

Development and Validation of the iDI: A Short Self-Rating Disability Instrument for Low Back Pain Disorders

Global Spine Journal
 2017, Vol. 7(2) 123-132
 © The Author(s) 2017
 Reprints and permission:
sagepub.com/journalsPermissions.nav
 DOI: 10.1177/2192568217694006
journals.sagepub.com/home/gsj


Cornelia Rolli Salathé, PhD¹, Achim Elfering, PhD¹, Alexander Tuschel, MD, MSc², Michael Ogon, MD, PhD², H. Michael Mayer, MD, PhD³, and Norbert Boos, MD, eMBA^{4,5}

Abstract

Study Design: Cross-sectional and longitudinal validation study.

Objective: Development and validation of a short, reliable, and valid questionnaire for the assessment of low back pain-related disability.

Methods: The iDI was created in a stepwise procedure: (1) its development was based on the literature and theoretical consideration; (2) outcome data were collected and evaluated in a pilot study; (3) final validations were performed based on an international multicenter spine surgery outcome study including 514 patients; (4) the iDI was programmed for a tablet computer (iPad) and tested for its clinical practicability.

Results: The final version of the iDI comprises of 8 simple questions related to different aspects of disability with a 5-point Likert-type answer scale. The iDI compared very well to the Oswestry Disability Index in terms of reliability and validity. The iDI was demonstrated to be suitable for data assessment on a tablet computer (iPad).

Conclusions: The iDI is a short, valid, and practicable tool that facilitates routine quality assessment in terms of low back pain-related disability.

Keywords

low back pain, disability, questionnaire, validation, electronic data assessment

Introduction

Routine quality measurement of spinal treatments and their documentation must become part of our daily clinical practice if we want to improve the care for our patients.¹ Additionally, we will be more and more scrutinized by health care stakeholders to justify the amount of money spent in an area of medicine that is predominately focusing on an improvement of health-related quality of life rather than long-term survival.² Although the need for a quality assessment in spinal surgery is realized for many years, we are far away from a routine outcome assessment in our daily care. The prerequisite for the widespread use of quality management in daily clinical practice relies not only on data validity and reliability but also on the simplicity of data collection and handling. Particularly, the basic principle of “less is more” applies in this context.³ Recent studies argued that the length of a questionnaire influenced patients’ response rates and influenced data quality.^{4,5} If we

can generate a small comprehensive outcome data set for each patient, it is of more value than a set of sophisticated data with a lot of missing values. The validity and usability of very short scales, for example, neurological stroke scales, is already demonstrated in other clinical disciplines such as neurology in an emergency setting, where every second counts. Collection

¹ University of Bern, Bern, Switzerland

² Orthopaedic Hospital Speising, Vienna, Austria

³ Orthopaedic Clinic Munich-Harlaching, Spine Center, Munich, Germany

⁴ University Hospital Balgrist, University of Zurich, Zurich, Switzerland

⁵ Prodosro Centre for Spinal Medicine, Zurich, Switzerland

Corresponding Author:

Cornelia Rolli Salathé, Department of Work and Organizational Psychology, Institute of Psychology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland.
 Email: cornelia.rolli@psy.unibe.ch



This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (<http://www.creativecommons.org/licenses/by-nc-nd/4.0/>) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

of missing data is very cumbersome and often not possible in retrospect.

The purpose of this project was to generate a valid, reliable, and simple outcome tool for daily clinical application using modern information technology with a minimum of questions. A further goal was to make data assessment as simple as possible. Due to the wealth of data collected in our multinational study in 3 large German-speaking spine centers, this article summarizes the results on the self-reported disability domain of this new outcome tool. In this study, we specifically focused on the question whether a score for self-reported disability can be generated that reduces completion time and improves data consistency without compromising reliability and validity when compared with the most widely used Oswestry Disability Index (ODI).^{6,7} In this context, the reader should see the development and evaluation of an outcome instrument, we coined iDI, Internet-suitable disability index.

Methods

Design

In analyzing a total of 514 patients, the cross-sectional and longitudinal validation includes a 3-step validation process. First, iDI questions were created according to recommendations from international guidelines⁸ regarding similar aspects of back-related disability as used by the ODI, which was considered as the gold standard. After development and refinements, 3 data collection waves were conducted to measure and validate the psychometric properties.

