Send Orders for Reprints to reprints@benthamscience.net

Development and Initial Validation of the Satisfaction and Recovery Index (SRI) for Measurement of Recovery from Musculoskeletal Trauma

David M. Walton[*,1], Joy C. MacDermid[2,3], Mathew Pulickal[1], Amber Rollack[1] and Jennifer Veitch[1]

[1]*School of Physical Therapy, Western University, London Ontario, Canada*

[2]*School of Physical Therapy, McMaster University, Hamilton Ontario, Canada*

[3]*Clinical Research Lab, Hand and Upper Limb Centre, St. Joseph's Hospital, London Ontario, Canada*

Abstract: *Background*: There is a need for a generic patient-reported outcome (PRO) that is patient-centric and offers sound properties for measuring the process and state of recovery from musculoskeletal trauma. This study describes the construction and initial validation of a new tool for this purpose.

Methods: A prototype tool was constructed through input of academic and clinical experts and patient representatives. After evaluation of individual items, a 9-item Satisfaction and Recovery Index (SRI) was subject to psychometric evaluation drawn from classical test theory. Subjects were recruited through online and clinical populations, from those reporting pain or disability from musculoskeletal trauma. The full sample (N = 129) completed the prototype tool and a corresponding region-specific disability measure. A subsample (N = 46) also completed the Short-Form 12 version 2 (SF12vs). Of that, a second subsample (N = 29) repeated all measures 3 months later.

Results: A single factor 'health-related satisfaction' was extracted that explained 71.1% of scale variance, Cronbach's alpha = 0.95. A priori hypotheses for cross-sectional correlations with region-specific disability measures and the generic Short-form 12 component scores were supported. The SRI tool was equally responsive to change, and able to discriminate between recovered/non-recovered subjects, at a level similar to that of the region-specific measures and generally better than the SF-12 subscales.

Conclusion: The new SRI tool, as a measure of health-related satisfaction, shows promise in this initial evaluation of its properties. It is generic, patient-centered, and shows overall measurement properties similar to that of region-specific measures while allowing the potential benefit of comparison between clinical conditions. Despite early promising results, additional properties need to be explored before the tool can be endorsed for routine clinical use.

Keywords: Patient-reported outcomes, psychometrics, recovery, responsiveness, validity.

INTRODUCTION

Patient-reported outcomes (PROs) are recognized as a vital part of patient-oriented care [1]. Accepting that the patient's opinion is the one that matters most, PROs are often collected through use of standardized self-report tools with sound measurement properties, including adequate empirical evidence for reliability and validity. Increasingly, policy decisions and clinical behaviors are being influenced by the results of research using PROs as primary outcomes. This highlights the need for conceptually meaningful and statistically sound measurement tools.

Outcomes of rehabilitation research and clinical intervention are often collected using region-specific disability scales, from here forward referred to as Regional Measures (RM). These are commonly constructed using an item generation process to develop lists of items that address symptoms and functional impairments common to dysfunction in a particular region (e.g., pain in the neck, low back, upper or lower extremity). In many cases, scale items are chosen through consultation with clinical experts (e.g., [2,3]). Reliance on experts to generate items is more common than inclusion of patients, although some scales do both (e.g., [4]). Occasionally, items are generated by adapting those from scales intended for other regions (e.g., the Neck Disability Index [5] adapted from the Oswestry Disability Index [6]) or by collecting items from several other scales into a single aggregate measure (e.g., [7]). All approaches have merits and drawbacks, and it is difficult to endorse one method of scale development as clearly superior in all cases.

In contrast to regional measures, more generic tools have also been endorsed for clinical and research use. These include scales that focus on the functional impact of a symptom or condition, such as the Brief Pain Inventory [8] or Pain Disability Index [9]. Other scales focus even more broadly on constructs such as Health-Related Quality of Life (HRQoL, e.g. the EuroQoL 5-D [10]) or health status (e.g. the Medical Outcomes Survey Short-Form (SF) 36 [11] or

*Address correspondence to this author at the School of Physical Therapy Rm. EC1443, Western University, 1201 Western Rd., London, ON N6G 1H1, Canada; Tel: +1-519-661-2111, Ext. 80145; Fax: +1-519-661-3866; E-mail: dwalton5@uwo.ca

1874-3250/14

SF-12 [12]). Both regional and generic PROs hold value for different purposes and often demonstrate different measurement properties. Regional measures are directly relevant to a specific condition and multiple studies have demonstrated higher responsiveness than generic health status measures [13-15]. Conversely, the advantage of generic measures is that they allow for comparisons across clinical populations and hence can reflect the relative contribution of different health problems to disease burden [16, 17]. Generic measures that include preference weighting offer the added opportunity for health economic evaluations [18].

Despite the availability of many validated PROs, uptake in clinical practice is often low [19-23]. Barriers to implementation can include the length of measures, their lack of relevance to the clinical population, difficulty in obtaining copyrighted scales, or complicated scoring algorithms, amongst others [24, 25]. Many scales include specific activities that gravitate towards tasks of daily life, allowing them to be answered by the majority of respondents. In order to avoid floor effects, the items often reflect tasks of lower levels of demand and may not represent domains important to higher-functioning patients. Additionally, the underlying assumption of many such scales that are arithmetically summed is that all functional items on a scale are equally important to all respondents, thus all are given equal weighting in the total score. This assumption may result in measures that can no longer measure real change (limited range of variability), a higher burden on patients (due to non-informative items) and overall lack of validity in measuring the intended health construct. Walton *et al.* [26]. provide a more detailed discussion of these issues.

