Accuracy of the Örebro Musculoskeletal Pain Questionnaire and Work Assessment Triage Tool for selecting interventions in workers with spinal conditions

Eloi Gergelé, Eric Parent and Douglas P. Gross

*Department of Physical Therapy, University of Alberta, Edmonton, AB, Canada

**Grenoble Alpes University, Grenoble, France

Received 14 June 2020
Accepted 17 November 2020

Abstract.

BACKGROUND: Accurate clinical decision support tools may help clinicians select appropriate interventions for patients with spinal conditions. The Örebro Musculoskeletal Pain Questionnaire (ÖMPQ) is a screening questionnaire extensively studied as a predictive tool. The Work Assessment Triage Tool (WATT) is a clinical decision support tool developed to help select interventions for injured workers.

OBJECTIVE: To compare the classification accuracy of the ÖMPQ and WATT to clinician recommendations for selecting interventions leading to a successful return to work in patients with spinal conditions.

METHODS: A secondary analysis was undertaken of data from injured workers with spinal conditions assessed between 2013 and 2016. We considered it a success if the workers did not receive wage replacement benefits 30 days after assessment. Analysis included positive likelihood ratio (LR+) as an indicator of predictive accuracy.

RESULTS: Within the database, 2,872 patients had complete data on the ÖMPQ, WATT, and clinician recommendations. At 30 days, the ÖMPQ was most accurate for identifying treatments that lead to successful outcomes with a LR+ = 1.51 (95% Confidence Interval 1.26–1.82) compared to 1.05 (95% Confidence Interval 1.02–1.09) for clinicians, and 0.85 (95% Confidence Interval 0.79–0.91) for the WATT.

CONCLUSIONS: All tool recommendations had poor accuracy, however the ÖMPQ demonstrated significantly better results.

Keywords: Spine, decision making, return to work, occupational health, rehabilitation

1. Introduction

Spinal conditions, and especially low back pain, have a major public health impact on the working population as they are leading causes of disability and productivity loss [1,2]. Improved management strategies are needed, especially clinical tools that can help clinicians select effective treatments based on individual worker characteristics (i.e., clinical decision support tools) [3–5]. Accurate and validated clinical screening and decision support tools would allow clinicians to select the most appropriate interventions to facilitate timely recovery and return to work in individuals off work due to spinal conditions [6–8].

The Örebro Musculoskeletal Pain Questionnaire
(ÖMPQ) is a screening questionnaire that has been extensively studied as a screening and predictive tool in the work rehabilitation field [9,10]. The ÖMPQ has been shown to have a 89% sensitivity in the prediction of longer-term (more than 30 days) sick leave in patients with back pain as well as high reliability (Intraclass Correlation Coefficient = 0.98) [9,10]. However, its use as a clinical decision support tool has not been studied. Based on the ÖMPQ total score, patients are placed into one of three categories illustrating the risk of chronicity: low, moderate or high. High risk indicates the individual is more likely to experience delayed recovery and return to work, which has been demonstrated in workers with low back pain [11–13]. Theoretically, higher risk patients should warrant more intensive interventions, however, the accuracy and validity of the ÖMPQ as a clinical decision support tool for selecting treatments has not been previously studied.

The Work Assessment Triage Tool (WATT) is another clinical decision-support tool designed using state-of-the-art machine learning techniques to help select rehabilitation interventions for injured workers [14]. In the original WATT development study conducted on a sample that included a variety of work-related musculoskeletal disorders, classification accuracy of the WATT for selecting successful treatments was found to be higher than recommendations of clinicians [14]. However, in external validation studies [15,16], accuracy of the WATT was more modest and further development and testing of the WATT was recommended.

Additional accuracy and validity testing of the ÖMPQ and WATT are needed in workers with spinal conditions to determine if either can classify patients into appropriate intervention groups better than recommendations of clinicians. Therefore, the objective of this study was to compare the classification accuracy of the ÖMPQ and WATT to clinician recommendations for selecting appropriate interventions leading to a successful return to work in patients with spinal conditions.