Measures

The iDI comprised 8 questions covering walking, sitting, standing, lifting, self-care, sleeping, social life, and traveling. The response format was a 5-point Likert-type scale with grouping answers “not at all,” “somewhat,” “moderately,” “strongly,” and “extreme.” A 5-point Likert-type scale was chosen because of its usability and its validity.⁹ For each of the 8 items, a maximum of 4 points was attainable (no disability = 0, maximum item-specific disability = 4). The total sum was divided by the total possible sum (ie, $n = 32$), multiplied by 100. In our multicenter study, ODI,^{7,10} Roland & Morris Disability Questionnaire (RMDQ),^{11,12} and EuroQol-5 Dimensions Index (EQ-5D)^{13,14} were used as reference scales. All instruments were validated in the German language.

Data Collection and Participants

Pilot Study. Participants included in the pilot project ($n = 118$) suffered from low back pain undergoing nonoperative as well as operative treatment in a spine center in Switzerland (Figure 1). Exclusion criteria were patients with pregnancy, tumor, infection, severe comorbidity compromising overall well-being (which was particularly marked by the treating physician as an activity-limiting comorbidity in the Sangha-Index¹⁵), and unwillingness to complete questionnaires. All participants

completed both questionnaires on spot (paper-and-pencil version): participants randomly received either the iDI or ODI. After returning the completed questionnaire, the participants were given the other questionnaire. Test-retest reliability was assessed by asking participants to fill out a second copy of the iDI or ODI, respectively, within 24 hours¹⁶ and to return the questionnaire the following day in a preaddressed envelope.

Multicenter Outcome Study. All participants ($n = 306$) in this multinational study suffered from degenerative lumbar spinal disorders attending 3 spine centers in Germany, Austria, and Switzerland (Figure 1). Inclusion and exclusion criteria were identical to the pilot study. After giving informed consent, all participants responded to an entire questionnaire set before treatment (baseline) and after 6 months (follow-up). Each participating center sent the completed questionnaires sets to our data assessment center. Six months later, a study coworker contacted all participants by telephone and asked if they still agreed to participate in the follow-up assessment as initially consented. When agreeing, the patients received the questionnaire by mail (paper-and-pencil version) with a preaddressed answer envelope. If patients did not respond within 2 to 4 weeks, they were contacted and reminded again by phone. Thereafter, no further attempt was made. The study protocol was approved by the institutional review boards of the participating hospitals.

Electronic Assessment of iDI. The programming of the electronic version was custom made for use on an iPad mini using iOS in a web-based mode. On the starting page, the hospital staff filled in patient identification number, sex, and date of birth. The next page contained general information on the use of the program. At this stage, the iPad was handed out to the patient who could then advance to the next page. All following pages displayed only one iDI question per page with the 5 response options. Only after responding to the question would the next page appear, allowing for a complete data set. A forward/backward option allowed correcting a question if needed. After answering the final question, the last page appeared with thanks for study participation. The data were automatically transferred to a server. A study coworker was available to assist patients if needed. The number of patients needing help as well as intervention reasons and questionnaire completion time were recorded. After sampling data electronically, the ODI paper-and-pencil version was handed out to the patient. Again, the completion time was recorded. For the test-retest assessment, the study participants received an email within 24 hours with a link to the iDI to be completed for retest measures.

Statistical Analysis

To gain an overview of the test qualities of the iDI, statistical analyses have been conducted in all 3 data sets. All analyses were performed using SPSS version 22.0 (IBM SPSS Inc, Chicago, IL).

