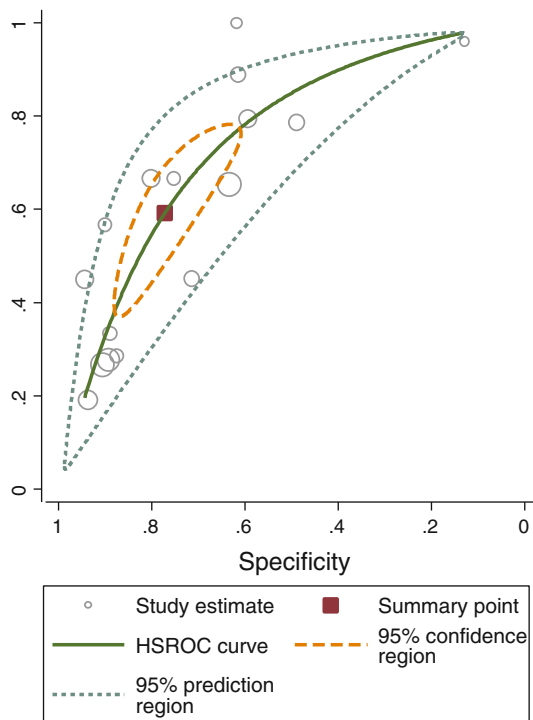


Fig. 2 Summary ROC plot: ALBPSQ or ÖMPSQ with different outcomes

statistical summary of these data may lead to a biased summary estimate [32], (2) different outcomes most probably have different predictors; e.g., work loss most likely has a different predictor than function, and (3) interpretation of the summary statistics is difficult when different outcomes are pooled. Nevertheless, as a separate analysis of the different outcome was not possible because of the low number of studies for each outcome, we presume that pooling of the different outcomes is the best approach to summarize the overall predictive value of the two questionnaires.

Furthermore, from the results of this study, it is not possible to define the best threshold (cutoff value). The high heterogeneity limits the information of the summary values for sensitivity and specificity, as the clinician cannot estimate the influence of a given setting on sensitivity and specificity at the recommended threshold of 105.

Reasons for heterogeneity might be as follows: (1) predictive values differ between patients with back pain or those with other pain sources, as shown, e.g., in the study by Margison et al. [29], and most of the studies had a different case-mix. (2) Outcomes were different, and even the work-related outcomes were assessed differently in different studies. (3) Time points of follow-up assessments differed among the studies. Although sensitivity and specificity are not inherently influenced by the prevalence of the outcomes, it has been demonstrated that the patient

spectrum clearly influences the sensitivity and specificity [33]. (4) Treatment may successfully address predictive factors and thus make the “static” predictive model less predictive (to adjust for this, one could assess the predictors several times during a follow-up period and adjust the predictive model). (5) The examined studies were rather small with few cases, which can lead to different sensitivities and specificities due to sample variability (i.e., by chance).

Several published and ongoing studies have used one of the two questionnaires in conjunction with other (potential) predictors. Further systematic reviews will indicate whether these combinations improve the predictive value.

In the case of the ALBPSQ and ÖMPSQ, dichotomizing the prediction (good/bad outcome) will lead either to a high sensitivity with a relative lower specificity or to a high specificity with a relative lower sensitivity, as shown in Figs. 1 and 2. However, more promising would be to use the items of the questionnaire as a prediction rule (a probability calculation with a formula giving a weight to each question in the questionnaire). One could argue that case managers or medical doctors need cutoff values to be able to make a clear and objective decision. However, probability scores would allow us to integrate further information, such as the “gut feeling” of the case manager. Furthermore, one might argue that if the workload is high and the medical doctor could not assess all patients in detail, he/she needs to have clear cutoff values to decide on which patient should be assessed thoroughly. However, a clear cutoff value could also be based on probability models; e.g., one could refer to in-depth assessment of all patients with a risk probability of over 90%. Furthermore, this cutoff value can be adapted according to resources, i.e., can be increased when resources are sparse.

A second solution would be to define two thresholds, a higher one for high specificity, and a lower one for high sensitivity. This would allow ruling out the risk for prolonged sick leave or function-related problems in a patient with a value below the lower threshold, while ruling in the risk in a patient with a value higher than the upper threshold. For all patients with values between the two thresholds, no decision can be made and further tests must be carried out.