The vast majority of PROs offer standardized response options and scoring metrics. Anchors such as "no difficulty" to "unable to do", or "no pain" to "worst pain" reflect this. However, this scoring metric may not adequately reflect the extent to which an item is a concern for the patient. For example, there may be tasks that patients are unable to do but that are of no real concern to them (e.g., driving, throwing, or running). Conversely, some respondents may be able to perform a task but require strict pacing or adherence to medication regimens in order to do so, that lead to overall dissatisfaction despite the ability to perform. In prognosis research, regional or generic measures are commonly used to dichotomize outcomes as 'good' or 'bad', 'recovered' or 'not recovered' based on norm-based cut-scores inferred from previous sample means [27-29]. However, the same argument holds true: that just because a respondent *can* walk a certain distance may not necessarily indicate satisfaction with that ability, similarly, an inability to walk that same distance may be inconsequential to another respondent.

One response to the challenge of item and response relevance in standardized measures has been the use of scales with patient-generated items, such as the Patient Specific Functional Scale [30] or the Canadian Occupational Performance Measure [31]. Such tools are by definition patient-centered, and appear useful for individual case management as they can be used to set treatment goals and monitor change. As a result they have been shown to be more responsive than regional measures with standardized items [30,32]. However, a key limitation of these measures is that scores cannot be compared between patients owing to the highly individualized nature of each patient's scale, thereby posing conceptual problems for research purposes that rely on sample means. There appears to be a need for a conceptually and psychometrically sound measurement tool that is patient-centred while permitting comparison between patients and conditions.

The purpose of this study was to add to the pool of existing measures by creating and evaluating a new scale that would measure a unique construct of importance-weighted "health-related satisfaction". This construct considers different domains of recovery from musculoskeletal (MSK) trauma, and allows weighting by relative importance to the individual. The goal was to create a generic PRO that offered balance between standardization and patient-centeredness. This is meant to be broadly applicable to all patients following MSK trauma, and potentially beyond to any disorder where recovery is possible. The specific objectives in this study were to describe the following:

1. The developmental process of the tool and the content validity as determined by expert review and patient-reported item importance.
2. The factor structure of the tool.
3. The construct validity determined by comparing the new tool to established condition-relevant RMs or a common generic health-status measure, the short-form (SF)-12.
4. The comparative responsiveness of the new tool, condition-relevant RMs, and generic SF-12 to identify meaningful change following routine physiotherapy over a 3-month interval.

METHODS

Tool Construction

The construct to be measured was importance-weighted Health-Related Satisfaction (HRS). At the time of undertaking this project, HRS lacked a clear conceptual framework in the literature. For this purpose, the construct of HRS was developed through a series of focus groups with patients (N = 35, mean age 41y, 69% female) currently or recently experiencing neck pain of traumatic origin. Using a nominal question of *"How will you know when you are recovered?"* focus sessions and one-on-one interviews were conducted in both Canada and Australia, the transcripts of which were analyzed thematically to identify meta-themes that influenced the construct of 'satisfactory recovery'. The results of this work have been published elsewhere [33]. Concurrently, the authors described a conceptual framework for the notion of post-traumatic recovery that could be summarized as a 'satisfying end to the injury experience', which drew from current theories of happiness, health, and human potential. This has also been published previously [26]. Using these sources, 15 items were generated by the co-authors that tapped the new construct of HRS. The items were sent to 10 international experts in post-traumatic rehabilitation (5 countries, mean years in the field = 8) who provided feedback on the appropriateness of included items, and whether any domains were missing. This exercise led to the removal of 2 items that were not deemed adequately

important for this population. Subsequently, 6 patients of a local tertiary care centre for disorders of the upper extremity participated in a cognitive interviewing session using a 'think aloud' approach to identify problems in wording and response structure. Through this process an additional 3 items were collapsed with other items while the response structure was retained. The prototype tool therefore consisted of 10 items that all appeared to tap the construct of HRS. Each item was scored on two 11-point scales, an importance scale (0 = not important to me at all, 10 = extremely important to me) and a satisfaction scale (0 = not satisfied at all, 10 = completely satisfied). The prototype tool was reviewed by a professional technical editor to ensure grammar, spelling and format were correct. It was then translated into French using the independent forward/backward translation approach endorsed by Beaton and colleagues [34] prior to deployment.

Empirical Evaluation of Tool Properties

Data were collected in two ways. An online recruitment procedure was undertaken using advertisements through the *Google AdWords*(Google Inc.) and the *Facebook Ads*(Facebook Inc.) platforms. Clicking on an advertisement on either platform would take potential subjects to a letter of information describing the study and how their data would be used. Those who wished to participate provided their email address, to which a separate link was sent that took respondents to a secure survey platform (www.LimeService.com). Electronic consent was obtained explicitly by clicking an 'I consent to participate' radio button in order to proceed. The survey then used routing logic to screen respondents using the following inclusion/exclusion criteria: age at least 18 years, able to read and understand English or French at a minimum grade 6 level, and currently experiencing pain of traumatic origin (collision, trip, fall, awkward lift, hit by an object or person) in the neck, lower back, upper or lower extremities. In order to avoid bias especially regarding the item about future growth, those with chronic progressive comorbidities were excluded (e.g. end-stage cancer, multiple sclerosis, amyotrophic lateral sclerosis, end-stage liver, kidney or heart disease). Consenting subjects provided basic demographics (age, sex, work status prior to injury, current work status, area of body injured and time since injury). They also answered the prototype 10-item questionnaire and one of 4 RMs dependent on their area of injury (described below). Finally, questions meant to indicate general recovery status were answered: frequency of use of analgesic medications for injury-related symptoms over the past week, current indemnity benefits status, current legal status, currently require care for their injury, and finally an 11-point Numeric Rating Scale (NRS) asking: 'To what extent have you recovered from your injury?', with response options of 0% (not recovered at all) to 100% (completely recovered) in 10% intervals. The order with which the new tool and the RM were presented was randomized for each respondent. This was a one-time cross-sectional study with no follow-up.