2. Materials and methods

2.1. Study design

We conducted a historical comparative cohort study on patients with spinal conditions using data from the administrative and clinical databases of the Workers’ Compensation Board – Alberta (WCB-Alberta). This is a secondary analysis of a previously collected dataset used for evaluating the WATT among workers with a variety of musculoskeletal conditions. Details of the data collection procedures and variables in the dataset are available elsewhere [15]. In brief, all clinical measures in the dataset, including the ÖMPQ and WATT, were collected during routine clinical care as part of a return-to-work assessment conducted to determine patients’ ability to return to work or need for further treatment. We followed the Standards for QUality Improvement Reporting Excellence (SQUIRE 2.0) guidelines to report our results [17].

2.2. Population

Data were extracted from the WCB-Alberta database on patients considered for rehabilitation between January 2013 and December 2016. We used ICD9 diagnostic codes to identify patients with specific or nonspecific spinal conditions. Relevant diagnostic codes that were considered as spinal conditions and the actual number of cases in our study are provided in Supplementary Table 1. We excluded patients who did not have a spinal condition or who did not have a complete ÖMPQ score.

2.3. Procedures

This study was approved by the University of Alberta Health Research Ethics Board. No patients were recruited/consented since this study relied on archived program evaluation data from the WCB-Alberta database. More details on the data extraction process are provided elsewhere [15]. From the existing data, we created the intermediate variables needed to perform the analysis.

2.4. Measures

2.4.1. ÖMPQ

The ÖMPQ is a 25-item questionnaire that evaluates psychosocial and work-related barriers to recovery [9, 12]. ÖMPQ scores range between 0 and 210, but the overall score is categorized using cut-offs to represent a patient’s level of risk of delayed recovery (i.e., low, medium, or high risk). Originally, the ÖMPQ cutoffs were defined as 90 and 105 to discriminate low, medium and high risk of long-term disability [9]. Cutoffs used by the WCB-Alberta are slightly higher because the jurisdiction has found their population typically scores higher on the tool. The values of the WCB-Alberta
Table 1

<table>
<thead>
<tr>
<th>Cutoffs</th>
<th>Low risk</th>
<th>Moderate risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original ÖMPQ score cut-off:</td>
<td>&lt; 90</td>
<td>90 to 105</td>
<td>&gt; 105</td>
</tr>
<tr>
<td>Alternate Alberta ÖMPQ score cut-off:</td>
<td>&lt; 105</td>
<td>105 to 130</td>
<td>&gt; 130</td>
</tr>
<tr>
<td>Treatment recommended:</td>
<td>No further rehabilitation, community single service provider (physical therapy, chiropractic, other)</td>
<td>Functional restoration, workplace-based intervention, hybrid of these two programs</td>
<td>Complex interdisciplinary biopsychosocial rehabilitation</td>
</tr>
</tbody>
</table>

Cutoffs are 105 and 130 to discriminate low, medium and high risks. We conducted our analysis with both the original and the WCB-Alberta cutoffs.

For use as a clinical decision support tool, the different ÖMPQ risk categories lead to different treatment recommendations. Within the WCB-Alberta system [18], various rehabilitation interventions were available including: 1) single service rehabilitation provider (i.e., physiotherapy, chiropractic, other provider); 2) functional restoration; 3) workplace-based intervention; 4) hybrid functional restoration/workplace-based intervention; 5) complex interdisciplinary biopsychosocial rehabilitation; or 6) no further rehabilitation. These programs have been described in greater detail elsewhere [14,18]. According to the WCB-Alberta, the complex interdisciplinary biopsychosocial program is designed for patients with “significant barriers to return to work” [19]. We considered this complex interdisciplinary program as the only intervention adequate for patients in the ÖMPQ ‘high risk’ category. Functional restoration, workplace-based intervention, hybrid functional restoration/workplace-based intervention have been described as interventions used “to address the medical, functional, musculoskeletal, psychosocial, and vocational needs of the worker” [20]. We considered these interventions adequate for patients in the ÖMPQ moderate risk category. We considered the “single service interventions” and “no treatment” options as adequate for patients in the ÖMPQ low risk category. The two ÖMPQ risk level cut-offs used and corresponding treatment recommendations are shown in Table 1.

