Original Paper

The Online Patient Satisfaction Index for Patients With Low Back Pain: Development, Reliability, and Validation Study

Tamana Afzali[1], MSc; Henrik Hein Lauridsen[2], PhD; Janus Laust Thomsen[1], PhD; Jan Hartvigsen[2,3], PhD; Martin Bach Jensen[1], PhD; Allan Riis[1,4], PhD

[1]Research Unit for General Practice in Aalborg, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[2]Departments of Sports Science and Clinical Biomechanics, Centre for Muscle and Joint Health, University of Southern Denmark, Odense, Denmark

[3]Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark

[4]Department of Physiotherapy, University College Northern Denmark, Aalborg, Denmark

Corresponding Author:
Allan Riis, PhD
Research Unit for General Practice in Aalborg
Department of Clinical Medicine
Aalborg University
Fyrkildevej 7
Aalborg, 9220
Denmark
Phone: 45 20823660
Email: ariis@dcm.aau.dk

Abstract

Background: Low back pain is highly prevalent, and most often, a specific causative factor cannot be identified. Therefore, for most patients, their low back pain is labeled as nonspecific. Patient education and information are recommended for all these patients. The internet is an accessible source of medical information on low back pain. Approximately 50% of patients with low back pain search the internet for health and medical advice. Patient satisfaction with education and information is important in relation to patients' levels of inclination to use web-based information and their trust in the information they find. Although patients who are satisfied with the information they retrieve use the internet as a supplementary source of information, dissatisfied patients tend to avoid using the internet. Consumers' loyalty to a product is often applied to evaluate their satisfaction. Consumers have been shown to be good ambassadors for a service when they are willing to recommend the service to a friend or colleague. When consumers are willing to recommend a service to a friend or colleague, they are also likely to be future users of the service. To the best of our knowledge, no multi-item instrument exists to specifically evaluate satisfaction with information delivered on the web for people with low back pain.

Objective: This study aims to report on the development, reliability testing, and construct validity testing of the Online Patient Satisfaction Index to measure patients' satisfaction with web-based information for low back pain.

Methods: This is a cross-sectional validation study of the Online Patient Satisfaction Index. The index was developed with experts and assessed for face validity. It was subsequently administered to 150 adults with nonspecific low back pain. Of these, 46% (70/150) were randomly assigned to participate in a reliability test using an intraclass correlation coefficient of agreement. Construct validity was evaluated by hypothesis testing based on a web app (MyBack) and Wikipedia on low back pain.

Results: The index includes 8 items. The median score (range 0-24) based on the MyBack website was 20 (IQR 18-22), and the median score for Wikipedia was 12 (IQR 8-15). The entire score range was used. Overall, 53 participants completed a retest, of which 39 (74%) were stable in their satisfaction with the home page and were included in the analysis for reliability. Intraclass correlation coefficient of agreement was estimated to be 0.82 (95% CI 0.68-0.90). Two hypothesized correlations for construct validity were confirmed through an analysis using complete data.

Conclusions: The index had good face validity, excellent reliability, and good construct validity and can be used to measure satisfaction with the provision of web-based information regarding nonspecific low back pain among people willing to access the internet to obtain health information.

Trial Registration: ClinicalTrials.gov NCT03449004; https://clinicaltrials.gov/ct2/show/NCT03449004

XSL•FO
RenderX

Introduction

Background

Low back pain (LBP) is highly prevalent and is the most frequent reason for patients to consult general practice in Denmark [1,2]. LBP affects men and women of all ages [3] and is rarely caused by one specific factor [4,5]. Therefore, for most patients, their LBP is labeled as nonspecific, that is, a nociceptive source is not well established, and causes are multifactorial [6,7]. Patient education and information are generally recommended for people seeking care for nonspecific LBP [8]. However, delivering evidence-based information can be time-consuming and cumbersome, which can be a challenge during the available consultation time in general practice [9].

The internet is an accessible source of medical information for patients and it offers a range of information provided by a variety of sources. It has been reported that approximately 50% of patients search the internet for health and medical advice [10], and evidence suggests that this is increasing [11]. The advancement of new technologies offers more opportunities for delivering patient information on private computers, tablets, and smartphones, and web-based information can be considered an inexpensive solution to extend the treatment in general practice [12]. Therefore, future optimization of web-based information delivery has the potential to increase the delivery of evidence-based information about LBP, which may, in turn, lead to better patient outcomes [13,14].

Patient satisfaction is important in the use of web-based information and the degree to which patients rely on information from the internet [15]. Although patients who are satisfied with the information they retrieve use the internet as a supplementary source of information, dissatisfied patients tend to avoid using the internet [11]. Consumers' loyalty to a product is often applied to evaluate their satisfaction [16]. Consumers have been shown to be good ambassadors for a service or product when they are willing to recommend the service to a friend or colleague [16,17]. When consumers are willing to recommend a service to a friend or colleague, they are also likely to be future users of the service [16,17].

Objectives

To the best of our knowledge, no multi-item instrument to specifically evaluate satisfaction with information delivered on the web for people with LBP exists. This study aims to report on the development and validation of the Online Patient Satisfaction Index (OPSI), a self-reported measure to evaluate patients' satisfaction with web-based information for LBP.

Methods

Overview

This study was registered at ClinicalTrials.gov (ID: NCT03449004). The study follows the Consensus-based Standards for the Selection of Health Measurement Instruments Taxonomy [18]. The *Methods* section consists of 2 subsections: development of the OPSI and validation of the OPSI.