A second sample was subsequently recruited through 1 of 8 Physiotherapy clinics located across Canada. Subjects were eligible if they presented for rehabilitation following a traumatic injury to the neck, low back, upper or lower extremity. Eligibility criteria were consistent with that of the first sample. In addition to the prototype tool, the RMs, demographics and recovery indicators, these subjects also completed the Medical Outcomes Survey Short-Form 12 version 2 (SF-12v2) [12] through the same on-line platform. A request to complete the same set of questionnaires was sent 3 months later through an email link. Non-responders to the first request were sent a second and, where necessary, third request. Those not responding after 3 requests were considered lost to follow up for the 3-month follow-up period.

Measures Used

1. The prototype HRS tool was composed of 10 items tapping the following domains of health-related satisfaction: basic needs, cognitive function, physical fitness, ability to fulfill life roles, intimate relationships, connection with the community at large, independence, spontaneity, positive emotions, and potential for future growth. Each item received a rating of both personal importance and health-related satisfaction in that area. A weighted score was calculated as described in the analysis section below.

2. Regional Measures: These were chosen based on the areas injured as reported by the respondent. They were: Neck (Neck Disability Index) [5], Low back (Roland Morris Low Back Disability Questionnaire) [35], Lower extremity (Lower Extremity Functional Scale) [2], Upper extremity (Upper Extremity Functional Index [3], or Patient-Rated Wrist Evaluation [36] for wrist-specific disorders). Each scale has been subject to considerable psychometric evaluation previously and supported as adequately valid and reliable to be endorsed for routine clinical and research use [2, 3, 35, 37]. Scores on the LEFS and UEFI were reversed (80 - score), and then scores on each scale were converted to a percentage (summed score / maximum possible score x 100). In this way, scores on the regional measures were oriented such that a higher number indicated greater disability. Where multiple body regions were injured, the highest disability score was chosen for that subject.

3. The Short-Form 12v2: The SF-12 is a shortened version of the full SF-36. Both are considered to be valid measures of generic health status [12], providing scores across 8 subscales and 2 component summary scores, physical and mental. Version 2 of the SF-12 includes the same items as the original, but offers expanded response options on some items [38]. A transformation algorithm is required in order to calculate the subscale and component scores. For the purpose of this study, only the Physical Component Score (PCS) and Mental Component Score (MCS) were used. A higher score on each component score indicates better health.

ANALYSIS

The sample included 3 groups for statistical analysis: 1. The full sample (FULL) to evaluate content validity, factor structure and internal consistency; 2. A cross-sectional sample (CROSS) to evaluate cross-sectional convergent and

divergent validity; 3. A longitudinal sample (LONG) to evaluate responsiveness and known-groups validity. Characteristics of each sample were calculated and reported descriptively (mean and standard deviation, frequencies). Subsequent analyses are described in turn.

Individual Item Analysis

The purpose of this stage was to identify items to be earmarked for potential removal based on individual performance. Using data from the FULL sample, an inter-item correlation matrix including the 10 prototype items was constructed, with correlations of $r > 0.90$ indicating potential redundancies to be explored further. Content validity was partly supported by the nature of item generation (focus groups, expert opinion), and statistically was evaluated through calculating the mean and median importance score for each of the 10 prototype items. It was expected that all 10 items would be important, leading to an *a priori* threshold of a mean importance of 8.0 or greater as indicating adequate importance.

As an evaluative tool, all items were individually expected to meet a minimum ability to detect change. The LONG sample was used to identify underperforming items for detecting change over time. Weighted scores for each of the 10 items were first calculated by the following formula:

Weighted score = (Satisfaction x Importance) / 10

Change scores were calculated for each of the 10 items from baseline (T1) to 3-month follow-up (T2). Change (T2 - T1) was also calculated on the single 0-10 recovery item. Based on previous research using 11-point numeric rating scales [39], a change score of 2 points on the recovery NRS was considered a clinically relevant improvement in perceived recovery status. The LONG sample was dichotomized into those who had shown meaningful recovery over the 3 month span and those who had not. Ability to discriminate between those two groups was evaluated by creation of 10 individual Receiving Operating Characteristic (ROC) curves, one for each of the 10 items, and calculating the area under that curve (AUC). The null AUC was set at 0.50, meaning that each item should individually have an AUC with a lower 95% confidence limit greater than 0.50 in order to warrant retention. At this point, no items were removed but underperforming items were earmarked.

Finally, convergence with global recovery status was evaluated through correlation (Pearson's r) of each item with the recovery NRS at T1(baseline) using the CROSS sample. Any item that failed to show a significant correlation with recovery NRS was earmarked for removal.

Item removal was undertaken conservatively, out of respect for the rigorous generation process. Only those items that a) had a mean importance rating <8.0 / 10, b) were unable to identify meaningful change over 3 months (lower limit of AUC 95%CI <0.50), and c) did not show a significant ($p>0.05$) cross-sectional association with recovery status were removed.

Evaluation of Scale Properties

The remaining items were collectively referred to as the Satisfaction and Recovery Index (SRI) and were subjected to further evaluation of test properties drawn from Classical Test Theory. Missing responses were replaced with the mean if no more than 1 item was missing. Dummy items were added to the scale as dependency checks ('enter a '4' in this column') to ensure attention. Only those datasets that satisfied the dependency check and were not missing >1 response were retained for analysis. The overall SRI score was calculated by the formula:

SRI score = [(Sum of weighted scores as calculated above)/(Sum of importance scores only)] x 100

In this way, satisfaction in areas that were deemed more important was weighted heavier in the overall score than were those deemed less important. This had the added of benefit of meaning that *change* in areas of importance led to greater change in total score than did change in areas of less importance.