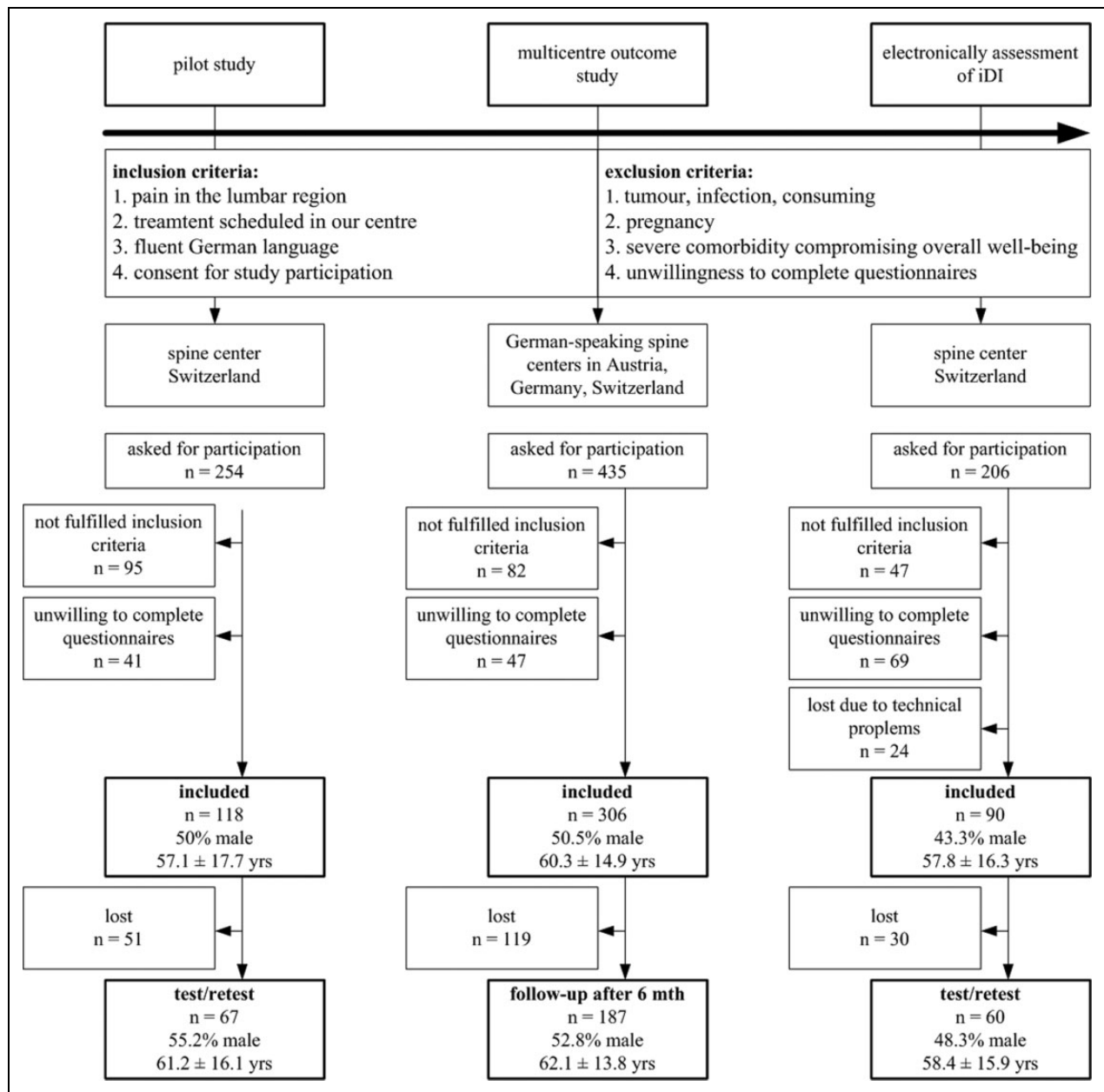


Figure 1. Flow of participants.

Data Quality. The paper-and-pencil versions of iDI and ODI were not considered in the analyses if more than 2 responses were missing. In case of missing 1 or 2 items, the denominator was adapted according to the total possible sum. Floor and ceiling effects were acknowledged if more than 15% of all patients reported highest or lowest score possible.¹⁷

Reliability Measures. Internal consistency was assessed using Cronbach's α .¹⁷ A value between 0.8 and 0.9 was considered as acceptable, and more than 0.9 as high.¹⁸ With a test-retest analysis, the extent to which the same results were obtained on repeated measures when no changes are expected have been analyzed. The differences in mean values for repeated trials were checked with the intraclass correlation coefficient

(ICC_{agreement}) and the standard error of the mean (SEM).¹⁷ If there is a perfect agreement between the 2 measures, the ICC is 1 and the SEM is 0. The SEM was also used to indicate the minimum detectable change, MDC_{95%}. At $P < .05$, it is calculated with the formula $1.96 \times \sqrt{2} \times \text{SEM}$ and represents the smallest score change that can be interpreted as a real change. Furthermore, expecting a simple structure with one underlying factor, we conducted an exploratory factor analysis with an orthogonal rotation method in form of Varimax.¹⁹

Validation Measures. The convergent validity as well as the divergent validity were assessed by testing the correlations between the ODI, iDI, and matching references scales.¹⁷ With regard to the convergent validity, thus the comparison between

analogue reference scales, ODI and iDI were compared to the RMDQ scale.^{11,12} In order to compare the scales that measure the opposite constructs (divergent validity), ODI and iDI were compared to the EQ-5D.^{13,14}

Usability. In terms of usability of the electronic version, completion time (in seconds) as well as iPad handling problems were assessed using a 5-point Likert-type scale (ie, 1 = without problems, without problems after short instruction, with little support, only with support, 5 = not at all). The completion times were compared using a paired *t* test, ICC_{agreement}, and SEM.¹⁷

Results

Overall, 514 patients have been included in all 3 studies. One-hundred and eighteen participated in the pilot study (test-retest 56.8%), 306 in the multicenter outcome study (follow-up 61.1%), and 90 in the electronically iDI assessment (test-retest 66.7%; Figure 1). Studies samples characteristics are shown in Table 1.

