Determining the Reliability and Convergent Validity of a Return-to-Work Status Questionnaire
ABSTRACT

**Background:** In occupational rehabilitation programs, return-to-work is a key outcome measure; however, the studies either used different definitions for return-to-work or do not provide their definition. In order to provide a solution to this issue, we developed a self-report return-to-work measure.

**Objective:** We investigated the reliability and validity of a self-report return-to-work questionnaire in a cohort of workers with a work-related injury.

**Methods:** Two research assistants independently administered the baseline questionnaires and a follow-up questionnaire. The questionnaires contained work-related questions (e.g., currently working, if duties changed) that were used to create a four-category work status measure. Pain-related and a recovery questions were also asked. We obtained loss of earnings data from the compensation board. The short-term reliability and convergent validity were assessed.

**Results:** We recruited 75 workers, and 57 completed the test-re-test baseline questionnaire, and 51 completed the follow-up. The mean age was 45.4 years and 57% were female. The participants had a mixture of musculoskeletal injuries. Most were in the acute stage, but 17% of the participants were injured for more than a year. The short-term reliability of current working status had a kappa value of 0.90. Participants who were not working had higher levels of pain-related disability than those who were working. The kappa value for the agreement between self-reported working status and administrative data on receiving any loss of earnings payment was around 0.65.

**Conclusions:** Our study provides evidence of reliability and validity for a new return-to-work measure.

**Keywords:** return-to-work; occupational injury; questionnaire

Keywords: Measurement properties, self-report measure, workers’ compensation data
1.0 Introduction

Return to work is an important outcome measure for assessing the effectiveness of occupational rehabilitation programs. To our knowledge, there is no standard way to measure return-to-work. This is may be the result of the different definitions that have been applied to represent the concept of returning to work. Several studies have used a dichotomous definition of either returned to work versus not returned to work, or working versus not working to define it [1, 2]. Other studies defined returned-to-work based on whether the injured individual’s working status met certain criteria such as returned to the same job and same employer, returned to full duties and full hours, or returned for at least 60 consecutive days[3-6]. Finally, there are studies that have a precise definition of return-to-work based on the occupational legislation in their countries. For example, Dutch studies typically define that the worker has returned if they have “return to work in own or equal work with equal pay, for at least four weeks” [7, 8].

Return-to-work has typically been measured using either self-report or administrative data. Self-report definitions are usually vague and prone to misclassification bias [9, 10]. The studies that used administrative databases had clearly defined definitions, but the definitions vary from date of claim closure to the number of compensated lost-work days [11-13]. Krause et al. [1999] determined that using different administrative definitions from workers’ compensation databases for return-to-work has an impact on the mean number of disability days, which varied across studies from 75 to 337 days[9].

Regardless of the data source, there is a need to determine the reliability and validity of a return-to-work outcome measure. Therefore, our objective is to investigate the measurement properties of a new self-report return-to-work questionnaire in a cohort of workers who were seeking rehabilitation care for their musculoskeletal-related (MSK) occupational injury. We specifically assess the short-term test-retest reliability and the convergent validity of our questionnaire.

2.0 Methods

2.1 Setting

Workers with an injury were recruited from nine private clinics belonging to a rehabilitation company located in the greater Toronto area of Ontario, Canada. The clinics assess and treat compensation board claimants, traffic-injury
claimants and other types of clients who have injuries that require outpatient rehabilitation. Every clinic has at least one physiotherapist and one kinesiologist on staff. Chiropractic care and massage therapy were available at some of the clinics.

2.2 Inclusion and Exclusion Criteria

Workers with an injury were eligible for the study if they attended one of the clinics, were able to comprehend English, were at least 18 years old, were actively receiving treatment for their work-related injury, and were receiving compensation benefits from the sole compensation board in the province. We did not have an inclusion criterion for time since injury, as we were interested in testing our measures in clients with varying days since injury. Workers with an injury who failed to provide informed consent were excluded from the study.

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (University Health Network’s Research Ethics Board) and with the Helsinki Declaration of 1975, as revised in 2000. The study received ethics approval from the University Health Network. Informed consent was obtained from all participants.

2.3 Recruitment and Data Collection

Research assistants employed by a teaching hospital and not affiliated with the compensation board traveled to each rehabilitation clinic. The clinic staff identified the compensation board’s clients (i.e., their rehabilitation care was being paid by the compensation board) and asked their permission to introduce them to the research assistant. The research assistant explained the study, assessed eligibility, obtained informed consent, administered the initial baseline questionnaire (B1) and collected the compensation’s board identification number of the participant (figure 1). At least three to five days after enrolling in the study, another research assistant who was blinded to the participant’s initial baseline results (B1) telephoned the participant to administer the second baseline questionnaire (B2). Approximately six weeks later, the follow-up questionnaire (F) was administered by telephone. The research assistant did not have access to the participant’s previous questionnaire(s) (B1 and B2) during the follow-up interview.
2.4 Study Measures

Three questionnaire booklets were used in this study: first baseline (B1), second baseline (B2) and follow-up (F). The booklets differed slightly with respect to what questionnaires/questions were included (table 1).