The following steps were subsequently conducted to test the factor structure, cross-sectional validity, responsiveness and known-groups discriminative properties of the tool:

Exploratory Factor Analysis (EFA) using Horn's Parallel Analysis technique [40] and Varimax rotation, allowing some degree of inter-correlation between factors. The FULL sample was used for this analysis, assuming sampling adequacy [Kaiser-Meier-Olkin (KMO) statistic ≥ 0.70 [41] and Bartlett's test of sphericity $p<0.05$ [42]]. Factor retention was based on eigenvalue, and internal consistency of identified factors was estimated using Cronbach's α where a value of > 0.70 was considered desirable [43].

Cross-sectional construct validity was evaluated using the CROSS sample. The dependent variable was the SRI, and the independent variables were the SF12v2 Physical Component Summary score (PCS), the SF12v2 Mental Component Summary score (MCS), and the score on the RM(% disability). A moderate (r = 0.4 to 0.7) positive association was expected between the SRI and the PCS. A small (r = 0.2 to 0.4) but significant positive association was expected between the SRI and the MCS. A moderate (r = -0.4 to -0.7) negative correlation was expected between the SRI and RMs. In all cases Pearson's r was used after assumptions of normality which were adequately satisfied through the Kruskal-Wallis test. Where normality was absent, Spearman's rho was used instead.

Longitudinal responsiveness was evaluated using the LONG sample. It was already known that each individual item was significantly able to detect change by virtue of the item retention/removal steps outlined above. The ability of the overall scale to identify change was compared to the same ability for the RM and SF12 component scales. Meaningful change was again operationalized as a change in recovery NRS of at least 2 points from T1 to T2, and ROC curves for each scale were constructed. The AUC (plotting *change* in recovery NRS against *change* in each scale) was compared across the 4 scales, with significant differences in responsiveness identified by non-overlapping 95% confidence intervals of the AUC.

RESULTS

The FULL sample was composed of 135 subjects. Of those, 9 subjects failed the dependency check and 3 provided incomplete data of >10% (>1 item) of the new scale. The

final FULL sample was therefore 123 subjects. Of those, 50 were asked to complete an additional set of scales including the SF12v2 and recovery indicators of which 46 (92%) provided complete data (the CROSS group). That same sample was invited to complete the scales a second time 3 months later, and 29 (63%) complied, forming the LONG group. The characteristics of the FULL sample and two sub-groups are presented in Table **1**.

Individual Item Analysis

The inter-item correlation matrix revealed no obvious redundancies, with the highest correlation being r = 0.84 between items 9 (feeling positive emotions) and 10 (having potential to grow in the future). Mean importance scores are reported in Table **2**. Item 6 'Connection with your community' had a mean importance score of 7.3, where all others were 8.1 or greater. Correlation of each item's weighted satisfaction score with the recovery NRS revealed that all of the items were significantly correlated with current recovery NRS (r = 0.33 to 0.54, $p < 0.05$) save for 'Connection with your community' (r = 0.22, $p > 0.10$). Finally for this stage of item analysis, change in recovery NRS from T1 to T2 revealed 18/29 subjects improved a meaningful amount while none worsened. Using the change scores for each item over that same span, AUC analysis revealed that all items save for 'Connection with your community' and 'Intimacy' were able to significantly discriminate between the groups (Table **3**). Item 6 'Connection with your community' failed all 3 tests for retention, and was therefore removed for subsequent analyses. The remaining 9 items formed the prototype SRI tool.

Scale Properties

Weighted responses from all 123 subjects in the FULL sample were entered into EFA after satisfying assumptions of sampling adequacy (KMO statistic = 0.92, Bartlett's χ^2 = 956.5, df = 36, $p < 0.01$). A single factor, 'Health-related satisfaction' (eigenvalue = 6.40) was extracted that explained 71.1% of total scale variance. Factor loadings ranged from 0.81 (Items 2 'Mentally sharp' and 7 'Being spontaneous) to 0.90 (Item 4: 'Fulfilling life roles). Cronbach's alpha of the overall scale indicated excellent internal consistency (α =0.95).

Table **3** shows the results of the cross-sectional correlational analyses. In each case, the *a priori* hypotheses for magnitude of association were supported. However, while meeting the hypothesized magnitude, the correlation

Table 1. Characteristics of the 3 samples used in the different analyses.

	FULL (N = 123)	CROSS (N = 46)	LONG (N = 29)
Sex (% female)	58.4%	58.7%	72.4%
Age (mean, SD)	42.1 (12.9)	38.1 (13.5)	42.6 (13.6)
Duration of symptoms			
<3 months	49.0%	100.0%	100.0%
≥3 months	51.0%	0.0%	0.0%
Region affected (%)*			
Neck	63.6%	58.7%	44.9%
Low back	38.3%	23.9%	13.8%
Lower extremity	18.7%	28.3%	27.6%
Upper extremity	12.1%	43.5%	37.9%
Mechanism of injury (%)			
Motor vehicle accident	42.9%	53.3%	44.8%
Other type of impact	21.9%	13.3%	6.9%
Fall, trip or slip	4.8%	11.1%	31.0%
Awkward lifting/twisting	5.7%	13.3%	6.9%
Other	23.8%	8.9%	10.3%
Medicolegal status (%)			
Motor vehicle insurance	39.1%	39.1%	31.0%
Worker's compensation	19.5%	13.0%	6.9%
Current litigation	9.5%	8.7%	0.0%
Work status at inception			
Full pre-injury	36.3%	36.7%	56.3%
Modified return	42.2%	43.3%	37.5%
Disability leave	21.6%	20.0%	6.3%
Off for other reasons	2.0%	6.7%	0.0%

*: Respondents could choose more than 1 region affected if applicable.

Table 2. Individual item analysis results.