Development of the OPSI

Previous Work

A qualitative interview study had previously identified patients' preferences for the content, design, and functionality of a web application with evidence-based information and advice for patients with LBP consulting general practice [19]. This study identified a set of important domains to address web-based information and advice for patients with LBP in Denmark and highlighted the importance of the following domains: design, readability, customization, credibility, usability, and coping [19]. On the basis of these findings, 8 specific items related to these domains were identified as important for patient satisfaction with web-based information for LBP. Design and readability were represented with 2 items; the other domains were presented with 1 item each.

Development Process

A total of 8 items were combined into the first version of the index. The content of the 8 items in the first version came from an interview study [19], after which the authors made a draft version where experts (not members of the author group) and 10 patients provided input. Thereafter, the reliability and validity were tested. All items initially had response options ranging from 0 to 10, where 0 indicated strongly dissatisfied and 10 indicated strongly satisfied. The first version of the index was then tested for face validity by discussing the wording of the items with 7 experts. The experts were personally invited among colleagues but were outside the author group (1 academic and 1 researcher experienced in written communication, 2 researchers with expertise in musculoskeletal disorders, and 3 researchers with expertise within the development of questionnaires). This process was carried out through 2 rounds, where the first draft was discussed among the experts and subsequently modified. After round 1, the questions were reformulated to reduce jargon and for better wording of the items, and the order of the items was rearranged to create a better flow of the index. This revised index was then discussed with the same experts until a consensus on the final version was reached. Importantly, the 0 to 10 response rate scale about satisfaction was found to be difficult to use, and therefore, the response scale was changed from the numerical rating scale to a categorical scale about satisfaction with 4 response options: *Very Much, Quite a bit, A little,* and *Not at all* (Figure 1).

Figure 1. The Online Patient Satisfaction Index, translated to English from Danish.

Here is a questionnaire about your satisfaction with the website. Please tick a box for every question.

If you find it hard to select an answer, we recommend that you choose the answer, which fits you best.

	Very Much	Quite a bit	A little	Not at all
1. Do you like the website's appearance?	☐	☐	☐	☐
2. Is it easy to find your way around the website?	☐	☐	☐	☐
3. Do you understand the content?	☐	☐	☐	☐
4. Is the language suitable for you?	☐	☐	☐	☐
5. Is the content relevant for you?	☐	☐	☐	☐
6. Do you trust the website?	☐	☐	☐	☐
7. Did you find what you were looking for?	☐	☐	☐	☐
8. Is the website a good tool to help you with your back problem?	☐	☐	☐	☐

Conceptual Framework

The OPSI is based on a formative model in which the construct (satisfaction) is the result of patients' experiences with different aspects of satisfaction. For example, the items relating to design, credibility, and readability can all have an impact on patients' satisfaction with the web-based information, whereas higher satisfaction with a home page does not necessarily lead to patients finding it more customized to their needs [20].

Face Validity

The OPSI was pilot tested for face validity on 10 respondents with nonspecific LBP from the Sano Centre and 10 respondents with LBP recruited from social media. The Sano Centre is a training and rehabilitation center for people with a high degree of musculoskeletal pain or disability. First, they were asked to fill in baseline characteristics on paper for age, sex, pain duration (>12 weeks), pain intensity, curiosity to find new knowledge (0-10), and frequency of internet searching for health-related information (monthly or more). With 1 researcher (AR or TA) present, respondents were asked to search for information on an existing website (The Patient Handbook) [21]. The publicly available Patient Handbook has previously been found to be trustworthy and a preferred site among Danes searching for

information about LBP [19]. The author group was not involved in developing the design or choosing the content of the Patient Handbook. After assessing the Patient Handbook for 10 minutes, respondents were asked to complete the OPSI and were encouraged to comment openly on the process and content. Their thoughts and comments were noted on paper by the researcher. This was done to optimize the content validity of the items by reducing ambiguity, avoiding double-barreled questions, reducing jargon terms, reducing the length of the items, checking the existence of irrelevant items, and patients were asked if there was a lack of any items related to patient satisfaction with web-based information. We specifically asked about the feasibility of the questions, their understanding of the items, and the reasons for their choice of response options. Respondents' thoughts and suggestions about the index were discussed between TA and AR, and the index was revised and ready for validation among a larger population of respondents with nonspecific LBP. Figure 1 shows the English version of the OPSI, which has been forward-backward translated from Danish using the method suggested by Beaton et al [22] with modifications to stage 4. In stage 4, 1 native English-speaking researcher in expertise in musculoskeletal disorders achieved consensus with 2 native Danes holding a master's degree in

English. Although stages 1 to 3 were conducted at personal meetings, stage 4 was conducted on the web with TA as a facilitator. The Danish version is shown in Figure 2.

Figure 2. The Online Patient Satisfaction Index in Danish. Each item contributes 0 to 3 points, giving a total Online Patient Satisfaction Index score between 0 and 24.

Nedenunder finder du spørgsmål angående din tilfredshed med hjemmesiden.
Du bedes sætte ét kryds ved hvert spørgsmål. Hvis du har svært ved at vælge dit svar, opfordrer vi til at du vælger det svar, som passer bedst.