Missing Data and Normality of Score Distribution

Missing data occurred in all 3 studies (Table 2) and included most frequently the ODI Item 8 regarding sex life (31.8%). All other missing items were at random. Participants lost to test-retest or follow-up did not significantly differ from the participants included in the longitudinal analysis. Overall, the ODI scored lower showing more and higher floor effects than the iDI.

Item Quality and Reliability Measures

In total, 49 of 306 data sets had to be excluded because more than 2 item answers were missing in the multicenter outcome study. The item means were comparable for all 3 data collections (Table 3). With regard to item quality aspects, most item difficulties lie between a medium difficulty ($r_{diff} = 0.2-0.8$) indicating that most patients answered correctly to the items, and the discriminatory power of all items were found above $r = 0.5$.

Furthermore, Cronbach's α revealed that the strength of the relationship between the items within the test instrument were comparable for both questionnaires (Table 3). The extent to which the same results are obtained on repeated measures when no changes are expected have been analyzed with a test-retest analysis; that is, ICC_{agreement} demonstrated comparable values for both questionnaires, while the SEM revealed slightly higher measures for the ODI. MDC_{95%} exceeded 20% in the multicenter outcome study for both questionnaires. With regard to the exploratory factor analysis, a simple structure was detected as expected in both questionnaires (ODI and iDI). Explained variance is slightly higher for the iDI compared with the ODI. However, item loadings (Table 4) are comparable for both questionnaires. The simple structure was identified by 2 methods: first by scree plot and second as eigenvalues greater than 2.

Considering that a factor loading is a correlation coefficient, a factor loading above 0.6 (would equal a 0.6 correlation coefficient) is commonly accepted.¹⁸

Construct Validity

The hypothesis regarding convergent validity (ie, both questionnaires are expected to correlate moderately to highly in a positive manner to the RMDQ) was confirmed (Table 5). Similarly, the hypothesis regarding divergent validity (ie, both questionnaires are expected to correlate moderately in a negative manner to the EQ-5D) was confirmed.

Usability

With regard to completion time in the electronic assessment, a sample of $n = 63$ filled out the ODI in 137.0 seconds (mean, SD ± 53.3), while they answered to the iDI in 70.8 seconds (mean, SD ± 27.0 ; $t = 10.4$, $P < .001$). The SEM scored high when a low score is targeted (SEM = 24.6), while ICC demonstrated a low agreement (ICC_{agreement} = 0.22; 95% confidence interval = -0.17 to 0.52). The mean gain in time was 66.3 seconds (Table 6). Considering the problems handling the electronic device, the mean value of 2.1 (SD ± 1.4) and the median value of 1 demonstrated an unproblematic handling. Only 3 out of 63 patients could not handle the electronic device due to peripheral neuropathies and needed full assistance by our staff. Sixteen elderly patients required some help in the beginning but managed to get along afterwards.

Discussion

The prerequisite common to all well-functioning industries is competition. In healthy competition, product and service quality rise steadily, innovation leads to new and better approaches, uncompetitive providers eventually go out of market, and costs are driven down for better quality.²⁰ In medicine, the strategy for improvement in health care must be as well linked to a value-based approach, that is, health outcome achieved for the dollars spent.²¹⁻²³ However, outcome assessment in medicine particularly in spinal medicine is rather complex. The vast majority of spinal disorders are not life-threatening conditions but are associated with a compromised quality of life. Therefore, a simple (dichotomous) outcome endpoint is not applicable, for example, survived/not survived or prosthesis in situ/removed. Outcome assessment with regard to an improvement of health-related quality of life in spinal medicine predominately relies on 5 pillars, that is, reduction of pain, self-reported disability, and pain medication as well as an improvement of quality of life and work capacity.^{24,25} Reduction of pain and self-reported disability are the most important domains related to a good outcome.²⁴ The assessment of pain by a visual analogue or numeric scale (10 points) is widely recommended and meanwhile standard in most institutions.²⁶ In terms of the assessment of self-reported disability related to lumbar disorders, most centers use the ODI,⁶ RMDQ,¹¹ or the

Table 1. Demographics.