Strictly, the use of an overall score only makes sense for unidimensional scales [34]. This is questionable regarding the ALBPSQ or ÖMPSQ, as they were constructed with items from questionnaires covering different constructs (dimensions). Therefore, the use of sub-scores may improve prediction. There is some evidence that subscales have better predictive values; e.g., in [24]: The ALBPSQ subscales, “pain” and “function/pain/psychology and fear-avoidance beliefs,” showed better predictive values than the total score (area under the ROC curve: 0.641 (total

score), 0.855 (function/pain/psychology and fear-avoidance beliefs), and 0.817 (pain)).

Even if we assumed that a summary score for this questionnaire was sound and decided not to use a probability model, we would still be unable to make a clear recommendation on the cutoff value because of the heterogeneity of the cutoff value used in the studies. One approach to deal with this uncertainty about the cutoff value might be to build a database including about 100 patients in a given setting and to calculate the optimal cutoff value for this setting. For example, Margison et al. [29] decided to use three cutoff values after the evaluation of their patients: <99 for the classification as “patient with mainly biomedical pain generator;” 140–147 as “high risk for prolonged disability;” and >147 as “very high risk for prolonged disability.”

Conclusion

Practitioners should not use thresholds of the summary score of the ALBPSQ or ÖMPSQ to identify individuals at risk of developing chronic pain and disability. Instead, they should pay more attention to high values for specific items in the scale that highlight an individual’s problems in the respective domain. We recommend the use of an individual risk profile, instead of the summary score. Researchers should integrate the questionnaire into the probability models with other risk factors and test the models in large cohort studies.

Acknowledgment We thank Debra A. Dunstan, Margreth Grotle, Hans Heneweer, Philip Gabel, Christopher Maher, and Douglas Margison for the additional data provided. We thank Douglas Margison and Lucas Bachmann for their valuable input on preliminary drafts of the manuscript.

Conflict of interest None.

References

1. Becker A, Held H, Redaelli M, Strauch K, Chenot JF, Leonhardt C, Keller S, Baum E, Pflingsten M, Hildebrandt J, Basler HD, Kochen MM, Donner-Banzhoff N (2010) Low back pain in primary care: costs of care and prediction of future health care utilization. *Spine (Phila Pa 1976)* 35:1714–1720
2. Wieser S, Horisberger B, Schmidhauser S, Eisenring C, Brugger U, Ruckstuhl A, Dietrich J, Mannion AF, Elfering A, Tamcan O, Muller U (2010) Cost of low back pain in Switzerland in 2005. *Eur J Health Econ*. doi:10.1007/s10198-010-0258-y
3. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G (2006) European guidelines for the management of chronic nonspecific low back pain. *Eur Spine J* 15(Suppl 2):S192–S300. doi:10.1007/s00586-006-1072-1
4. Koes BW, van Tulder M, Lin CW, Macedo LG, McAuley J, Maher C (2010) An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. *Eur Spine J*. doi:10.1007/s00586-010-1502-y
5. Hilfiker R, Bachmann LM, Heitz CA, Lorenz T, Joronen H, Klipstein A (2007) Value of predictive instruments to determine persisting restriction of function in patients with subacute non-specific low back pain: systematic review. *Eur Spine J* 16:1755–1775. doi:10.1007/s00586-007-0433-8
6. Linton SJ (2001) Occupational psychological factors increase the risk for back pain: a systematic review. *J Occup Rehabil* 11:53–66
7. Linton SJ, 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
8. Kendall NAS, Linton SJ, Main CJ (1997) Guide to assessing psycho-social yellow flags in acute low back pain: risk factors for long-term disability and work loss. Accident Compensation Corporation and the New Zealand Guidelines Group, Wellington
9. Australian Acute Musculoskeletal Pain Guidelines Group (2003) Evidence-based management of acute musculoskeletal pain. Australian Academic Press, Brown Hills
10. Linton SJ, Boersma K (2003) Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Orebro Musculoskeletal Pain Questionnaire. *Clin J Pain* 19:80
11. WorkCoverSA (2010) Managing acute-subacute low back pain: clinical practice guide. In: WorkCoverSA (ed) Adelaide, South Australia
12. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y (2008) Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol* 167:362–368
13. Straus SE, Richardson WS, Glasziou P, Haynes RB (2005) Evidence-based medicine: how to practice and teach EBM. Elsevier Churchill Livingstone, Edinburgh
14. Hockings RL, McAuley JH, Maher CG (2008) A systematic review of the predictive ability of the Orebro Musculoskeletal Pain Questionnaire. *Spine (Phila Pa 1976)* 33:E494–E500. doi:10.1097/BRS.0b013e31817ba3bb
15. Hayden JA, Cote P, Bombardier C (2006) Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 144:427–437. doi:10.1093/ajcp/144/6/427
16. Harbord R (ed) (2008) metandi: Stata module for meta-analysis of diagnostic accuracy. In: Statistical Software Components. Boston College Department of Economics, Boston
17. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 8:239–251. doi:10.1093/biostatistics/kx1004
18. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwiderman AH (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 58:982–990
19. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327:557–560. doi:10.1136/bmj.327.7414.557
20. Dunstan DA, Covic T, Tyson GA, Lennie IG (2005) Does the Orebro Musculoskeletal Pain Questionnaire predict outcomes following a work-related compensable injury? *Int J Rehabil Res* 28:369–370. doi:10.1007/s00586-010-1509-4
21. Gabel CP, Melloh M, Yelland M, Burkett B, Roiko A (2011) Predictive ability of a modified Orebro Musculoskeletal Pain Questionnaire in an acute/subacute low back pain working population. *Eur Spine J* 20:449–457. doi:10.1007/s00586-010-1509-4
22. Grimmer-Somers K, Prior M, Robertson J (2008) Yellow flag scores in a compensable New Zealand cohort suffering acute low back pain. *J Pain Res* 1:15–25