	Meget	Noget	Lidt	Slet ikke
1. Kan du lide hjemmesidens udseende?	☐	☐	☐	☐
2. Er det nemt at finde rundt på hjemmesiden?	☐	☐	☐	☐
3. Forstår du indholdet?	☐	☐	☐	☐
4. Er sproget målrettet til dig?	☐	☐	☐	☐
5. Er indholdet relevant for dig?	☐	☐	☐	☐
6. Stoler du på hjemmesiden?	☐	☐	☐	☐
7. Fandt du det, du ledte efter?	☐	☐	☐	☐
8. Er hjemmesiden et godt værktøj til at hjælpe dig med dine rygproblemer?	☐	☐	☐	☐

Construct Validity and Reliability of the OPSI

Data from the validation study were gathered using paper versions of the OPSI, and TA entered the data in REDCap (Research Electronic Data Capture, Vanderbilt University) [23].

Participants

Sample sizes for validation studies are recommended to contain more than 100 participants and at least 50 participants per subgroup [20]. Consequently, the required sample size was set at 150 participants. The sample population for the development of the OPSI was recruited from Sano Centre in Aarhus, Denmark, and via social media (Facebook, LinkedIn, and Twitter) to obtain a case-mix. The patients had a 4-week stay at the center, and the training and rehabilitation course was patient-centered and tailored to the patients' needs. Patients were offered specialized therapy, and most received an iPad as part of their training program [24]. Hence, most people were familiar with the internet and electronic devices to manage their pain. Respondents from social media were also expected to be familiar with the use of the internet and electronic devices. Consequently, the sample population was expected to be heterogeneous regarding their levels of pain and functional disability.

Respondents from Sano and social media were eligible for inclusion if they had nonspecific LBP (with or without leg pain) of any pain intensity during the previous year and were older than 18 years. Exclusion criteria were as follows: no internet access, pregnancy, inability to speak Danish as their native language, diagnosis of spinal stenosis, or signs of a serious underlying disease (signs of fracture, cauda equina syndrome, malignancy, osteoporosis, or spondyloarthritis).

Procedures

The study was registered by the Danish Data Protection Agency (J.nr. 2017-41-5222). Ethics approval was not required following Danish law. Respondents received verbal project information from TA or AR, and informed consent was signed by the respondents and the assessor. Testing was performed at the Sano Centre, public libraries, or in respondents' homes with either TA or AR present. Initially, respondents filled in baseline questions on paper and were encouraged to navigate and search for a new home page, MyBack, for 10 minutes. The MyBack home page is in Danish and contains information about LBP to guide patients with self-management [25]. The content and design were developed by researchers with systematic input from patients and general practitioners [19,26,27]. Respondents

then filled in paper versions of the OPSI to measure their satisfaction with MyBack together with other questions about satisfaction and their functional disability. If any item was left blank, the assessor encouraged respondents to choose the response that was most suitable for them. Thereafter, respondents navigated Wikipedia with information about LBP for 10 minutes with minimum help from the assessor. We assumed that most respondents would be more satisfied with MyBack than with information about LBP on Wikipedia. As the amount of information regarding LBP is limited in the Danish version of Wikipedia, we considered it difficult to read, not addressed to a particular group of people, and the sources are unknown and might, therefore, not be considered trustworthy by respondents. Thereafter, respondents filled out the OPSI for the Wikipedia page.

Among the respondents, 70 were randomized to be invited to participate in the retest after a minimum of 7 days and a maximum of 4 weeks. Randomization and allocation numbers were provided by a researcher who was not involved in this study. A publicly available home page was used to generate 150 numbers with *yes* or *no* and sent the document with the allocation numbers to the assessors [28]. Respondents randomized to *yes* were invited to participate in a retest. Respondents who agreed to take part in the retest searched and navigated the MyBack home page a second time after a minimum of 7 days with the same assessor present (TA or AR). After filling out paper versions of the OPSI, respondents were encouraged to respond to 1 question about the stability of their satisfaction: Do you think your satisfaction with the home page has changed since last time? (Answer options: yes/no/don't know). Their replies were discussed with the assessor to validate their responses and if responding *no* or *don't know*, they were considered stable and included in the retest analysis.

Measurement Tools

A web app can be considered a product, and we assume that users are satisfied when they are more likely to recommend a web app to a friend or colleague. *The Ultimate Question* was used as the primary outcome measurement to compare measurement properties with the OPSI. *The Ultimate Question* is often applied to measure costumers' satisfaction with products or services using a single question: How likely are you to recommend the website to others? The question can be answered using a 0 to 10 scale, where 0 is the least likely to recommend and 10 is the most likely to recommend [16,17]. Respondents replying 10 or 9 are considered promotors and are likely to buy or use the product or service again, respondents replying 8 or 7 are considered passive, and respondents replying 6-0 are considered detractors [16,17]. Subtracting the percentage of detractors from the percentage of promoters yields the net promoter score (NPS).

For this study, we dichotomized the scores by defining the responses 10 or 9 as *promoters* of the web-based information and, therefore, considered satisfied, whereas we defined responses from 8 to 0 as *nonpromotors* and thereby not satisfied [16,17]. A Global Rating Scale (GRS) of satisfaction with the web app was applied (*What is your overall satisfaction with the website?*) with response options ranging from 10 (very satisfied) to 0 (not at all satisfied).

The Danish version of the Roland-Morris Disability Questionnaire (RMDQ) was used to measure LBP related function [29]. The Danish version consists of 23 items, and the sum score ranges from 0 (no disability) to 23 (maximum disability) [29].