2.4.1 Return-to-work questionnaire

The baseline and follow-up return-to-work questionnaires were tested. They were developed by the researchers based on a systematic review of the different return-to-work definitions used in the literature and conducting a variety of focus groups with clinicians and discussions with the compensation board. The developed questionnaires were pilot-tested in a large hospital rehabilitation center for over a year prior to initiating this study. During the first couple of months, the research assistants administering the questionnaire provided us with feedback with regards to unclear questions/response categories, and they were modified accordingly.

The baseline questionnaire has 24-items and elicits the following information: date of injury, days off because of the injury, current work status, reasons for not working and claim-related questions. The follow-up questionnaire is composed of 12-items addressing current working status, expectations for returning to work and reasons for not working. Nine of the items in the follow-up questionnaire are the same as the baseline questionnaire and five of them are used to generate the return-to-work status (Figure 1). The participant’s return-to-work status could be defined as working with no modifications, working with modifications, tried to go back to work, in labor market retraining (LMR), gone back to school, or not working. However, for this study, we had too few participants who stated that they had been enrolled in LMR, and we did not have any participants who stated they had gone back to school. Therefore, we decided to include these participants in the “not working” category rather than keep them in their own category.

2.4.2 Current Work Status Question

The last question on each questionnaire asked the participant to classify their current working status. The question was as follows: “Would you classify yourself currently as….” with the following response categories: 1) working at the same level (time, duties) as before the injury; 2) working not at the same level (time, duties) as before the injury;
3) tried to go back to work; 4) not working at all; or 5) in labor market re-entry or in school. This question was used to determine the internal reliability of the questionnaire.

2.4.3 Pain Intensity

As the participants could have been receiving treatment for different injuries involving different body part(s), we decided not to have separate pain intensity questions for each body part. Therefore, the average self-reported pain in the past week from their injury, or injuries, was scored on a 0 to 10 numerical rating scale. A score of “0” meant they had no pain at all and a score of “10” meant pain as bad as could be [14]. This pain intensity scale has been found to have excellent psychometric properties [15].

2.4.4 Pain Disability Index

The pain disability index (PDI) is a seven-item questionnaire, which measures the effect of an individual’s perceived pain on their daily living activities on a 0 to 10 numerical rating scale [16]. The score can range from 0 to 70, where higher scores represent more perceived pain. If one question was missing in the PDI, the missing value was replaced with the participant’s average score across the six other questions. If more than one question was missing, a score could not be calculated. The PDI has been found to be reliable, valid and sensitive to change. The internal consistency is 0.87 and the test-retest correlation at one week is 0.91 [17]. The pain disability score can be dichotomized as follows: high disability, a score of 44 or more and low disability, a score of less than 44 [16].

2.4.5 Global perceived recovery question

The participants were asked to rate their response to how well they felt they were recovering from their injuries since the last interview using the seven ordinal response categories of much better, much improved, slightly improved, no change, slightly worse, much worse and worse than ever. This question has been used to measure recovery in randomized clinical trials and in cohort studies of musculoskeletal disorders [18, 19]. It has been found to have excellent short-term test-retest reliability [20] and criterion validity [21].

2.5 Compensation board data
We obtained compensation board data for participants who had a lost time claim (i.e., occupational-injury/disease that resulted in either being off work for more than one day, loss of wages, or permanent disability/impairment). The following information was extracted from the administrative databases: part of body and nature of injury codes, occupation, loss of earnings, and labor market re-entry (LMR) expenditures. The injury was described using the CSA (Canadian Standards Association) part of body codes and nature of injury description codes. The occupation at the time of the injury was coded using the National Occupational Classification (NOC) system, which classifies occupations based on their skill level and skill type[22]. We reported the major group of the NOC system.

2.6 Data Analysis

2.6.1 Test-retest reliability

Short-term reliability was calculated by the Kappa statistic and its 95% confidence interval using our baseline administrations of the questionnaire (i.e., B1 and B2). This analysis was restricted to participants who reported no change in recovery at baseline, and those who had no change in baseline pain intensity between the two administrations of the questionnaire. We also calculated the level of agreement and percentage of agreement for the full sample. We had hypothesized that participants who reported the same self-reported recovery category or pain intensity within one point difference of each other would have short-term reliability between good to excellent (i.e., kappa >0.80).

2.6.2 Internal reliability

The correlation between the derived current working status measure and the single question on work status was calculated for each questionnaire (i.e., B1, B2 and F). Spearman correlation coefficients with Fisher’s Z transformation 95% confidence intervals were calculated. We hypothesized that the two measures would be highly correlated with each other.