2.4.2. WATT

WATT is an 18-item computerized clinical decision-support tool that provides rehabilitation recommendations based on an algorithm developed using machine learning [14]. The tool provides recommendations using individual worker-level characteristics for different rehabilitation program possibilities available within the WCB-Alberta jurisdiction (listed above). Validity of the WATT has been studied in injured workers with a variety of musculoskeletal conditions. In the original development study, the WATT outperformed recommendations made by clinicians (Receiver Operating Characteristic accuracy of 0.94 for the WATT versus 0.86 for clinicians) [14]. However, external validity demonstrated more modest accuracy, and further development and testing of the WATT is needed especially within specific patient groups such as those with spinal conditions where it has not been evaluated [15,16].

2.4.3. Clinician recommendations

The dataset also contained information on the treatment recommendations made by clinicians conducting the assessments of workers with spinal conditions. These clinicians were trained physical therapists, occupational therapists or exercise therapists using a standardized battery of clinical tools that was consistent across workers as it was conducted in the context of a formal functional capacity evaluation. Therapists could recommend any rehabilitation program available within the jurisdiction (see Table 1) based on their clinical assessment findings. The clinicians did not have access to recommendations of the WATT, but they completed and scored the ÖMPQ as part of their standardized battery of assessments. However, they were not trained on using the ÖMPQ to inform treatment decisions but used it for prediction purposes only. Additionally, they made decisions using information from the entire clinical encounter and battery of assessments performed, of which the ÖMPQ was only one tool.

2.4.4. ‘Matching’ strategy

For each of the decision making strategies (WATT, ÖMPQ, clinician recommendations) we considered that the recommendation and the intervention ‘matched’ if a low risk intervention (single service interventions and no treatment) was recommended and actually undertaken, if a moderate risk intervention (functional restoration program, workplace-based intervention, hybrid functional restoration/workplace-based intervention) was recommended and actually undertaken, or if a high risk intervention (complex, psychological) was recommended and actually undertaken. We considered
that the recommendation and intervention undertaken were ‘unmatched’ if the recommendation and intervention undertaken were not in the same risk level (i.e., high risk intervention was recommended but low risk intervention undertaken).

2.5. Outcome

Our reference criterion was based on whether the intervention actually undertaken led to a successful complete or partial return to work. Return to work was measured using a surrogate indicator, reception of wage replacement benefits, and was available on 100% of our sample. Workers who were not receiving a full day of wage replacement benefits 30 days following assessment were considered to have returned to work. This was our main indicator of treatment success since wage replacement benefits are a commonly used measure within workers’ compensation jurisdictions and was used to evaluate the WATT tool previously [14,21]. We also considered other potential wage replacement outcomes including whether the workers received wage replacement benefits (full or partial benefits) for up to 180 days after assessment in the follow-up year.

2.6. Analysis

The sample was described using appropriate descriptive statistics for nominal, ordinal, and continuous variables. We used available demographic data to describe the patients with spinal conditions and a complete ÖMPQ score.

To assess the matching of risk-based recommended treatments to those actually undertaken versus whether that treatment was successful, we used the contingency table shown in Table 2. Columns reported whether the patients had successfully returned to work or not within 30 days and the rows reported the frequencies of whether the actual treatment received matched the recommended treatment or not. Sensitivity was calculated as the number of cases where the recommendation and the actual intervention matched and led to a successful return to work divided by the total number of cases with successful return to work. Specificity was calculated as the number of cases where the recommendation and the intervention mismatched and did not lead to a successful return to work divided by the total number of cases without successful return to work. Sensitivity and specificity were compared between decision tools and clinician recommendations using 95% confidence intervals (95% CI). The overall accuracy for each recommendation was calculated as the number of patients correctly classified divided by the total number of patients. Receiver Operating Characteristic (ROC) values were also calculated. These parameters were calculated overall for the ÖMPQ, WATT, and clinician recommendations, as well as separately within each risk level.