Statistical Evaluation

Reliability was assessed by studying the difference between the OPSI at baseline and OPSI after a minimum of 7 days using a 2-way mixed-effect analysis of variance model with interaction for absolute agreement as the intraclass correlation coefficient of agreement ($ICC_{agreement}$) [30]. Where an $ICC_{agreement} >0.75$ can be interpreted as excellent, 0.4 to 0.75 indicate fair to poor, and values <0.4 indicate poor reliability [31]. Measurement error was assessed using the limits of agreement proposed by Bland and Altman [32]. The smallest detectable change ($SDC_{consistency}$) was calculated as follows:

$$SDC_{consistency} = 1.96 \times SD_{difference}$$

where $SD_{difference}$ is the SD of the difference between the test and retest. This equals the limits of agreement without systematic errors [33]. Construct validity was evaluated by hypothesis testing of the size and direction of correlations between the OPSI score for MyBack and the NPS, GRS, RMDQ, and OPSI score for Wikipedia about LBP using Spearman rank correlation coefficient. CIs were estimated by bootstrapping with 5000 replications. Correlations between 0.3 and 0.5 were considered weak, and correlations >0.5 were considered strong [34].

Hypothesis 1

We hypothesized that being categorized as *promotors* (scoring 9 or 10 on the NPS) would be positively and strongly (>0.5) correlated with higher OPSI scores (convergent validity).

Hypothesis 2

We hypothesized that respondent scores from 9 to 10 on the GRS would be positively and strongly (>0.5) correlated with higher OPSI scores (convergent validity).

Hypothesis 3

We hypothesized that RMDQ scores would be positively and weakly (<0.3) correlated with higher OPSI scores (discriminant validity).

Hypothesis 4

We hypothesized that the OPSI score for MyBack and the OPSI score based on the Wikipedia website would be negatively and weakly (−0.3 to 0) correlated with higher OPSI scores for OPSI based on MyBack (discriminant validity).

All analyses were conducted using Stata 16.0 (Stata Corp).

Results

A total of 150 participants were recruited between March 6, 2018, and May 10, 2019. The mean age of the participants was

48.7 (SD 12.9) years, and 67.3% (101/150) were women. Most (146/150, 97.3%) had experienced pain for >12 weeks with an average score of 8 (range, 0-10) for having an interest in finding new information on the internet (Table 1).

Table 1. Characteristics of the study population (N=150).

Baseline characteristics	Patients
Age (years), mean (SD)	48.7 (12.9)
Female, n (%)	101 (67.3)
Education level, bachelor's degree or more, n (%)	55 (36.7)
Employment status, n (%)	
Working full-time or part-time	66 (44)
On sick leave or leave of absence	39 (26)
Unemployed	8 (5.3)
Retired	37 (24.7)
Pain duration >12 weeks, n (%)	146 (97.3)
Pain intensity, 0-10, mean (SD)	5.4 (2.1)
RMDQ[a] score, mean (SD)	11.8 (5.1)
Contact with GP[b] about LBP[c] during the past 1 year, n (%)	146 (97.3)
Curious about finding new information, 0-10, median (IQR)	8 (7-9)
Use of internet about health, n (%)	
Daily	9 (6)
Weekly	48 (32)
Less than weekly	10 (6.7)
Monthly	39 (26)
Less than monthly	44 (29.3)
OPSI[d] home page, median (IQR)	20 (18-22)[e]
OPSI Wikipedia, median (IQR)	12 (8-15)[f]
OPSI home page retest, median (IQR)	20 (17-22)[g]

[a]RMDQ: Roland-Morris Disability Questionnaire.

[b]GP: general practitioner.

[c]LBP: low back pain.

[d]OPSI: Online Patient Satisfaction Index.

[e,f,g]Distribution of the OPSI scores were nonnormal, which were identified on histograms and q-norm visualizations.

A total of 70 randomly chosen participants were invited to the retest, 53 accepted whereas 39 answered *no* or *don't know* to the question about the stability of their satisfaction. These 39 were considered stable and were included in a retest analysis to evaluate the stability of the OPSI (Figure 3).

Figure 3. Flowchart of respondents. A total of 150 respondents were included in the validity analyses, and 39 were included in the reliability analysis. LBP: low back pain.
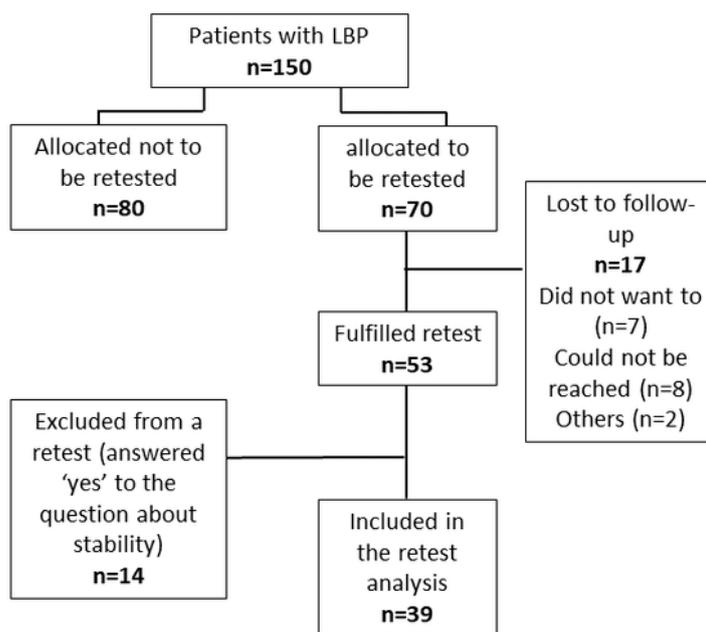
For the MyBack website, the OPSI score ranged from 724, and for the Wikipedia website, the entire OPSI score (0-24) range was used. The mean OPSI score of promoters for the MyBack website was 21.68 (95% CI 21.14-22.22), and the mean OPSI score of the nonpromoters was 18 (95% CI 17.25-18.75). The response rate was 100% for both websites. $ICC_{agreement}$ was estimated at 0.82 (95% CI 0.68-0.90). The $SDC_{consistency}$ was estimated at 4.71.