2.6.3 Convergent Validity

Typically, convergent validity is tested against an external measure that has shown to be valid in the same population. Given the study cohort was composed of a heterogeneous group of musculoskeletal injuries, we were unable to use a site-specific functional measure and therefore used the generic PDI. For each category of the two working status variables, the mean and 95% confidence intervals of the PDI was calculated. We hypothesized that
the not working group would have higher PDI scores than the other groups because they may not be working because of their pain-related disability.

The PDI can discriminate between individuals who have low pain-related disability from those with high pain-related disability[23] using the cut-point of 44. We hypothesized that the participants who were classified as working would have on average a PDI score less than 44 and those who were not working their PDI scores on average would be greater than 44. In order to examine this hypothesis, the derived current working status and the single work status question were both dichotomized into not working or working. We conducted one-sample t tests to determine if the PDI scores for the two groups of participants was either less than or greater than 44.

We also examined whether participants were receiving any loss of earnings payment at the time of the first baseline interview (B1) and at the time of the follow-up interview (F). If the participant received any amount of money from the compensation board during the month prior to the baseline interview, we considered them as currently receiving loss of earnings payments. For the follow-up questionnaire, participants who received at least one payment were considered as receiving loss of earnings. We calculated the agreement between receiving loss of earnings payments and the dichotomous working status variables. We hypothesized good agreement between the two measures.

2.7 Sample Size Calculations
We conducted two sample size calculations. The first was for the test-retest reliability of the baseline return-to-work. We assumed that 20% of the participants were working at some level of capacity at the time of enrollment. Using a two-tailed test at the 90% level of power, we would need to recruit 14 workers with an injury to detect a kappa value of 0.80[24].

The second calculation was to determine convergent validity. We assumed a null kappa value of 0 and decided that we would like to detect a kappa value of 0.6, which is reflective of our hypothesis of good agreement. Using a two-tailed test at 90% power, we needed to recruit 30 workers with an injury. The sample size was increased to 68 workers with an injury to account for those who would decline to have their compensation board’s data linked to their questionnaire data, and those who could not be reached at the six-week follow-up interview.
3.0 Results

3.1 Study flow and characteristics of the participants

One hundred and two workers with an injury were invited to participate in the study and 75(73.5%) consented to participate (figure 2). Two participants had their injury claim subsequently denied by the compensation board. In addition, we were unable to obtain compensation board data for three participants because of missing or inaccurate claim numbers. These five participants were excluded from any analysis that involved data from the compensation board.

The mean age was 45.4 years and 57% were female (table 2). More than 50% of the participants had completed more than high school education. Most participants were working full-time at the time of their injury. At the time of the baseline questionnaire, 44% had been off work for three months or less, 39% had been off for four to 12 months and 17% had been off work for more than one year. The majority of the injuries were a result of a sprain or strain, or soft-tissue tear, and the injuries predominately occurred in the back or in the upper extremities.

Fifty-one participants (68%) (figure 1) completed the follow-up questionnaire on average 86.3 days after their initial baseline questionnaire. The amount of disability associated with pain decreased significantly from baseline (mean = -8.1 95% CI (-12.6 to -3.6)), and there was a significant decrease in pain intensity (mean = -1.0 95% CI (-1.7 to -0.3)). Twenty-one participants reported being either much improved or completely better from their work-related injury at their follow-up interview.

3.2 Test re-test Reliability

The B2 questionnaire was administered to 76% of the participants (n = 57) (figure 1) on average 8.3 days after B1 questionnaire. Twenty-seven (48.2%) participants reported no change in their perceived level of recovery from their work-related injury and 12 (21.4%) reported either much improved or completely better. The percentage agreement between B1 and B2 for the generated work status was 92.6% and the level of agreement was kappa = 0.90 (95% CI 0.76 to 1.00) (table 3). We collapsed the generated work status down to working (n=16 (59.3%)) versus not (n=11 (40.7%)) and found perfect agreement between B1 and B2 with a kappa value of 1.0. Similarly, the collapsed single work status question yielded perfect agreement and a kappa value of 1.0, but the four nominal-level questions had lower percentage agreement and kappa (81% and 0.71, respectively) than the generated work status.
The pain intensity did not change between B1 and B2 for 16 (28.1%) participants (table 3). The agreement for the generated working status had a kappa value of 0.7. The single work status question had the same percentage of agreement (81%), but the kappa statistic could not be calculated because of an empty cell. Collapsing both the generated and single work status question down to working (n=10 (63%)) versus not (n=6 (37%)) resulted in perfect agreement for each of them.

The agreement for the generated working status between B1 and B2 for the entire sample was lower than for the subset of participants who reported no change in their condition between the two administrations (kappa = 0.63; 95% CI 0.40 0.80). This lower level of agreement is likely the result of changing working status, given that 11 out of 15 (73.3%) improved their working status level between B1 and B2.