N = 123	Mean	Median	Range	Recovery Correlation	AUC (95% CI)
1. Meeting your most basic needs	9.6	10	5 - 10	0.53**	0.82 (0.67, 0.97)
2. Being mentally sharp	9.5	10	7 - 10	0.56**	0.87 (0.73, 1.00)
3. Being physically fit compared to others like you	9.1	10	5 - 10	0.53**	0.84 (0.67, 1.00)
4. Fulfilling your life roles	9.5	10	3 - 10	0.45**	0.74 (0.57, 0.93)
5. Intimate relationships	9.2	10	4 - 10	0.38*	0.68 (0.48, 0.89)
6. Being connected with your community at large	**7.3**	**8**	**0 - 10**	**0.19**	**0.51 (0.25, 0.77)**
7. Being independent	9.5	10	2 - 10	0.48**	0.90 (0.79, 1.00)
8. Being spontaneous	8.1	8	3 - 10	0.51**	0.84 (0.69, 1.00)
9. Feeling positive emotions	9.3	10	6 - 10	0.34*	0.89 (0.77, 1.00)
10. Feeling like you've got the potential to achieve new or greater things in the future	9.2	10	5 - 10	0.42**	0.81 (0.65, 0.98)

Columns 2-4: Importance ratings for each of the 10 items. 0 = not important to me at all, 10 = extremely important. Column 5: Pearson's r correlation coefficient between each item and score on a recovery NRS (0 = not recovered at all, 10 = completely recovered). Column 6: The area under the Receiver Operating Characteristic curve (AUC) for discriminating between changed and stable recovery status over a 3-month period. Bold indicates the single item that failed all 3 tests for retention and was removed from the final tool.
* = correlation significant at the $p < 0.05$ level.
** = correlation significant at the $p < 0.01$ level.

between the SRI and SF12v2 MCS was not statistically significant ($r = 0.28$, $p = 0.14$).

Responsiveness estimates are shown in Table **4**. Both the SRI and RMs showed significant ability to discriminate between the improved/not improved groups at 3 months (AUC = 0.82 and 0.79, respectively). Neither of the SF-12v2 subscales were able to discriminate between groups, although 6 subjects had to be excluded from this analysis due to missing data or obvious response bias on the SF-12 at T2 (e.g. scoring lowest or highest numbers all the way down even on reverse-scored items). This meant that confidence intervals were wider than desirable for the PCS and MCS analyses. Fig. (**1**) shows the receiver operating characteristic (ROC) curves for all 4 scales. Both the SRI and RMs showed evidence of significantly greater responsiveness in this sample compared to the MCS subscale, by virtue of non-overlapping confidence intervals and point estimates.

Table 3. Cross-sectional means (column 2) and associations (column 3).

N = 46	Mean (SD, Range)	Pearson's r
SRI	63.4% (25.5, 2.0% to 100.0%)	
Region-specific disability	39.6% (23.3, 0.0% to 89.5%)	-0.67**
SF12 PCS	64.7% (18.2, 29.3 to 99.5)	0.45*
SF12 MCS	73.2% (10.5, 53.0 to 94.2)	0.28

*: correlation with SRI is significant at the $p < 0.05$ level. **: correlation with SRI is significant at the $p < 0.01$ level.

DISCUSSION

A new, importance-weighted health-related satisfaction scale has been created and its measurement properties appear sound during developmental evaluation. It is composed of a single factor with excellent internal consistency and performs comparably as an evaluative outcome measure when compared to region-specific disability scales, and better than the generic SF-12v2 health status measure in this sample of ambulatory community-dwelling people with traumatic musculoskeletal injuries. The items are generic enough to allow comparison across clinical conditions while the use of standardized items overcomes conceptual challenges with the use of scales composed of patient-generated items. Furthermore, importance-weighting overcomes the conceptual challenges posed by many regional measures that necessarily assume equal importance of all items to all people.

Table 4. Responsiveness (area under the Receiver Operating Characteristic curve (AUC) when change score on each scale was plotted against clinically meaningful improvement on the recovery NRS) estimates of the tools under evaluation.

	Responsiveness AUC (95%CI)
SRI	0.82 (0.67, 0.97)
Region-specific disability	0.79 (0.62, 0.96)
SF12 PCS	0.69 (0.42, 0.86)
SF12 MCS	0.50 (0.25, 0.70)

SRI = Satisfaction and Recovery Index; SF12 PCS = Physical Component Summary score of SF-12; SF12 MCS = Mental Component Summary score of the SF-12.

To our knowledge, this is the first published scale to measure importance-weighted health-related satisfaction. In its current form, the SRI is composed of 9 items that all appear to represent important influences on our respondents' sense of recovery and satisfaction. It is phrased in a positive direction, accessible and easy to score, sensitive to change