23. Grotle M, Vollestad NK, Brox JI (2006) 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* 22:458–467. doi:[10.1097/01.ajp.0000208243.33498.cb](https://doi.org/10.1097/01.ajp.0000208243.33498.cb)
24. Heneweer H, Aufdemkampe G, van Tulder MW, Kiers H, Stappaerts KH, Vanhees L (2007) Psychosocial variables in patients with (sub) acute low back pain: an inception cohort in primary care physical therapy in The Netherlands. *Spine (Phila Pa 1976)* 32:586–592. doi:[10.1097/01.brs.0000256447.72623.56](https://doi.org/10.1097/01.brs.0000256447.72623.56)
25. Hurley DA, Dusoir TE, McDonough SM, Moore AP, Baxter GD (2001) 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* 17:256–263
26. Jellema P, van der Windt DA, van der Horst HE, Stalman WA, Bouter LM (2007) Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *Br J Gen Pract* 57:15–22
27. Linton SJ, Nicholas M, Macdonald S (2010) Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire. *Spine (Phila Pa 1976)*. doi:[10.1097/BRS.0b013e3181f8f775](https://doi.org/10.1097/BRS.0b013e3181f8f775)
28. Maher CG, Grotle M (2009) Evaluation of the predictive validity of the Orebro Musculoskeletal Pain Screening Questionnaire. *Clin J Pain* 25:666–670. doi:[10.1097/AJP.0b013e3181a08732](https://doi.org/10.1097/AJP.0b013e3181a08732)
29. Margison DA, French DJ (2007) Predicting treatment failure in the subacute injury phase using the Orebro Musculoskeletal Pain Questionnaire: an observational prospective study in a workers' compensation system. *J Occup Environ Med* 49:59–67. doi:[10.1097/JOM.0b013e31802db51e](https://doi.org/10.1097/JOM.0b013e31802db51e)
30. Vos CJ, Verhagen AP, Koes BW (2009) The ability of the Acute Low Back Pain Screening Questionnaire to predict sick leave in patients with acute neck pain. *J Manipulative Physiol Ther* 32:178–183. doi:[10.1016/j.jmpt.2009.02.004](https://doi.org/10.1016/j.jmpt.2009.02.004)
31. Westman A, Linton SJ, Ohrvik J, Wahlen P, Leppert J (2008) Do psychosocial factors predict disability and health at a 3-year follow-up for patients with non-acute musculoskeletal pain? A validation of the Orebro Musculoskeletal Pain Screening Questionnaire. *Eur J Pain* 12:641–649. doi:[10.1016/j.ejpain.2007.10.007](https://doi.org/10.1016/j.ejpain.2007.10.007)
32. Puhan MA, Soesilo I, Guyatt GH, Schunemann HJ (2006) Combining scores from different patient reported outcome measures in meta-analyses: when is it justified? *Health Qual Life Outcomes* 4:94. doi:[10.1186/1477-7525-4-94](https://doi.org/10.1186/1477-7525-4-94)
33. Leeflang MM, Bossuyt PM, Irwig L (2009) Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol* 62:5–12. doi:[10.1016/j.jclinepi.2008.04.007](https://doi.org/10.1016/j.jclinepi.2008.04.007)
34. Streiner DL (2003) Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *J Pers Assess* 80:217–222