3.3 Internal Reliability

There was excellent internal reliability of the generated work status against the single work status question for the three questionnaires (all correlations > 0.8) (table 4). There were two patterns of discrepancies between these two. The first pattern was whether or not the participants considered themselves as working with or without modifications. Participants were defined as working at the same level using the generated work status, but considered themselves as working with modifications using the single work status question (B1: 8(11%); B2: 5(9%); F: 7(14%)). Conversely, participants were defined as working not at the same level using the generated work status, but were defined as working with no modifications using the single question (B1: 2(3%); B2: 4(7%); F: 2(4%).

The second pattern was whether or not the participant was classified as either “tried to go back to work” or “not working.” Participants were defined as tried to go back to work based on the generated working status variable, but classified themselves as not working using the single question (B1: 8(11%); B2: 9(16%); F: 6(12)). The reverse was true for two participants (3%) for B1. There were no participants who were classified as not working using the generated working status, but they considered themselves to have tried to go back to work using the single question for B2 and F.
3.4 Convergent Validity

On average, the participants reported some pain disability associated with their daily activities at baseline. Participants who were working had significantly lower PDI scores than those classified as not working (table 5) and on average, their PDI score was less than 44, which indicates less disability. At the follow-up interview, the PDI scores were slightly lower than the baseline scores, which meant, on average, they were having less pain-related disability than at baseline. Those working at the follow-up interview had PDI scores less than 44 on average.

We obtained compensation board data on 70 participants from the date of injury until one month after their last interview. At the baseline interview, 35 participants had received loss of earnings payments at least once after their workplace injury. Of these participants, 32 (91.4%) reported they had one day or more off work. There were 20 participants who were classified as “no loss time claims” by the compensation board. Five of these participants (25.0%) reported they had taken at least one day off work for their work-related injury, but it is unknown if the workers had these days covered by their employers. At the time of the B1 interview, there were 23 participants (32.9%) currently receiving loss of earning payments from the compensation board and 18 of them (78.3%) self-reported they were not working. Of the 47 participants who were not currently receiving loss of earnings payments, 41 participants (87.2%) self-reported they were working. The level of agreement between baseline derived current work status and receiving LOE benefits from the compensation board had a kappa value of 0.65 (95% CI 0.46 to 0.84). At the time of the follow-up interview, there were 16 participants (31.4%) who were receiving loss of earnings payments, and of these, 12 self-reported they were currently not working. The level of agreement was again similar (kappa =0.64 (95% CI 0.41 to 0.87)).

4.0 Discussion

This study assessed the reliability and validity of a working status questionnaire, which we developed as a result of a literature search, conducting interviews with various stakeholders and pilot testing at a large university hospital rehabilitation program. We found adequate short-term reliability of current working status that was derived using five questions, as hypothesized. Collapsing working status down to working versus not working resulted in perfect agreement. Participants’ generated working status were associated with their amount of pain-related disability. As hypothesized, we found that working participants had less disability than the participants who were not working.
We also found a kappa value of around 0.65 between working status measures and receiving any loss of earnings payment from the compensation’s board. Using the compensation’s board data, we were unable to derive a four-level return to work status variable, as the payments were either stable over the long-term, or fluctuated severely within 30-day periods.

To our knowledge this is the first study that has examined short-term reliability using the test-retest method on a four-level working status measure. Previous studies have examined test-retest reliability of whether workers with an injury had taken any time off as the result of a musculoskeletal complaint and the number of sick days[25, 26]. Both of these studies also found high levels of agreements, but their source populations were different. They were examining any sickness absences during a time-period in an occupational cohort rather than examining their current working status in an occupational cohort who had experienced a work-related injury.

We used two different anchors for defining a ‘stable’ subset of the study participants to assess the short-term reliability. Although these anchors were not directly related to current work status, they both measure aspects of the participant’s life that reflect an overall change in health status, and would ultimately be reflected in a change in their working status. Comparing the two anchors, we found slightly different subsets of participants, with 30% belonging to both subsets. The minimal clinical meaningful difference for pain measured on an 11-point numerical rating scale is slightly over one-point for acute patients and two points for chronic patients[27, 28]. Our cohort was composed of a mixture of participants who had been recently injured and others who were injured more than three months ago. Changing the definition of stable to a change in pain of less than two points, we found a 51% overlap between the two subsets. However, the level of agreement between the working status and the revised ‘stable’ pain group dropped to moderate levels.

Participants who self-reported they had returned to work reported lower functional disability, which was expected for convergent validity. Similar findings were found in another Ontario cohort of workers with back or upper extremity MSK disorders. [29]. The Roland Morris disability questionnaire, a measure for determining functional disability in the lower back, was significantly higher for the workers with injuries who reported not working or had fluctuating periods of working, compared to workers with injuries who successfully returned to work one-month.
post injury. In addition, those who successfully returned to work reported higher physical and mental functioning according to the MOS Short Form 12 (SF12) and lower depressive symptomology.