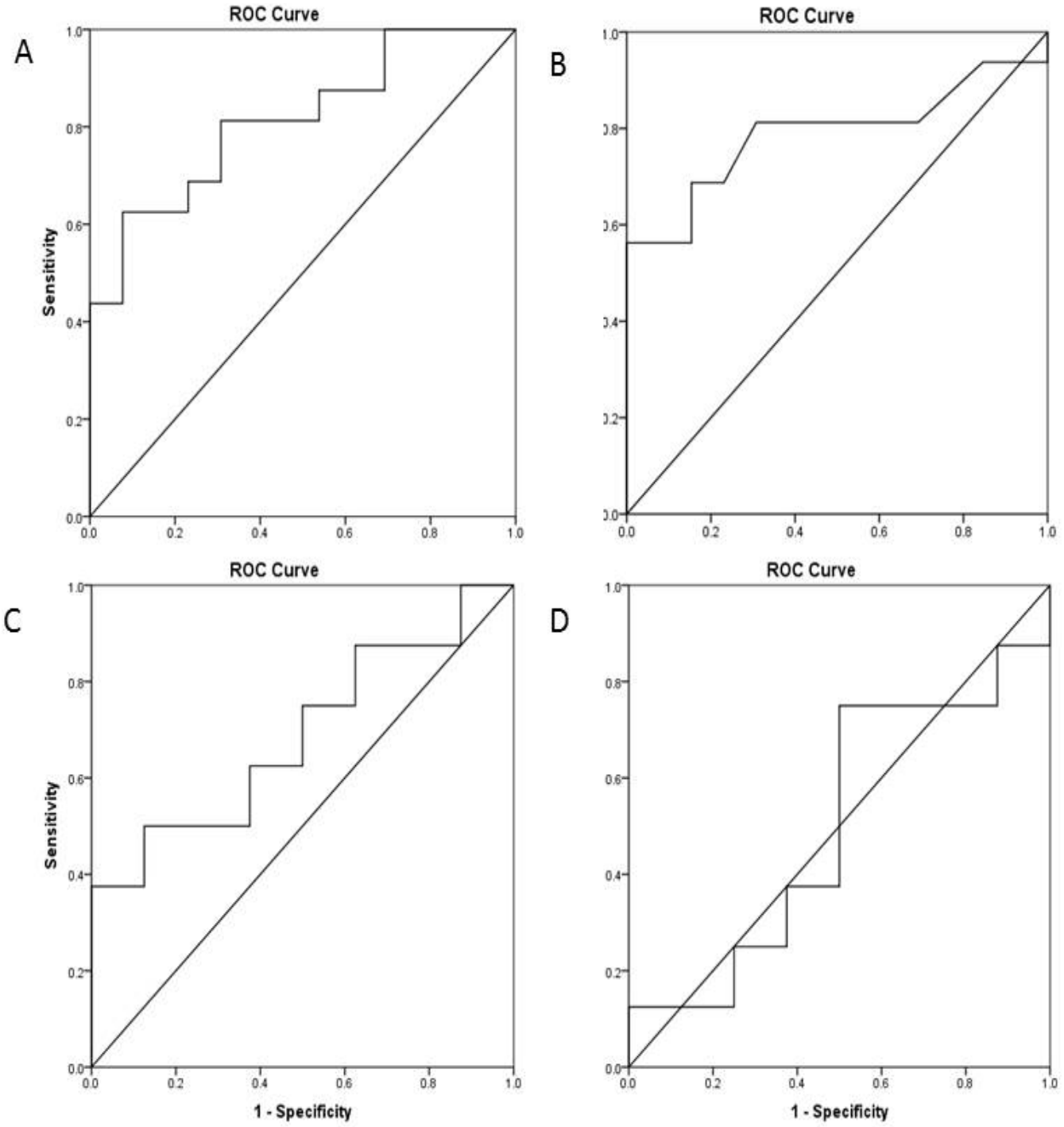

Fig. (1). Receiver Operating Characteristic (ROC) curves showing discriminatory ability of **A**: Satisfaction and Recovery Index, **B**: Regional Measures, **C**: SF-12 Physical Component Subscore, **D**: SF-12 Mental Component Subscore to detect change based on global recovery score (≥2 / 10 point change). The diagonal line represents an area under the curve (AUC) of 0.50.

during the *process* of recovery, and adequately associated with the *state* of recovery. The focus on satisfaction is considered to be unique in comparison to tools that focus on concepts like symptoms or function that are variably associated with recovery. Hence, this measure is proposed to fill a niche not addressed by current tools. The potential value of this measure as a replacement for, or addition to, current tools could not be defined by this study but requires evaluation across multiple contexts.

The potential for the SRI to be useful in clinical practice and research is currently based on conceptual rationale. The SRI may have particular relevance in research where health status has to be dichotomized as‘recovered’ or ‘not recovered’. ‘Good’ or ‘bad’ outcomes for prognostic research are often based on a cut-score on a regional condition-specific measure. In MSK trauma research, ‘return to pre-injury status’is often considered a desirable outcome, but it is difficult to operationalize since pre-injury data are

rarely available. Further, a return to pre-injury status itself becomes less compelling to indicate recovery as the duration of a condition increases. The SRI has potential application to provide a measure of treatment response and recovery that recognizes a respondent may be satisfied regardless of how their current state compares with their pre-injury one.

Since validity is context-specific and based on indirect interpretation it is not a result, but a process. Overall, the results of this study provide a sound start to that process by providing evidence for the content, structural (factor), cross-sectional, and longitudinal validity of the SRI. A series of *a priori* hypotheses were defined prior to initiating data collection, to define how the SRI was expected to relate to regional and general health status measures. In all cases save for one, those were supported, which suggests that the SRI measures the construct of recovery as expected. The exception was the association between the Mental Component Summary score of the SF-12v2, which was of the expected magnitude, but not statistically significant. Post-hoc power analysis indicated that the analysis was adequately powered for the expected correlation (power > 0.80), suggesting that the lack of statistical association was not a function of sample size. A logical question is whether the association between these tools may be different in other populations with a wider range of mental health problems. Table **3** indicates that, on average, the sample was not highly disabled by their injuries (mean regional measure of 39.6% disability). While consistent with the sampling frame, this means that the results are most easily generalizable to other samples that are moderately disabled by their injuries. The function of the SRI in more severely affected patients or those with significant comorbidities has to be considered currently unknown.