Limits of agreement were estimated to be –4.11 to 5.13 with a mean difference of 0.509 (95% CI –0.127 to 1.146). The mean difference is close to zero with the CI overlapping zero; hence, there is a negligible systematic difference between the baseline measurement and the retest measurement. In addition, the difference between baseline and follow-up measurements did not seem to depend on the level of satisfaction (Figure 4).

Figure 4. Comparison of test-scores and retest-scores. Online Patient Satisfaction Index for MyBack at the initial test and retested after 1 week.
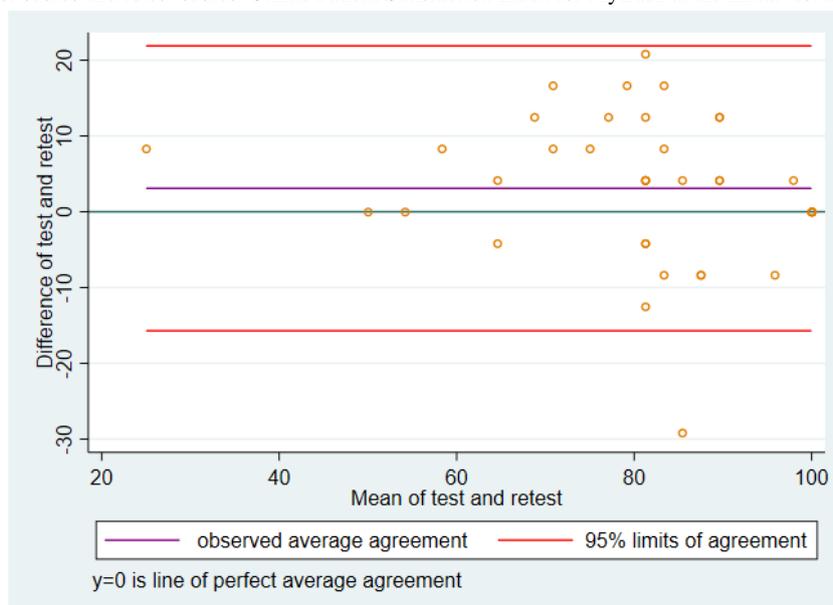
Two hypotheses were confirmed (Table 2).

Table 2. Hypothesis testing of construct validity.

Characteristics	Hypotheses	Correlations[a]	
		Expected, r	Observed[b], r (95% CI)
NPS[c]	Hypothesis A[d]	>0.5	0.58 (0.47 to 0.69)
GRS[e]	Hypothesis A	>0.5	0.70 (0.60 to 0.80)
RMDQ[f]	Hypothesis B[g]	<0.3	−0.14 (−0.30 to 0.02)
OPSI$_{wikipedia}$[h]	Hypothesis C[i]	−0.3 to 0	−0.04 (−0.20 to 0.13)

[a]Spearman rank correlation coefficient.

[b]Correlations from confirmed hypotheses.

[c]NPS: net promoter score.

[d]Scales are expected to measure the same construct. The correlation was expected to be positive and strong (>0.5).

[e]GRS: Global Rating Scale. [f]RMDQ: Roland-Morris Disability Questionnaire.

[g]Scales are not expected to measure the same construct. The correlation was expected to be positive and weak (<0.3).

[h]OPSI: Online Patient Satisfaction Index.

[i]Scales are expected to measure the same construct. The correlation is expected to be negative and weak (−0.3 to 0).

Discussion

Principal Findings

The OPSI is easy to apply as each of the 8 items is scored from 0 to 3, resulting in an overall score between 0 and 24 points. The index was found to have good face validity, excellent reliability, and good construct validity among our sample of participants with long-standing nonspecific LBP.

Recommendations to the OPSI Users

We recommend the use of proportional recalculation to convert the index score to a 0 to 100 scale, as it accounts for items with missing scores [35]. The OPSI was found to have excellent reliability in measuring satisfaction at 1 time point. Thus, it is suitable to measure satisfaction at 1 time point for 1 person or to compare satisfaction between groups. However, we did not test for responsiveness among participants experiencing a change in satisfaction over time; consequently, we cannot recommend using the OPSI to measure changes over time. Whether this can be recommended in the future, needs to be supported by the evaluation of responsiveness.

Limitations

The item *Do you trust the website?* is expected to capture both trust in the content and trust in security and handling of data. During development and face validity, splitting this item into two was not mentioned by participants. This may be explained by the high level of trust in health care authorities handling data in Denmark, but we do not know if this is the case. In other cultures, trust in the content and trust in the handling of data by the provider may be considered as 2 different issues and thereby require 2 items to be properly captured.

The development and evaluation of the OPSI was based on a formative model, which is 1 of the 2 conceptual frameworks, the other being a reflective model. A reflective model assumes interrelatedness between items and thereby item correlations. However, a formative model does not assume item correlations, and this is a limitation, as common statistical methods to describe the relationships between items and the construct were not applicable in this study [20].