Our main comparison between recommendation sources was based on the positive likelihood ratio (LR+). LR+ was considered useful for identifying how the decision-making tools improved the initial probability of return to work (i.e., prevalence). It is interpreted as how importantly does following a recommendation change the initial probability (prevalence) of return to work. The size of the LR+ is interpreted as follows: 1 no difference, 2 small, 5 moderate, 10 large [22]. In addition, this parameter takes into account both sensitivity and specificity. It is calculated by the formula $\text{LR}^+ = \frac{\text{sensitivity}}{\text{1-specificity}}$. We are interested in the LR+ because we want to know how much we increase the probability of return to work by using the appropriate intervention (i.e., by following a tool’s recommendation). The LR− is a secondary outcome because we are less interested in knowing the probability of return to work when we don’t follow a tool’s recommendation. The 95% CI was used to determine if the different tools had significantly different results (no superposition/overlapping).

In addition to completing analyses on the entire sample, we conducted sub-group analyses within relevant occupational and diagnostic groups. We examined results within the occupational group of ‘Trades, Transport and Equipment Operators and Related Occupations’ since this group had adequate sample size and has traditionally been at high risk of experiencing spinal conditions. We also examined results separately for workers with neck, back, non-specific, and specific diagnoses (see Supplementary Table 1 for relevant ICD9 codes representing these groups). Analyses were performed using IBM© SPSS® Statistics version 26 (Armonk, NY, USA).

Table 2

<table>
<thead>
<tr>
<th>Matched</th>
<th>Successful intervention</th>
<th>Not matched</th>
<th>Unsuccessful intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive (TP)</td>
<td>False negative (FN)</td>
<td>False positive (FP)</td>
<td>True negative (TN)</td>
</tr>
</tbody>
</table>

Sensitivity = TP/(TP + FN), Specificity = TN/(FP + TN), Overall accuracy = (TP + TN)/(TP + TN + FP + FN), Positive likelihood ratio = sensitivity/(1-specificity), Negative likelihood ratio = (1-sensitivity)/specificity.
3. Results

3.1. Patient characteristics

In the full database, 8,747 injured workers had a spinal condition. All of these had data on the clinician and WATT recommendations. However, only 2,872 (33%) had complete ÖMPQ data. In 32.8% of the cases, patients had low back pain (i.e. ICD9 diagnosis of lumbago, backache, or lumbar sprain). In 82.7% of cases, patients had a non-specific diagnosis. For those with spinal conditions and complete ÖMPQ data, descriptive characteristics are presented in Table 3.

3.2. Treatment recommendations

The tools’ and clinicians’ recommendations cross-tabulated against the level of risk assigned to the interventions actually undertaken by each worker are shown in Table 4. Clinicians rarely recommended interventions considered suitable for ‘high risk’ workers (1.4% of recommendations made), whereas the WATT rarely recommended interventions considered suitable for ‘low risk’ workers (0.2% of recommendations made). The ÖMPQ was more balanced across the three risk levels, with some differences seen between the original and WCB-Alberta classifications.
Table 4

Cross-tabulation of the risk levels identified by each source compared to levels of risk for programs actually undertaken