Our study found agreement between the self-reported working status and loss of earnings to be around 0.65 (kappa value). It is important to note that administrative outcome measures of return-to-work are proxy measures because the termination of benefits may not necessarily mean returned to work. For example, benefits might be terminated if a worker does not comply with recommendations, such as attending rehabilitation or other return-to-work activities. Moreover, the termination of benefits might not mean that the worker has recovered from his/her injuries[30]. A study of 205 workers with musculoskeletal injuries in the United States found that workers, who returned to full duties according to termination of compensation benefits, reported some form of disability for at least six months after the injury[31]. On the other hand, employers may not pay the full salary of a worker after returning to work because they are unable to perform all of their employment duties or cannot work the full workday. As the result of this, compensation boards may provide partial loss of earnings to a worker.

Our study has several strengths. The generalizability of our results should be excellent because we recruited from community-based clinics that see the typical worker with work-related injuries in the province. Also, participants had different types of injuries and varying times since injury. Finally, we pilot-tested the survey in a different setting for a very long period of time where it went under a couple of revisions during the first few months.

Our study has some limitations. The length of time between administrations of the two baseline questionnaires was on average longer than three to five days. We compensated for this by excluding workers who reported a change in recovery during this time. Also, we were unable to derive all of the categories in our working status measure as fewer than five workers stated they were in the labor market re-entry program. This program is typically offered to workers who have been off work for more than one-year, which would apply to approximately 17% of our cohort members.

5.0 Conclusion
Our new return-to-work questionnaire has adequate short-term reliability and convergent validity with overall pain disability. Future studies should incorporate a self-reported return-to-work measure as an outcome measure because we found it to be reliable and valid.

Conflict of Interest
All authors declare no conflict of interests.