It is possible that weighting by preference only adds complexity and may not improve the performance of the scale, such as was found for the UK oral health-related quality of life measure [44]. However, other scales, including the COPM, have seen value in this additional step [31]. With respect to the SRI specifically, the middle column of Table **2** provides arguably the most compelling reason to retain the importance-weighting: while the sample median importance for each item was high, the range clearly indicates that not all items are equally important to all respondents. While these minor deviations are obscured by statistics that use group means for analysis, they may be very meaningful at the level of the individual patient when used in clinical practice. Answering the items twice does add burden, although the simplicity of the scale and its structure are thought to make this additional burden small, with time to completion for most subjects <5 minutes. Further, it should not be assumed that importance on each item is a stable trait - the field of Quality of Life research has provided ample evidence for the phenomenon of response shift [45,46], which is assumed to be a function of shifting priorities (or importance) of life domains over the course of living with chronic disease. Thus, the SRI might provide a tool for measuring response shift in addition to recovery status. For this reason, it is recommended that both the importance and satisfaction ratings on each item are collected at each assessment period, allowing finer interpretation of patient reports.

The ability of the regional measures to respond to change in our sample is generally in keeping with the results of previous research on each scale independently [2, 3, 47, 48]. Similarly, the relative inability of the SF-12 component scores to detect change in this population has been reported previously, especially the MCS [48, 49]. While measurement of recovery across patient-derived domains was not *more* responsive than the regional measures, the fact that a generic measure was compared favorably to targeted RMs across a range of MSK injuries is an important finding when this has long been considered a strength of RMs. It also appears as though the information is not redundant, as indicated by the moderate correlation between the two, suggesting that there may be value in capturing *both* a targeted region-specific scale and the more generic SRI in research and clinical contexts in order to provide a more detailed description of a patient's status.

One key scale property that has yet to be evaluated is short-term test-retest reliability. While reliability was good in unreported data collected during development, a fully powered reliability study is needed, and is currently underway. An additional limitation to those already discussed is the higher-than-desired 37% rate of dropouts in the LONG sample. Dropouts were different from completers in potentially important ways. Dropouts were on average younger, more likely to be male, and less likely to be working full time at inception. The remaining sample was too small to examine differences in scale properties by age, sex or work status, but until shown otherwise readers should be aware that scale properties may differ in these potentially important clinical subgroups.

On balance, the preliminary results of the SRI evaluation are promising. The scale is provided by open access for free use. Its contribution to better understanding of outcomes in practice and research will require further study. The cost/benefit value of weighting the scale, and its short-term retest reliability are psychometric issues that require further study. Areas of research that might particularly benefit from this tool are outcomes of MSK injury, prognosis, response shift, and clinical measurement research. Implementation research, including case studies, may be needed to determine how it contributes to management of individual patients. Overall, the SRI is a brief measure of importance-weighted health-related satisfaction that has demonstrated initial evidence of content, structural, and construct validity, and that provides responsiveness to change that is similar to targeted regional outcome measures. The tool is provided in Appendix **A**.

AUTHORS CONTRIBUTIONS

DW conceptualized the study, analyzed the data, and prepared the first draft of the manuscript, JM co-created the tool, assisted in interpretation of results and data presentation, MP, AR and JV assisted in clinical data collection, interpretation of results, and provided important contributions to earlier drafts of the manuscript.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

DW was funded by a Doctoral Fellowship through the Canadian Institutes of Health Research when this work was initiated.

APPENDIX

Name: _____ Date: _____

Satisfaction and Recovery Index

Below are 10 areas of life that other people in pain have identified as influencing recovery and satisfaction. For each row, please indicate 1: how *important* that area is to you personally, and 2: how *satisfied* you currently feel in that area considering any interference from your injury or symptoms. Note that it is possible to feel satisfied in an area that is not important to you, or to feel dissatisfied in an area that is important to you. Use the following scale:

Importance:

0	1	2	3	4	5	6	7	8	9	10
Not important to me at all					Moderately important to me					Extremely important to me

Satisfaction:

0	1	2	3	4	5	6	7	8	9	10
Not satisfied at all (complete interference)										Completely satisfied (no interference)

	Importance (0-10)	Satisfaction (0-10)
1. Meeting your most basic needs (e.g., eating well, good sleep, good personal hygiene, etc…)		
2. Being mentally sharp (i.e., your ability to concentrate, remember or think quickly)		
3. Being physically fit (eg., strong, energetic or flexible) compared to other people of your age and sex		
4. Fulfilling your 'life roles' (e.g., being a spouse, friend, parent, coworker and/or volunteer)		
5. Intimate relationships, whether they be physical relationships or close personal relationships above the level of normal friendship		
6. For validation purposes, place a '4' in the Importance column, and a '6' in the Satisfaction column in this row		
7. Being independent (e.g., making your own decisions and being in control of your own life)		
8. Being spontaneous (doing things without having to plan)		
9. Feeling positive emotions (e.g. happiness, joy, self-esteem)		
10. Feeling like you've got the potential to achieve new or greater things in the future		

REFERENCES

[1] Canadian Institute on Patient-Oriented Research. Strategy on Patient-Oriented Research. US: National Academy of Sciences 2012.

[2] Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999; 79(4): 371-83.

[3] Stratford PW, Binkley JM, Stratford DM. Development and initial validation of the upper extremity functional index. Phys Can 2001; (Fall): 259-67.

[4] Pinfold M, Niere KR, O'Leary EF, Hoving JL, Green S, Buchbinder R. Validity and internal consistency of a whiplash-specific disability measure. Spine 2004; 29(3): 263-8.

[5] Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. J Manipulative Physiol Ther 1991; 14(7): 409-15.

[6] Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980; 66(8): 271-3.

[7] Hill JC, Dunn KM, Lewis M, *et al*. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008; 59(5): 632-41.

[8] Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singap 1994; 23(2): 129-38.

[9] Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. Arch Phys Med Rehabil 1987; 68(7): 438-41.

[10] Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997; 36(5): 551-9.

[11] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30(6): 473-83.

[12] Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996; 34(3): 220-33.

[13] Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995; 22(6): 1193-6.

[14] MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH. Responsiveness of the short form-36, disability of the arm, shoulder, and hand questionnaire, patient-rated wrist evaluation, and physical impairment measurements in evaluating recovery after a distal radius fracture. J Hand Surg Am 2000; 25(2): 330-40.