Variable	Pilot Study		Multicenter Outcome Study		Electronic iDI Assessment	
	Baseline (n = 118)	Retest (n = 67)	Baseline (n = 306)	Follow-up (n = 187)	Baseline (n = 90)	Retest (n = 60)
Age, mean ± SD	57.1 ± 17.7	61.2 ± 16.1	60.3 ± 14.9	62.1 ± 13.8	57.8 ± 16.3	58.4 ± 15.9
Female gender, n (%)	59 (50.0%)	30 (44.8%)	153 (49.5%)	88 (47.2%)	51 (56.7%)	31 (51.7%)
Marital status, n (%)						
Single			28 (9.1%)	21 (11.3%)		
Married/in partnership			206 (67.1%)	125 (66.6%)		
Divorced/widowed			72 (23.8%)	41 (22.1%)		
Data origin, n (%)						
Germany			55 (17.9%)	51 (27.4%)		
Austria			193 (62.9%)	90 (48.4%)		
Switzerland	118 (100%)	67 (56.8%)	59 (19.2%)	45 (24.2%)	90 (100%)	60 (66.7%)
Days of duration between data assessments, median, mean ± SD		6 days, 5.8 ± 4.3		180 days, 205.5 ± 45.5		2 days, 2.1 ± 3.2
Body mass index, mean ± SD			27.0 ± 4.8		26.1 ± 4.2	
Duration of LBP, n (%)						
Up to 4 weeks			9 (3%)			
5-12 weeks			11 (3.6%)			
3-12 month			35 (11.6%)			
More than 1 year			247 (81.8%)			
Frequency of LBP, n (%)						
Daily			199 (66.1%)	52 (28.0%)		
Several times/month to several times/week			77 (25.4%)	65 (46.8%)		
Few times/year or less			28 (8.5%)	66 (36.0%)		
Diagnosis, n (%)						
Disc herniation	20 (16.9%)		81 (26.4%)	62 (33.3%)	17 (18.9%)	
Spinal stenosis	26 (22.0%)		109 (35.5%)	62 (33.3%)	20 (22.2%)	
Degenerative spondylolithesis/stenosis	7 (5.9%)		50 (16.3%)	29 (15.6%)	17 (18.8%)	
Isthmic spondylolisthesis	4 (3.4%)		12 (3.9%)	7 (3.8%)	10 (8.8%)	
Degenerative disc disease	22 (18.6%)		63 (20.5%)	52 (28.0%)	7 (7.8%)	
Degenerative scoliosis	5 (4.2%)		17 (5.5%)	9 (4.9%)	7 (7.8%)	
Vertebral compression fracture	2 (1.7%)		22 (7.2%)	12 (6.5%)	6 (6.7%)	
Other	84 (71.2%)		38 (11.4%)	28 (14.1%)	29 (32.2%)	
Scheduled treatment, n (%)						
Discectomy			40 (13.0%)	33 (17.7%)	1 (1.1%)	
Decompression			79 (25.7%)	48 (25.8%)	21 (23.3%)	
Instrumented fusion			96 (31.3%)	56 (30.1%)	10 (11.1%)	
Dynamic instrumentation			6 (2.0%)	6 (3.2%)	1 (1.1%)	
Disc prosthesis			4 (1.3%)	3 (1.6%)	0	
Vertebroplasty/kyphoplasty			11 (3.6%)	5 (2.7%)	0	
Nonoperative care (physiotherapy, spinal injections, etc)			113 (36.8%)	64 (34.4%)	45 (49.9%)	
Highest education level, n (%)						
Compulsory education	81 (68.6%)		196 (63.8%)	125 (67.2%)		
High school	2 (1.7%)		34 (11.1%)	14 (7.5%)		
College/university	16 (13.6%)		62 (20.2%)	39 (21.0%)		
None	1 (0.8%)		7 (2.3%)	3 (1.6%)		
Current work ability, n (%)						
Able to work full-time			47 (15.3%)	49 (26.3%)		
Able to work part-time			38 (12.4%)	27 (14.5%)		
Unable to work due to LBP			64 (20.8%)	10 (5.4%)		
Unable to work due to other reasons			6 (2.0%)	4 (2.2%)		
Disability pension due to LBP			12 (3.9%)	10 (5.4%)		
Disability pension due to other reasons			11 (3.6%)	8 (4.3%)		
Retired			66 (21.5%)	41 (22.0%)		
Homemaker, student			24 (7.8%)	15 (8.1%)		
Not working due to unknown reasons			20 (6.5%)	19 (10.2%)		
No answer			19 (6.2%)	3 (1.6%)		

Abbreviation: LBP, low back pain.

Table 2. iDI Wording, Missing Data, End Effects, and Floor and Ceiling Effects^a.