Acknowledgement
[Edited for the review process]
References


Table 1: Overview of when the questionnaires were administered and their purpose
<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Time point of administration</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First baseline (B1)</td>
<td>Test retest</td>
</tr>
<tr>
<td></td>
<td>Second baseline (B2)</td>
<td>B1 and B2</td>
</tr>
<tr>
<td></td>
<td>Follow-up (F)</td>
<td></td>
</tr>
<tr>
<td>Return-to-work</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pain intensity</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Pain disability index</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Self-reported recovery</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Single work status question</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Notes: B1 = first baseline questionnaire; B2 = second baseline questionnaire; and F = follow-up questionnaire
Table 2: Characteristics of the participants
<table>
<thead>
<tr>
<th>Variable</th>
<th>B1 interview (N=75)</th>
<th>B2 interview (N=57)</th>
<th>F interview (N = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Age in years: mean (std)</td>
<td>45.4 (11.0)</td>
<td>46.9 (10.7)</td>
<td>47.4 (10.7)</td>
</tr>
<tr>
<td>Female</td>
<td>43 (57.3%)</td>
<td>33 (57.9%)</td>
<td>32 (62.8%)</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/common-law</td>
<td>49 (65.3%)</td>
<td>36 (63.2%)</td>
<td>33 (64.7%)</td>
</tr>
<tr>
<td>Single (never married)</td>
<td>13 (17.3%)</td>
<td>9 (15.8%)</td>
<td>8 (15.7%)</td>
</tr>
<tr>
<td>Separated/divorced</td>
<td>13 (17.3%)</td>
<td>12 (21.1%)</td>
<td>10 (19.6%)</td>
</tr>
<tr>
<td>Highest education level (n=60; n=46; n=40)a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some high school</td>
<td>8 (12.9%)</td>
<td>7 (15.2%)</td>
<td>4 (10.0%)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>14 (22.6%)</td>
<td>10 (21.7%)</td>
<td>7 (17.5%)</td>
</tr>
<tr>
<td>Some/completed trade school</td>
<td>14 (22.6%)</td>
<td>9 (19.6%)</td>
<td>10 (25.0%)</td>
</tr>
<tr>
<td>Community college</td>
<td>26 (41.9%)</td>
<td>20 (43.5%)</td>
<td>19 (47.5%)</td>
</tr>
<tr>
<td>Immigrated to Canada</td>
<td>29 (38.7%)</td>
<td>22 (38.6%)</td>
<td>21 (41.2%)</td>
</tr>
<tr>
<td>Does not speak English at home</td>
<td>10 (13.3%)</td>
<td>8 (14.0%)</td>
<td>6 (11.8%)</td>
</tr>
<tr>
<td>Current occupation (n=69; n=51; n=45)b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business, finance and administration</td>
<td>6 (8.7%)</td>
<td>5 (9.8%)</td>
<td>5 (11.1%)</td>
</tr>
<tr>
<td>Health</td>
<td>5 (7.3%)</td>
<td>3 (5.9%)</td>
<td>3 (6.7%)</td>
</tr>
<tr>
<td>Processing, manufacturing and utilities</td>
<td>7 (10.1%)</td>
<td>6 (11.8%)</td>
<td>5 (11.1%)</td>
</tr>
<tr>
<td>Sales and service</td>
<td>13 (18.8%)</td>
<td>8 (15.7%)</td>
<td>10 (22.2%)</td>
</tr>
<tr>
<td>Trades, transport, equipment operators and</td>
<td>13 (18.8%)</td>
<td>10 (19.6%)</td>
<td>8 (17.8%)</td>
</tr>
<tr>
<td>related occupations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5 (7.3%)</td>
<td>4 (7.8%)</td>
<td>3 (6.7%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>20 (29.0%)</td>
<td>15 (29.4%)</td>
<td>11 (24.4%)</td>
</tr>
<tr>
<td>Work status prior to injury</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>64 (85.3%)</td>
<td>49 (86.0%)</td>
<td>43 (84.3%)</td>
</tr>
<tr>
<td>Part-time</td>
<td>9 (12.0%)</td>
<td>7 (12.3%)</td>
<td>7 (13.7%)</td>
</tr>
<tr>
<td>Temporary/seasonal or other</td>
<td>2 (2.7%)</td>
<td>1 (1.8%)</td>
<td>1 (2.0%)</td>
</tr>
<tr>
<td>Time since injury in months: mean(std) (n=69; n=52; n=45)b</td>
<td>9.7 (17.5)</td>
<td>9.4 (16.4)</td>
<td>10.3 (17.3)</td>
</tr>
<tr>
<td>0 to 3 months</td>
<td>30 (43.5%)</td>
<td>24 (46.2%)</td>
<td>18 (40.0%)</td>
</tr>
<tr>
<td>4 to 12 months</td>
<td>27 (39.1%)</td>
<td>19 (36.5%)</td>
<td>18 (40.0%)</td>
</tr>
<tr>
<td>more than 12 months</td>
<td>12 (17.4%)</td>
<td>9 (17.3%)</td>
<td>9 (20.0%)</td>
</tr>
<tr>
<td>Type of injury (n = 70; n = 52; n = 46)b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fractures</td>
<td>5 (7.1%)</td>
<td>5 (9.6%)</td>
<td>5 (10.9%)</td>
</tr>
<tr>
<td>Sprains, strains, tears</td>
<td>33 (47.1%)</td>
<td>22 (42.3%)</td>
<td>21 (45.7%)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (17.1%)</td>
<td>10 (19.2%)</td>
<td>9 (19.6%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>20 (28.6%)</td>
<td>15 (28.9%)</td>
<td>11 (23.9%)</td>
</tr>
<tr>
<td>Part of body injured (n=69; n=52; n=46)b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back</td>
<td>19 (27.5%)</td>
<td>14 (26.9%)</td>
<td>14 (30.4%)</td>
</tr>
<tr>
<td>Upper extremities</td>
<td>16 (23.2%)</td>
<td>10 (19.2%)</td>
<td>10 (21.7%)</td>
</tr>
<tr>
<td>Lower extremities</td>
<td>8 (11.6%)</td>
<td>8 (15.4%)</td>
<td>8 (17.4%)</td>
</tr>
<tr>
<td>Multiple areas or abdominal</td>
<td>6 (8.7%)</td>
<td>5 (9.6%)</td>
<td>3 (6.5%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>20 (29.0%)</td>
<td>15 (28.9%)</td>
<td>11 (23.9%)</td>
</tr>
</tbody>
</table>

B1 = First baseline interview; B2 = Second baseline interview; F = Follow-up interview; std = standard deviation
a Sample sizes for the baseline, the re-administration of the baseline and the follow-up questionnaires.

b Unable to obtain data from the compensation board for five participants. Two participants had their injury claim subsequently denied and three participants had missing or inaccurate claim number.

c This information was not collected by the compensation board because the participant was classified as a "no lost time" claimant. These claimants did not miss a day of work other than the day of injury and required health care for their injury. The compensation board is responsible for paying the health care costs.