[15] Giesinger K, Hamilton DF, Jost B, Holzner B, Giesinger JM. Comparative responsiveness of outcome measures for total knee arthroplasty. Osteoarthr Cartil 2014; 22(2):184-9.

[16] Loza E, Abasolo L, Jover JA, Carmona L, EPISER Study Group. Burden of disease across chronic diseases: a health survey that measured prevalence, function, and quality of life. J Rheumatol 2008; 35(1): 159-65.

[17] Working Group on Health Outcomes for Older Persons with Multiple Chronic Conditions. Universal health outcome measures for older persons with multiple chronic conditions. J Am Geriatr Soc 2012; 60(12): 2333-41.

[18] Brazier JE, Rowen D, Mavranezouli I, *et al*. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technol Assess 2012; 16(32): 1-114.

[19] MacDermid JC, Wojkowski S, Kargus C, Marley M, Stevenson E. Hand therapist management of the lateral epicondylosis: a survey of expert opinion and practice patterns. J Hand Ther 2010; 23(1): 18-29; quiz 30.

[20] Macdermid JC, Vincent JI, Kieffer L, Kieffer A, Demaiter J, Macintosh S. A survey of practice patterns for rehabilitation post elbow fracture. Open Orthop J 2012; 6: 429-39.

[21] Michlovitz SL, LaStayo PC, Alzner S, Watson E. Distal radius fractures: therapy practice patterns. J Hand Ther 2001; 14(4): 249-57.

[22] Schmitt J, Di Fabio RP. The validity of prospective and retrospective global change criterion measures. Arch Phys Med Rehabil 2005; 86(12): 2270-6.

[23] Macdermid JC, Walton DM, Cote P, *et al*. Use of outcome measures in managing neck pain: an international multidisciplinary survey. Open Orthop J 2013; 7: 506-20.

[24] Mehta S, Grafton K. A survey on the use of outcome measures by musculoskeletal physiotherapist's in India. Physiother Theory Pract 2014; 30(2): 110-22.

[25] Antunes B, Harding R, Higginson IJ, on behalf of EUROIMPACT. Implementing patient-reported outcome measures in palliative care clinical practice: A systematic review of facilitators and barriers. Palliat Med 2014; 28(2): 158-75.

[26] Walton DM, Macdermid JC, Nielson W. Recovery from acute injury: Clinical, methodological and philosophical considerations. Disabil Rehabil 2010; 32(10): 864-74.

[27] Walton DM, Caroll LJ, Kasch H, *et al.* An overview of systematic reviews on prognostic factors in neck pain: results from the international collaboration on neck pain (ICON) project. Open Orthop 2013; 7(Suppl 4): 494-505.

[28] Walton DM, Macdermid JC, Giorgianni AA, Mascarenhas JC, West SC, Zammit CA. Risk Factors for persistent problems following acute whiplash injury: update of a systematic review and meta-analysis. J Orthop Sports Phys Ther 2013; 43(2): 31-43.

[29] Walton D. A review of the definitions of recovery used in prognostic studies on whiplash using an ICF framework. Dis Rehabil 2009; 31(12): 943-57.

[30] Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. J Orthop Sports Phys Ther 1998; 27(5): 331-8.

[31] Law M, Baptiste S, McColl MA, Opzoomer A, Polatajko H, Pollock N. The Canadian Occupational Performance Measure: An outcome measure for occupational therapy. Can J Occup Ther 1990; 57(2): 82-7.

[32] Horn KK, Jennings S, Richardson G, Vliet DV, Hefford C, Abbott JH. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. J Orthop Sports Phys Ther 2012; 42(1): 30-42.

[33] Walton DM, Macdermid JC, Taylor T, ICON. What does 'recovery' mean to people with neck pain? Results of a descriptive thematic analysis. Open Orthop J 2013; 7: 420-7.

[34] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000; 25(24): 3186-91.

[35] Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983; 8(2): 141-4.

[36] MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. J Orthop Trauma 1998; 12(8): 577-86.

[37] Vernon H. The Neck Disability Index: state-of-the-art, 1991-2008. J Manipulative Physiol Ther 2008; 31(7): 491-502.

[38] Ware JE, Kosinski M, Turner-Bowker DM, Gandek B. How to score version 2 of the SF-12 health survey (with a supplement documenting version 1). Lincoln RI, Quality Metric Incorporated 2002.

[39] Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004; 8(4): 283-91.

[40] Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965; 32: 179-85.

[41] Kaiser HF. An index of factorial simplicity. Psychometrika 1974; 39(1): 31-6.

[42] Bartlett MS. Test of significance in factor analysis. Br J Psychol 1950; 3: 77-85.

[43] Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ 1997; 314: 572.

[44] McGrath C, Bedi R. Why are we "weighting"? An assessment of a self-weighting approach to measuring oral health-related quality of life. Community Dent Oral Epidemiol 2004; 32(1): 19-24.

[45] Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MA, Fayers PM. The clinical significance of adaptation to changing health: a meta-analysis of response shift. Qual Life Res 2006; 15(9): 1533-50.

[46] Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 1999; 48(11): 1531-48.

[47] Walton DM, MacDermid JC. A brief 5-item version of the Neck Disability Index shows good psychometric properties. Health Qual Life Outcomes 2013; 11: 108-7525-11-108.

[48] Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. Med Care 2010; 48(11): 1007-14.

[49] Ebert JR, Smith A, Wood DJ, Ackland TR. A comparison of the responsiveness of 4 commonly used patient-reported outcome instruments at 5 years after matrix-induced autologous chondrocyte implantation. Am J Sports Med 2013; 41(12): 2791-9.

Received: April 9, 2014 Revised: August 22, 2014 Accepted: September 3, 2014