The Sano Centre receives a new cohort of patients every 4 weeks, most of whom have long-standing LBP, so it is convenient and easy to ask patients to participate in the study because they stay at the center. For participants volunteering over social media, the assessors had to make an appointment to meet at a convenient location, which was not always easy and straightforward. Consequently, this led to an unequal distribution of respondents between the 2 sites, with (131/150, 87.3%) from the Sano Centre and (19/150, 12.7%) from social media. This is a limitation of the study as patients from the Sano Centre were expected to have more severe symptoms, require extended information, and thus score a larger difference between the 2 home pages. In contrast, participants recruited from social media may be more frequent users of Wikipedia for other information seeking and thereby more satisfied with the shorter wiki format. The size and direction of the Spearman coefficient changed for patients recruited from social media. However, this could be due to the small number of patients recruited from social media.

The time between the test and retest in the analysis of reliability was between 1 and 2 weeks, with a maximum of 4 weeks, which might have overestimated the reliability of OPSI. The short time duration may have influenced the participants' responses as some may recall their previous baseline response and, therefore, repeat the answer. However, people with back pain often experience changes in symptoms, and the short period between tests can be considered a strength when collecting data from participants with possible fluctuations in symptoms within a few weeks [36]. We applied a stability question to determine whether the participants were considered stable. It is a strength that TA or AR discussed with the participants to ensure that participants understood the question on change in satisfaction regarding MyBack and to ensure that participants with a change were excluded from the retest. We used the construct of *satisfaction* as a reflection of whether the included respondents in the reliability study were stable, and this is in fact treating

XSL•FO
RenderX

the items as a reflective model. However, the alternatives would either be to ask the respondents if they had changed each item or not to ask them at all. The first alternative would cause analytic problems determining who were stable and who were not, as some respondents probably would have changed on 1 or perhaps 2 items but not the rest. Setting a cutoff point on an acceptable number of stable items would be arbitrary and probably misleading, and we also question the feasibility of doing it this way. The second alternative of not asking about s*tability* is, in our opinion, unacceptable, as the potential to introduce a bias is high. We therefore opted for the solution of implementing a *global change question* as we believe this is the best of the 3 options introducing least bias regarding choosing stable respondents. However, assuming *don't know* as stable is potentially a limitation.

The use of NPS as a comparator for the construct validity of OPSI to measure satisfaction might be a weakness. The NPS uses a proxy to assess the customer's overall satisfaction with a service or product [17]. NPS is based on only 1 question with a reply option from 0 to 10, which is categorized into 3 groups: promoters (10-9), passive (8-7), and detractors (6-0) [17]. When estimating satisfaction, the middle group (passive) was excluded from the analysis [17]. This is, in our opinion, a weakness that can limit the usability of the NPS as a comparator for construct validity. We also applied the GRS as a comparator for construct validity using a response scale ranging from 0 to 10. The GRS may be inflated due to recall bias [37] and motivational effects [38] in longitudinal studies. Furthermore, transition scores seldom show an ideal pattern of association between baseline and follow-up measures [37]. However, we used baseline GRS to compare with a baseline score for the OPSI, and using a transition score to compare baseline scores has previously been found feasible [37].

Comparison With Prior Work

Although we have found no previous studies evaluating satisfaction with web-based information on LBP, other studies have reported satisfaction with web-based delivered health care information and educational information. Hence, a recent study of satisfaction with a gamified medical course among medical students in Thailand found a mean satisfaction of 9.02 (SD 1.11) out of 10 [39]. Equal high satisfaction was recorded in a study about the satisfaction of SMS text containing educational material for patients undergoing prostate biopsy in the United States. This study found a mean satisfaction of 4.5 (SD 0.9) out of 5 [40]. Furthermore, a high level of satisfaction was found in a study of satisfaction with telephone support in patients with type 2 diabetes [41,42]. They recorded a mean of 4/5 on all 21 items related to satisfaction [42]. These previous studies indicate a problem with ceiling effects when measuring satisfaction with health care interventions. The median score of OPSI used on a home page, which we considered good, was 20 [18,22] out of a maximum of 24 points. This indicates that satisfied patients tended to reply at the high end. Furthermore, this is in line with the development and use of the NPS, where only customers scoring 9 or 10 on a scale from 0 to 10 are considered satisfied

with a product or a service to a degree where customers are likely to buy a product again or reuse the service [16].

The reliability test indicated that a change of 4.71 points is necessary to preclude measurement error. If performing proportional rescaling of OPSI to a scale from 0 to 100, a change of approximately 20 points is necessary to preclude measurement error. In another study, applying a 0 to 100 score, the minimal detectable change was found to be 32.8 points on the group level when measuring the individualized quality of life in patients with LBP [43]. In a reliability and responsiveness study evaluating the minimal detectable change in LBP disability, questionnaires found lower or similar minimal detectable changes of 8.6 (RMDQ), 15 (Oswestry Disability Questionnaire), and 22 (the 36-item Short Form health surveys physical functioning scale) [44]. Consequently, compared with other measures applied to patients with LBP, OPSI has a similar good reproducibility.

Unanswered Questions and Future Research

The OPSI was evaluated in a population with severe and disabling LBP. Consequently, the index will benefit from further validation in other populations with less severe symptoms. In addition, the OPSI can be tested as a more generic tool to assess satisfaction without including dissatisfaction. King et al [45] constructed a survey with 22 items about the usability of phone apps to support physical activity and rated it on a 6-point Likert-type scale [45]. This scale was later adapted into a 21-item questionnaire rated for agreement or disagreement on a 5-point Likert scale [46]. A questionnaire developed more recently to measure user satisfaction with mobile health (mHealth) apps also applied a Likert scale—from *strongly disagree* to *strongly agree* [47]. However, this questionnaire did not meet the criteria for unidimensionality [47]. The OPSI is not challenged by the same potential limitation, as only satisfaction is measured and dissatisfaction is not.