d Other was considered to be one of the following: bruises, contusions, disc herniations/sciatica, mensical tear, multiple traumatic injuries, and traumatic tendinitis.
Table 3 Test re-test reliability of working status in participants who were stable between the two administrations of the baseline questionnaire
<table>
<thead>
<tr>
<th>Generated work status</th>
<th>B1 n (%)</th>
<th>B2 n (%)</th>
<th>Percentage of agreement</th>
<th>kappa 95% CI</th>
<th>B1 n (%)</th>
<th>B2 n (%)</th>
<th>Percentage of agreement</th>
<th>kappa 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working with no modifications</td>
<td>5 (18.5%)</td>
<td>5 (18.5%)</td>
<td>92.6%</td>
<td>0.90 (0.76 1.00)</td>
<td>6 (37.5%)</td>
<td>5 (31.3%)</td>
<td>81.30%</td>
<td>0.74 (0.50 0.99)</td>
</tr>
<tr>
<td>Working with modifications</td>
<td>11 (40.7%)</td>
<td>11 (40.7%)</td>
<td>3 (18.8%)</td>
<td>5 (31.3%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tried to go back to work</td>
<td>5 (18.5%)</td>
<td>7 (25.9%)</td>
<td>3 (18.8%)</td>
<td>5 (31.3%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not working or LMR</td>
<td>6 (22.2%)</td>
<td>4 (14.8%)</td>
<td>3 (18.8%)</td>
<td>1 (6.3%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Single work status question</th>
<th>B1 n (%)</th>
<th>B2 n (%)</th>
<th>Percentage of agreement</th>
<th>kappa 95% CI</th>
<th>B1 n (%)</th>
<th>B2 n (%)</th>
<th>Percentage of agreement</th>
<th>kappa 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working at the same level</td>
<td>3 (11.1%)</td>
<td>3 (11.1%)</td>
<td>81.4%</td>
<td>0.71 (0.49 0.93)</td>
<td>1 (6.3%)</td>
<td>3 (18.8%)</td>
<td>81.30%</td>
<td>unable to calculate</td>
</tr>
<tr>
<td>Not working at the same level</td>
<td>13 (48.2%)</td>
<td>13 (48.2%)</td>
<td>9 (56.3%)</td>
<td>7 (43.8%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tried to go back to work</td>
<td>2 (7.4%)</td>
<td>1 (3.7%)</td>
<td>1 (6.3%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not working at all</td>
<td>9 (33.3%)</td>
<td>10 (37.0%)</td>
<td>5 (31.3%)</td>
<td>6 (37.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B1= first baseline interview; B2 = second baseline interview; CI = confidence interval; LMR = labor market retraining
Table 4 Internal Reliability of the Derived Current Working Status versus the Single Question for Each Questionnaire Administration
<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Working with no modifications n (%)</th>
<th>Working with modifications n (%)</th>
<th>Tried to go back to work n (%)</th>
<th>Not working n (%)</th>
<th>Correlation (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 (n=75)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generated</td>
<td>14 (18.7%)</td>
<td>35 (46.7%)</td>
<td>13 (17.3%)</td>
<td>13 (17.3%)</td>
<td>0.87 (0.81 0.92)</td>
</tr>
<tr>
<td>Single question</td>
<td>8 (10.7%)</td>
<td>41 (54.7%)</td>
<td>7 (9.3%)</td>
<td>19 (25.3%)</td>
<td></td>
</tr>
<tr>
<td>B2 (n = 56 a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generated</td>
<td>12 (21.4%)</td>
<td>27 (48.2%)</td>
<td>11 (19.6%)</td>
<td>6 (10.7%)</td>
<td>0.83 (0.72 0.90)</td>
</tr>
<tr>
<td>Single question</td>
<td>12 (21.4%)</td>
<td>28 (50.0%)</td>
<td>1 (1.8%)</td>
<td>15 (26.8%)</td>
<td></td>
</tr>
<tr>
<td>F (n=51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generated</td>
<td>16 (31.4%)</td>
<td>19 (37.3%)</td>
<td>8 (15.7%)</td>
<td>8 (15.7%)</td>
<td>0.89 (0.82 0.94)</td>
</tr>
<tr>
<td>Single question</td>
<td>11 (21.6%)</td>
<td>24 (47.1%)</td>
<td>2 (3.9%)</td>
<td>14 (27.5%)</td>
<td></td>
</tr>
</tbody>
</table>

B1 = first baseline interview; B2 = second baseline interview; F = follow-up interview; CI = confidence interval