Item	Pilot Study (n = 118)				Multicenter Outcome Study, Baseline (n = 306)				Electronic iDI Assessment, Baseline (n = 90)			
	Missing Data, n (%)	Mean (SD)	Low (%)	High (%)	Missing Data, n (%)	Mean (SD)	Low (%)	High (%)	Missing Data, n (%)	Mean (SD)	Low (%)	High (%)
ODI_1	0	2.5 (1.2)	9.3	3.4	0	2.7 (1.0)	2.3	2.9	1 (1.1)	2.4 (1.2)	4.7	6.3
ODI_2	0	1.0 (1.0)	39.8	0	0	1.3 (1.0)	29.1	1.0	1 (1.1)	0.8 (1.0)	50	1.6
ODI_3	1 (0.8)	2.5 (1.4)	8.5	4.3	1 (0.3)	2.8 (1.3)	5.2	5.2	3 (3.3)	2.1 (1.3)	8.1	1.6
ODI_4	0	1.3 (1.2)	25.4	1.7	0	1.8 (1.3)	19.9	1.3	1 (1.1)	1.0 (1.1)	43.8	4.7
ODI_5	1 (0.8)	2.0 (1.2)	16.1	0	1 (0.3)	1.9 (1.2)	14.8	0.7	2 (2.2)	1.6 (1.1)	23.8	1.6
ODI_6	2 (1.7)	2.3 (1.3)	5.9	3.4	1 (0.3)	2.8 (1.3)	4.9	4.9	3 (3.3)	2.0 (1.3)	9.5	1.6
ODI_7	0	1.5 (1.3)	24.6	2.5	2 (0.7)	1.6 (1.1)	12.5	1.0	1 (1.1)	1.6 (1.3)	12.5	6.3
ODI_8	51 (43.2)	1.5 (1.7)	41.8	3.0	88 (28.8)	2.1 (1.8)	28.4	11.9	21 (23.3)	1.1 (1.6)	56.8	6.8
ODI_9	0	2.0 (1.5)	23.7	2.5	2 (0.7)	2.4 (1.4)	15.8	2.6	2 (2.2)	1.8 (1.5)	28.6	1.6
ODI_10	2 (1.7)	1.8 (1.5)	21.6	6.0	6 (2.0)	2.3 (1.6)	11.7	11.7	2 (2.2)	1.8 (1.8)	25.4	12.7
iDI_2	0	1.4 (1.0)	21.2	0.8	1 (0.3)	1.4 (1.1)	24.9	2.0	0	1.0 (1.1)	42.2	0
iDI_3	1 (0.8)	2.4 (1.0)	4.3	9.4	1 (0.3)	2.7 (1.0)	2.3	20.0	0	2.5 (1.2)	7.8	21.1
iDI_4	0	2.2 (1.2)	8.5	11.0	3 (1.0)	2.4 (1.1)	6.6	13.9	0	1.9 (1.1)	13.3	8.9
iDI_5	0	2.0 (1.1)	9.3	6.8	1 (0.3)	2.1 (1.0)	6.9	5.9	0	1.8 (1.1)	13.3	5.6
iDI_6	0	2.2 (1.0)	3.4	9.3	5 (1.6)	2.6 (1.0)	3.0	16.9	0	2.1 (1.0)	7.8	7.8
iDI_7	0	1.6 (1.2)	22.9	7.6	15 (4.9)	1.7 (1.1)	17.2	5.2	0	1.4 (1.3)	27.8	7.8
iDI_9	0	1.7 (1.1)	16.1	2.5	16 (5.2)	1.8 (1.2)	17.6	6.6	0	1.8 (1.2)	18.9	8.9
iDI_10	0	1.9 (1.1)	11.0	5.9	4 (1.3)	2.4 (1.0)	5.3	12.6	0	2.2 (1.1)	7.8	12.2

^aLowest (%) indicates the percentage of patients scoring the lowest score possible (→ floor effects). Highest (%) indicates the percentage of patients scoring the highest score possible (→ ceiling effects). No ODI follow-up data assessed in the electronic iDI assessment.

Table 3. Reliability Measures.^a

	Questionnaire	n _{t1-t2}	Range	Mean (SD), t1	Mean (SD), t2	SEM	MDC	MDC (%)	ICC (95% CI)
Pilot study	Sum ODI	31	0-50	17.6 (10.1)	20.7 (10.3)	2.96	8.20	16.4	0.98 (0.97-0.99)
	Cronbach's α			0.92	0.93				
	Sum iDI	36	0-32	15.6 (6.4)	13.6 (6.0)	1.38	3.82	9.6	0.97 (0.93-0.98)
	Cronbach's α			0.88	0.87				
Multicenter Outcome Study	Sum ODI	191	0-50	20.7 (7.9)	13.4 (9.9)	8.24	22.82	45.6	0.50 (0.09-0.70)
	Cronbach's α			0.84	0.93				
	Sum iDI	185	0-32	16.9 (5.1)	10.7 (7.2)	6.28	17.40	43.5	0.48 (-0.01 to 0.70)
	Cronbach's α			0.79	0.92				
	Sum iDI	60	0-32	14.9 (5.8)	14.9 (6.9)	2.93	8.13	20.32	0.96 (0.93-0.97)
	Cronbach's α			0.80	0.87				

Abbreviations: SEM, standard error of mean; MDC, minimum detectable change; ICC, intraclass correlation coefficient; CI, confidence interval; ODI, Oswestry Disability Index.