<table>
<thead>
<tr>
<th>Source of the recommendation</th>
<th>Recommendation categories</th>
<th>Level of risk corresponding to the rehabilitation program actually undertaken</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low risk</td>
<td>Moderate risk</td>
</tr>
<tr>
<td>WATT</td>
<td>Low risk</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Moderate risk</td>
<td>1147</td>
<td>1133</td>
</tr>
<tr>
<td></td>
<td>High risk</td>
<td>273</td>
<td>284</td>
</tr>
<tr>
<td>Clinician recommendation</td>
<td>Low risk</td>
<td>861</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>Moderate risk</td>
<td>546</td>
<td>1263</td>
</tr>
<tr>
<td></td>
<td>High risk</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>ÖMPQ (original cut-offs)</td>
<td>Low risk</td>
<td>487</td>
<td>420</td>
</tr>
<tr>
<td></td>
<td>Moderate risk</td>
<td>328</td>
<td>373</td>
</tr>
<tr>
<td></td>
<td>High risk</td>
<td>611</td>
<td>624</td>
</tr>
<tr>
<td>ÖMPQ (WCB-Alberta cut-offs)</td>
<td>Low risk</td>
<td>793</td>
<td>771</td>
</tr>
<tr>
<td></td>
<td>Moderate risk</td>
<td>474</td>
<td>511</td>
</tr>
<tr>
<td></td>
<td>High risk</td>
<td>159</td>
<td>135</td>
</tr>
</tbody>
</table>

Table 5

Diagnostic accuracy statistics for whether matched recommendations/programs led to successful return to work 30 days after rehabilitation

<table>
<thead>
<tr>
<th>Tool</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive likelihood ratio</th>
<th>Negative likelihood ratio</th>
<th>Overall accuracy</th>
<th>Area under the ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician recommendation</td>
<td>0.75 (0.73–0.76)</td>
<td>0.26 (0.22–0.30)</td>
<td>1.01 (0.95–1.07)</td>
<td>0.97 (0.82–1.14)</td>
<td>0.67</td>
<td>0.50</td>
</tr>
<tr>
<td>WATT recommendation</td>
<td>0.40 (0.38–0.42)</td>
<td>0.57 (0.53–0.62)</td>
<td>0.93 (0.83–1.05)</td>
<td>1.05 (0.96–1.14)</td>
<td>0.43</td>
<td>0.49</td>
</tr>
<tr>
<td>ÖMPQ original cut-offs</td>
<td>0.33 (0.31–0.34)</td>
<td>0.78 (0.75–0.82)</td>
<td>1.51 (1.26–1.82)</td>
<td>0.86 (0.81–0.91)</td>
<td>0.40</td>
<td>0.56</td>
</tr>
<tr>
<td>ÖMPQ WCB-Alberta cut-offs</td>
<td>0.47 (0.45–0.49)</td>
<td>0.59 (0.54–0.63)</td>
<td>1.13 (1.01–1.27)</td>
<td>0.91 (0.83–0.99)</td>
<td>0.49</td>
<td>0.53</td>
</tr>
</tbody>
</table>

WATT = Work Assessment Triage Tool, ÖMPQ = Örebro Musculoskeletal Pain Questionnaire, ROC = Receiver Operating Characteristic. For each parameter the numbers represent: Value (95% confidence interval).

3.3. Tools’ accuracy

At 30 days after the assessment, all sensitivities and specificities were modest for ability to recommend interventions that lead to a successful outcome (see Table 5). The clinicians’ recommendations were the most sensitive (0.75, 95% CI 0.73–0.76) while the ÖMPQ with original cut-offs was the most specific (0.78, 95% CI 0.75–0.82). The LR+ (primary outcome) were small for each tool, but the LR+ of the ÖMPQ with the original cut-offs (1.51, 95% CI 1.26–1.82) was significantly higher than the WATT (0.93, 95% CI 0.83–1.05), the recommendations of clinicians (1.01, 95% CI 0.95–1.07) or the ÖMPQ with the alternate WCB-Alberta cut-offs (1.13, 95% CI 1.01–1.27) (Table 5). Matching interventions with low risk categories recommendations were much more accurate in predicting successful outcomes than the matching in the moderate and high risk categories. For example, for the ÖMPQ with the original cut-offs the LR+ were 3.80 (95% CI 2.53–5.7), 0.96 (95% CI 0.75–1.24), 0.16 (95% CI 0.08–0.35), respectively, for matching interventions to the low, moderate and high risk recommendation categories.

Results did not change substantially on any of the subgroup analyses with the exception of slightly improved sensitivity of clinician recommendations within the back pain group (sensitivity increased from 0.75 to 0.83). However, all Likelihood Ratios as well as the ÖMPQ and WATT results did not change meaningfully (less than 10% change).