* One participant did not complete the single work status question at their B2 interview and therefore, the sample size has been reduced to 56.
Table 5 Convergent Validity of the Work Status Variable Against Pain Disability Index (PDI) for the Baseline and Follow-up Questionnaires
<table>
<thead>
<tr>
<th>Derived current working status</th>
<th>B1 PDI score</th>
<th>F PDI score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean (95% CI)</td>
</tr>
<tr>
<td>Entire sample</td>
<td>75</td>
<td>36.6 (32.8 40.4)</td>
</tr>
<tr>
<td>Derived current working status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working with no modifications</td>
<td>14</td>
<td>29.4 (20.5 38.4)</td>
</tr>
<tr>
<td>Working with modifications</td>
<td>35</td>
<td>32.3 (27.0 37.5)</td>
</tr>
<tr>
<td>Tried to go back to work</td>
<td>13</td>
<td>47.8 (40.9 54.6)</td>
</tr>
<tr>
<td>Not working or LMR</td>
<td>13</td>
<td>44.8 (34.5 55.1)</td>
</tr>
<tr>
<td>Working</td>
<td>49</td>
<td>31.5 (27.1 35.8)</td>
</tr>
<tr>
<td>Not working</td>
<td>26</td>
<td>46.3 (40.5 52.0)</td>
</tr>
<tr>
<td>Single work status question</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working with no modifications</td>
<td>8</td>
<td>21.9 (10.3 33.6)</td>
</tr>
<tr>
<td>Working with modifications</td>
<td>41</td>
<td>33.3 (28.6 38.0)</td>
</tr>
<tr>
<td>Tried to go back to work</td>
<td>7</td>
<td>48.4 (36.0 60.9)</td>
</tr>
<tr>
<td>Not working or LMR</td>
<td>19</td>
<td>45.5 (38.3 52.6)</td>
</tr>
<tr>
<td>Working</td>
<td>49</td>
<td>31.5 (27.1 35.8)</td>
</tr>
<tr>
<td>Not working</td>
<td>26</td>
<td>46.3 (40.5 52.0)</td>
</tr>
</tbody>
</table>

B1 = first baseline interview; PDI = pain disability index; F = follow-up interview; CI = confidence interval; LMR = labor market retraining

- Mean PDI was significantly less than 44, which indicated that the participants had less disability
- Mean PDI was not significantly more than 44, which indicated that the participants did not have high disability
- One participant was missing two responses for the pain disability index and therefore, the questionnaire could not be scored.
- Two participants classified themselves as tried to go back to work and therefore, estimates are unstable.
Figure 1: Return-to-work questions and the algorithms used to generate the work status at each interview
Baseline return-to-work status questions and algorithm to generate work status

1. Have you tried to go back to work since your injury?
   - No
   - Yes

2. Are you currently working?
   - Yes
   - No
   **If no, please go to question 7**

3. Have your work hours been reduced because of your injury?
   - No
   - Yes
   - Do not know

4. What are your duties?
   - Regular duties
   - Modified Duties

5. Are you working at the same job with the same employer that you had prior to your injury?
   - No
   - Yes

6. Have your duties changed because of your injury?
   - No
   - Do not know
   - Yes

7. Have you had any job retraining or labour market retraining (LMR) since the injury?
   - Yes
   - No

8. Have you gone back to school since the injury?
   - Yes
   - No

Follow-up return-to-work questions and algorithm to generate work status

1. Are you currently working?
   - Yes
   - No
   **If no, please go to question 6**

2. Have your work hours been reduced because of your injury?
   - No
   - Yes
   - Do not know

3. What are your duties?
   - Regular duties
   - Modified Duties

4. Are you working at the same job with the same employer that you had prior to your injury?
   - No
   - Yes

5. Have your duties changed because of your injury?
   - No
   - Do not know
   - Yes

6. Why are you currently not working? (Check all relevant boxes)
   - Employer doesn’t want me
   - Family responsibilities
   - Employer no longer exists
   - Other illnesses
   - Gone back to school
   - Retired
   - Job-related training/re-training
   - Unable to work because of the injury
   - Laid off
   - Other (please specify)

7. Have you attempted to return to work since our first interview with you?
   - Yes
   - No

Working with no modifications
- Answered YES to question 2; answered NO to question 3; answered REGULAR DUTIES to question 4

Working with modifications
- Answered YES to question 2; either answered YES or DO NOT KNOW to questions 3 or 6 or answered MODIFIED DUTIES to question 4

Tried to go back to work
- Answered YES to question 1; answered NO to question 2
- In labour market retraining or gone back to school
- Answered NO to question 2; Answered YES to either question 7 or 8
- Not working
- Answered NO to question 1; answered NO to question 2

Working with no modifications
- Answered YES to question 1; answered NO to question 2; answered REGULAR DUTIES to question 3

Working with modifications
- Answered YES to question 1; answered either YES or DO NOT KNOW to questions 2 or 5 or answered MODIFIED DUTIES to question 3

Tried to go back to work
- Answered NO to question 1; answered YES to question 7
- In labour market retraining or gone back to school
- Answered NO to question 1; answered either JOB-RELATED TRAINING/RE-TRAINING or GONE BACK TO SCHOOL to question 6
- Not working
- Answered NO to question 1; answered NO to question 7
Figure 2 Participant’s flow diagram
Assessed for eligibility
N = 102

Excluded n = 27
- Not interested in the study (n= 8)
- Concerns with the compensation board (n=5)
- No time to participate (n=4)
- Check the letter of consent with a family member (n=4)
- Language difficulties (n=4)
- No phone (n=1)
- Other reason (n=1)

First baseline interview
n = 75

Second baseline interview
n = 57

Follow-up interview
n = 51