We did not collect specific data regarding health literacy; however, either volunteering to participate on the internet or participating in a 4-week course at Sano indicates a health-interest in seeking health information among participants. This interest is supported by participants' baseline characteristics regarding the use of the internet to search for health information. Engaging all people, particularly those with low health literacy, in assessing health information can be a challenge [48]. Nevertheless, future research is needed to evaluate whether the OPSI can also be used by people with low health literacy. The index was developed and formally evaluated in Danish; however, it was translated into English using established guidelines [22]. The OPSI is potentially applicable in all Western countries, but future studies need to evaluate validity and reliability among other cultures and in other languages.

Conclusions

The OPSI showed good face validity, excellent reliability, and good construct validity and can be used when measuring satisfaction with the provision of information regarding LBP among people willing to access the internet for health information.

Conflicts of Interest

None declared.

References

1. Kent PM, Keating JL. The epidemiology of low back pain in primary care. Chiropr Osteopat 2005 Jul 26;13:13 [FREE Full text] [doi: 10.1186/1746-1340-13-13] [Medline: 16045795]
2. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet 2018 Jun 09;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
3. Hoy D, Brooks P, Blyth F, Buchbinder R. The Epidemiology of low back pain. Best Pract Res Clin Rheumatol 2010 Dec;24(6):769-781. [doi: 10.1016/j.berh.2010.10.002] [Medline: 21665125]
4. Bardin LD, King P, Maher CG. Diagnostic triage for low back pain: a practical approach for primary care. Med J Aust 2017 Apr 03;206(6):268-273. [doi: 10.5694/mja16.00828] [Medline: 28359011]
5. Lemeunier N, Leboeuf-Yde C, Gagey O. The natural course of low back pain: a systematic critical literature review. Chiropr Man Therap 2012 Oct 17;20(1):33 [FREE Full text] [doi: 10.1186/2045-709X-20-33] [Medline: 23075327]
6. Balagué F, Mannion AF, Pellisé F, Cedraschi C. Non-specific low back pain. Lancet 2012 Feb 04;379(9814):482-491. [doi: 10.1016/S0140-6736(11)60610-7] [Medline: 21982256]
7. Vlaeyen JW, Maher CG, Wiech K, Van Zundert J, Meloto CB, Diatchenko L, et al. Low back pain. Nat Rev Dis Primers 2018 Dec 13;4(1):52. [doi: 10.1038/s41572-018-0052-1] [Medline: 30546064]
8. Low back pain and sciatica in over 16s: assessment and management. National Institute for Health and Care Excellence. 2016. URL: https://www.nice.org.uk/guidance/ng59 [accessed 2021-10-15]
9. Bishop FL, Dima AL, Ngui J, Little P, Moss-Morris R, Foster NE, et al. "Lovely Pie in the Sky Plans": a qualitative study of clinicians' perspectives on guidelines for managing low back pain in primary care in England. Spine (Phila Pa 1976) 2015 Dec;40(23):1842-1850. [doi: 10.1097/BRS.0000000000001215] [Medline: 26571064]
10. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
11. Tustin N. The role of patient satisfaction in online health information seeking. J Health Commun 2010 Jan;15(1):3-17. [doi: 10.1080/10810730903465491] [Medline: 20390974]
12. Tailored on-line information and advice to patients with low back pain in general practice: a protocol for a semi-structured interview study (Part of the ADVIN Back Trial). Aalborg University. 2016. URL: https://vbn.aau.dk/ws/portalfiles/portal/241172247/Interview_study_2909_2016.pdf [accessed 2021-10-15]
13. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 09;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
14. Buchbinder R, van Tulder M, Öberg B, Costa LM, Woolf A, Schoene M, Lancet Low Back Pain Series Working Group. Low back pain: a call for action. Lancet 2018 Jun 09;391(10137):2384-2388. [doi: 10.1016/S0140-6736(18)30488-4] [Medline: 29573871]
15. Al-Abri R, Al-Balushi A. Patient satisfaction survey as a tool towards quality improvement. Oman Med J 2014 Jan;29(1):3-7 [FREE Full text] [doi: 10.5001/omj.2014.02] [Medline: 24501659]
16. Reichheld F, Markey R. The Ultimate Question 2.0: How Net Promoter Companies Thrive in a Customer-driven World. Massachusetts: Harvard Business Press; 2011.
17. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54, 124. [Medline: 14712543]
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
19. Riis A, Hjelmager DM, Vinther LD, Rathleff MS, Hartvigsen J, Jensen MB. Preferences for web-based information material for low back pain: qualitative interview study on people consulting a general practitioner. JMIR Rehabil Assist Technol 2018 Apr 02;5(1):e7 [FREE Full text] [doi: 10.2196/rehab.8841] [Medline: 29610108]
20. Henrica CW, Lidwine BM, Dirk LKC. Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology). Cambridge, UK: Cambridge University Press; 2011.
21. The patient handbook. Sundhed. URL: https://www.sundhed.dk/borger/patienthaandbogen [accessed 2021-10-15]

XSL•FO

RenderX

22. 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 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]

23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

24. Velkommen til Sano. Sano Danish Rheumatism Association. URL: https://sanocenter.dk [accessed 2021-10-15]

25. Min ryg: information om rygsmerter. Aalborg University. 2017. URL: https://vbn.aau.dk/en/publications/min-ryg-information-om-rygsmerter [accessed 2021-10-15]