^a“n” refers to the number of patients entering the test-retest analysis. “Range” refers to the range of the answering scale with 0 referring to no disability. “t1” refers to the baseline measure, t2 to the follow-up measure. t2 in the pilot study occurred within 3 days, in the outcome study within 180 days, and in the electronic assessment within 24 hours. No ODI follow-up data assessed in the electronic iDI assessment. “Cronbach’s α” represents the internal consistency—a value between 0.80 and 0.90 is regarded as acceptable, >0.9 as high internal consistency. “SEM_{agreement}” stands for the standard error of the mean with a smaller SEM indicating a more accurate assessment and therefore a better quality measure. “MDC_{individual}” is a responsiveness measure and stands for the smallest score change that can be interpreted as a real change and not measurement error (P < .05). “MDC%” refers to the MDC as percentage of maximum score. “ICC_{agreement}” is a reliability measure and stands for the intraclass correlation with a 2-way random effects model.

Table 4. Exploratory Factor Analysis.

	Pilot Study		Multicenter Outcome Study				Electronic iDI Assessment	
	Baseline		Baseline		Follow-up		Baseline	Follow-up
	ODI (n = 62)	iDI (n = 62)	ODI (n = 190)	iDI (n = 190)	ODI (n = 119)	iDI (n = 119)	ODI (n = 40)	iDI (n = 40)
2_personal care	.73	.76	.68	.65	.79	.80	.47	.54
3_lifting	.76	.78	.54	.66	.71	.81	.65	.63
4_walking	.65	.76	.65	.69	.75	.87	.81	.72
5_sitting	.71	.71	.58	.57	.69	.79	.52	.72
6_standing	.68	.68	.63	.66	.81	.82	.70	.79
7_sleeping	.74	.65	.48	.58	.73	.76	.36	.59
9_socialising	.81	.83	.71	.81	.86	.84	.72	.79
10_taveling	.78	.85	.77	.79	.91	.90	.78	.87
Eigenvalue	5.6	4.6	4.2	3.7	6.3	5.4	3.3	4.1
% of explained variance	56.3%	57.1%	42.0%	46.5%	63.2%	67.8%	41.5%	51.1%

Abbreviation: ODI, Oswestry Disability Index.

Table 5. Construct Validity^a.

	Multicenter Outcome Study			
	Baseline		Follow-up	
	P ^b RMDQ	P ^b EQ-5D	P ^b RMDQ	P ^b EQ-5D
ODI	0.60 (n = 340)	-0.70 (n = 311)	0.82 (n = 191)	-0.79 (n = 188)
iDI	0.57 (n = 313)	-0.65 (n = 293)	0.80 (n = 192)	-0.76 (n = 189)

Abbreviations: RMDQ, Roland-Morris Disability Questionnaire; EQ-5D, Euroqol-5 Dimensions Index; ODI, Oswestry Disability Index.

^aAll correlations are significant (P < .001, 2-tailed). Correlation coefficients: <0.3 = low; 0.3 to 0.6 = moderate; >0.6 = high.

^bSpearman’s rho.

North American Spine Society Score (NASS²⁷). The ODI is still the most frequently used instrument. Its main advantage is simplicity and the fact that it is available in multiple languages.

However, awkward phrasing as well as multiconceptual response categories exhibit its main disadvantages.²⁸ With regard to simplicity, the RMDQ is comparable to the ODI. Due

Table 6. Usability.^a

	n	Range	Mean (SD)	Median	Paired t test	SEM	ICC (95% CI)
Handling	63	1-5	2.1 (1.4)	1			
Without problems	33 (52.4%)						
Without problems after short instruction	8 (12.7%)						
With little support	7 (11.1%)						
Only with support	11 (17.5%)						
Not at all (due to physical handicap)	4 (6.3%)						
Time requirements ODI	63	30-270	137.0 (130.0)				
Time requirements iDI	64	30-180	70.8 (27.0)		10.4***	24.6*	0.22 (-0.17 to 0.52)*

Abbreviations: SEM, standard error of mean; ICC, intraclass correlation coefficient; CI, confidence interval; ODI, Oswestry Disability Index.