When we defined return to work success as receiving less than 180 days of compensation in the follow up year, results were not substantially different (results not shown) however there were no longer statistically significant differences between the tools. Overall, in most of the analyses the clinicians were the most sensitive (when return to work is a success clinical recommendations often matched actual treatments undertaken) and the ÖMPQ with the original cut-offs was the most specific (when recommended treatments don’t match those undertaken, the ÖMPQ had the highest number of failures).

4. Discussion

4.1. Main findings

Previous research indicates that screening questionnaires such as the ÖMPQ may help for predicting delayed return to work [9,12,23]. In this paper, we evaluated clinical decisions made based on the ÖMPQ risk
recommendations against a computer-based decision support tool (WATT) and clinician treatment recommendations made following a thorough assessment. Matching risk recommendations to the actual treatments undertaken resulted in overall poor accuracy, however the ÖMPQ had significantly better results than the WATT and clinician recommendations. Therefore, ÖMPQ may also be helpful for selecting appropriate interventions. In fact, based on our observed likelihood ratios, following the ÖMPQ’s recommendations could increase the return to work probability and not following it could decrease the return to work probability. The original cut-off showed the best accuracy.

To our knowledge this is one of the first times that the ÖMPQ has been tested as a clinical decision support tool. Clinicians knew the ÖMPQ results and may have used it in their decisions, however the ÖMPQ alone still resulted in higher accuracy than clinician recommendations. Further research is needed to explore optimal cut-offs and what additional information or measures could improve the accuracy of treatment recommendations.

Overall, the difference between the different recommendation sources is small and the LR+ were often low. However matching treatments provided by the ÖMPQ recommendations with the original cutoffs offered significantly better prediction of treatment success than the clinician recommendations alone (current standard practice). Even if the differences between tools are low, the consequences of using the tool are important. For example, if we consider a return to work prevalence of 68.2% [24] at 30 days, following the clinicians’ recommendation would increase the rate of return to work by only 1% (95% CI = 0.5–1.8%), whereas following the ÖMPQ with the original cutoffs would improve the likelihood of return to work by 8% (95% CI = 4.7–11.4%). When applied to an entire population of workers, this may result in sizeable cost savings through earlier return to work.

4.2. Strengths and limitations

Relying on historical archived data from WCB-Alberta allowed us to achieve a large sample size, however it also had some limitations and resulted in a large amount of missing data. We had no control over what variables could be measured and were limited to the variables in the existing dataset. In addition, the WATT and the clinicians normally have 6 treatment categories but in order to make it comparable with the ÖMPQ risk levels, we merged treatment options into only 3 categories. This increases the accuracy parameters of matching the recommendations but decreases precision (it recommends a group of interventions, not a single one). Additionally, the ÖMPQ treatment classifications used have not been previously studied and different classification strategies may perform better. The generalization of results is limited due to the restricted location of the population. The analysis has been restricted only to the 33% of the patients with spinal pain who had a complete ÖMPQ, which could lead to a selection bias. On the other hand, the study has a relatively high number of participants that could improve external validity of results. Another limitation is the use of wage replacement benefits as our main outcome measure. Reception of benefits is a surrogate indicator of actual return to work and is influenced by several individual and contextual factors. However, this outcome is commonly used in studies of injured workers and is an important outcome within workers’ compensation jurisdictions. Instead of wage replacement outcomes, future prospective studies in this area could include patient-reported outcome measures that were not available in our dataset.

5. Conclusion

To conclude, it is too early to recommend the use of the ÖMPQ by clinicians as a decision support tool. However, the ÖMPQ had significantly better accuracy than the clinician recommendations or the WATT. The ÖMPQ should be tested in a randomized controlled trial comparing outcomes between current practice and when making decisions informed by the ÖMPQ.

Acknowledgments

The Workers’ Compensation Board of Alberta provided funding and assisted with data collection for the original study from which data came.

Conflict of interest

The authors have no conflict of interest to report.

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/BMR-200169.
References