26. Hjelmager DM, Vinther LD, Poulsen SH, Petersen LS, Jensen MB, Riis A. Requirements for implementing online information material for patients with low back pain in general practice: an interview study. Scand J Prim Health Care 2019 Mar;37(1):60-68 [FREE Full text] [doi: 10.1080/02813432.2019.1569413] [Medline: 30700191]

27. Riis A, Hartvigsen J, Rathleff MS, Afzali T, Jensen MB. Comparing satisfaction with a participatory driven web-application and a standard website for patients with low back pain: a study protocol for a randomised controlled trial (part of the ADVIN Back Trial). Trials 2018 Jul 25;19(1):399 [FREE Full text] [doi: 10.1186/s13063-018-2795-0] [Medline: 30045749]

28. Random.org. URL: https://www.random.org/ [accessed 2021-10-28]

29. Albert HB, Jensen A, Dahl D, Rasmussen MN. [Criteria validation of the Roland Morris questionnaire. A Danish translation of the international scale for the assessment of functional level in patients with low back pain and sciatica]. Ugeskr Laeger 2003 Apr 28;165(18):1875-1880. [Medline: 12772398]

30. Qin S, Nelson L, McLeod L, Eremenco S, Coons SJ. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. Qual Life Res 2019 Apr;28(4):1029-1033 [FREE Full text] [doi: 10.1007/s11136-018-2076-0] [Medline: 30547346]

31. Fleiss J. The Design and Analysis of Clinical Experiments. USA: John Wiley & Sons; 1999.

32. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [doi: 10.1177/096228029900800204] [Medline: 10501650]

33. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006 Oct;59(10):1033-1039. [doi: 10.1016/j.jclinepi.2005.10.015] [Medline: 16980142]

34. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988.

35. Kent P, Lauridsen HH. Managing missing scores on the Roland Morris Disability Questionnaire. Spine (Phila Pa 1976) 2011 Oct 15;36(22):1878-1884. [doi: 10.1097/BRS.0b013e3181ffe53f] [Medline: 21304432]

36. Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? BMC Musculoskelet Disord 2016 May 21;17:220 [FREE Full text] [doi: 10.1186/s12891-016-1071-2] [Medline: 27209166]

37. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002 Sep;55(9):900-908. [doi: 10.1016/s0895-4356(02)00435-3] [Medline: 12393078]

38. Aseltine RH, Carlson KJ, Fowler FJ, Barry MJ. Comparing prospective and retrospective measures of treatment outcomes. Med Care 1995 Apr;33(4 Suppl):AS67-AS76. [Medline: 7536868]

39. Phungoen P, Promto S, Chanthawatthanarak S, Maneepong S, Apiratwarakul K, Kotruchin P, et al. Precourse preparation using a serious smartphone game on advanced life support knowledge and skills: randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16987 [FREE Full text] [doi: 10.2196/16987] [Medline: 32149711]

40. Balakrishnan AS, Nguyen HG, Shinohara K, Au Yeung R, Carroll PR, Odisho AY. A mobile health intervention for prostate biopsy patients reduces appointment cancellations: cohort study. J Med Internet Res 2019 Jun 02;21(6):e14094 [FREE Full text] [doi: 10.2196/14094] [Medline: 31199294]

41. McGloin H, O'Connell D, Glacken M, Mc Sharry P, Healy D, Winters-O'Donnell L, et al. Patient empowerment using electronic telemonitoring with telephone support in the transition to insulin therapy in adults with type 2 diabetes: observational, pre-post, mixed methods study. J Med Internet Res 2020 May 14;22(5):e16161 [FREE Full text] [doi: 10.2196/16161] [Medline: 32406854]

42. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]

43. Løchting I, Grotle M, Storheim K, Werner EL, Garratt AM. Individualized quality of life in patients with low back pain: reliability and validity of the Patient Generated Index. J Rehabil Med 2014 Sep;46(8):781-787 [FREE Full text] [doi: 10.2340/16501977-1826] [Medline: 25117176]

44. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002 Jan;82(1):8-24. [doi: 10.1093/ptj/82.1.8] [Medline: 11784274]

45. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]

XSL•FO
RenderX

46.   Mummah SA, Mathur M, King AC, Gardner CD, Sutton S. Mobile technology for vegetable consumption: a randomized
       controlled pilot study in overweight adults. JMIR Mhealth Uhealth 2016 May 18;4(2):e51 [FREE Full text] [doi:
       10.2196/mhealth.5146] [Medline: 27193036]

47.   Melin J, Bonn SE, Pendrill L, Trolle Lagerros Y. A questionnaire for assessing user satisfaction with mobile health apps:
       development using Rasch measurement theory. JMIR Mhealth Uhealth 2020 May 26;8(5):e15909 [FREE Full text] [doi:
       10.2196/15909] [Medline: 32452817]

48.   Riis A, Rathleff MS, Hartvigsen J, Thomsen JL, Afzali T, Jensen MB. Feasibility study on recruitment in general practice
       for a low back pain online information study (part of the ADVIN Back Trial). BMC Res Notes 2020 Jan 10;13(1):24 [FREE
       Full text] [doi: 10.1186/s13104-020-4894-8] [Medline: 31924258]

Abbreviations

GRS: Global Rating Scale
ICC: intraclass correlation coefficient
LBP: low back pain
NPS: net promotor score
OPSI: Online Patient Satisfaction Index
REDCap: Research Electronic Data Capture
RMDQ: Roland-Morris Disability Questionnaire
SDCconsistency: smallest detectable change

XSL•FO

RenderX