^a“n” refers to the number of patients entering the analysis. “Range” refers to the range of the answering scale with 1 referring to no problem (handling), to seconds (time), respectively. “Paired t test” value indicates if the mean time measures ODI versus iDI differ from a value of zero; hence, if they are alike. “SEM_{agreement}” stands for the standard error of the mean, with a smaller SEM indicating a more accurate assessment and therefore a better quality measure. “ICC_{agreement}” is a reliability measure and stands for the intraclass correlation with a 2-way random effects model. *P < .05, **P < .01, ***P < .001; two-tailed.

to its dichotomous response categories, only few information can be gathered per item.²⁸ The NASS is useful in measuring back pain, disability, and neurogenic symptoms. Its limitations include a rather narrow validation range.²⁸ When used in the context of routine quality management rather than scientific assessments of competing treatment methods, the questionnaire must be short, simple but valid and reliable, self-administered, and allow for easy electronic data sampling and management. Regarding these characteristics, the cited questionnaires leave room for improvements. In our study, we intended to develop an outcome tool facilitating disability assessment and allow for easy use on any touchscreen computer.

In 2011, the World Health Organization characterized disability as problems in human functioning that can be located in 3 areas: impairments, activity limitations, and participation restriction.²⁹ Hence, problems can be located in body function or structure, can cause difficulties in executing a task, or can hinder a person’s involvement in life situations. Measurement of disability is clearly different from measurement of pain and these 2 constructs should not be confounded. The items walking, sitting, standing, and lifting are part of most back pain disability assessment tools.^{28,30} The ODI, for example, encompasses additionally self-care, sleeping, social life, traveling, and sex life, which cover activity limitations and participation restriction. However, the latter item frequently leads to nonresponses. Many study participants do not answer this question because it compromises their intimacy and/or because of cultural and religious reasons. In our study, an average number of 31.8% did not answer this specific question across all data assessment waves. From a methodological point of view, missing data that are not random but reflect disagreement or reluctance to answer should be avoided because of result bias.³¹ Rather than reinventing the wheel, we opted to use similar aspects of disability than the ODI but omitted pain and sex life for the aforementioned reasons.

With regard to reliability, validity, and usability measures, the iDI is somewhat superior to the ODI. In contrast, Cronbach’s α of the ODI is slightly higher than the Cronbach’s α of iDI. However, an in-depth comparison of the ODI versus iDI

item quality, validity, and responsivity needs to be addressed in a subsequent study.

Unlike other ODI alterations,³⁰ not only the number of items but also the wording was thoroughly changed with the aim to create equidistance since disability measurements should have sufficient gradations.³² The drawback of a semantic differentiation of item expression is that intervals between statements cannot be presumed equal—not within the clusters nor between its items.³³ Therefore, the response options are often not equidistant (equal interval data level). In order to reach equidistance in response options, a scale should be symmetrical, have odd response options, and have a neutral center of the scale. Otherwise, response options cannot be equidistant.³⁴ As response format, we therefore opted for a 5-point Likert-type scale. This further leads to improved usability. The mean gain in completion time of more than a minute is superior to the ODI. With regard to the gain in completion time, 66 seconds may not seem a clinically relevant issue. However, the mean duration for answering the ODI is twice as long, which might make a difference in the patients’ perception and willingness to complete a questionnaire considering that disability is only one of the domains being assessed. In this context, a recent randomized controlled trial argued that the length of a questionnaire influenced patients’ response rates.⁴

New outcome tools should restrict nonresponses by software solution that will result in improvement of accurate data analysis.³⁵ In our iPad outcome tool application, nonresponses were restricted by the function that the questionnaire could not be completed unless all questions are answered. If software is applied that eliminates missing values, it would be unethical to force responses if the respondent is not willing to do so, for example, answering a question on sex life. Therefore, we included only items that fulfil this criterion. As expected, only very few patients (n = 3/63) had problems using the electronic device because of a physical handicap. All other individuals could handle the device without significant problems and/or assistance. This seems to be in line with other reports on data assessment with electronic devices in the elderly.^{36,37}

Some limitations have to be mentioned and discussed. In the beginning of the study, we encountered a technical problem

due to loss of Internet connectivity that resulted in program shutdowns. However, this was overcome by improving the WLAN connectivity in the treatment center.

When compared with the ODI, the iDI shows a potential upward bias due its 5-point Likert-type scale instead of a 6-point ordinal scale. Hence, the total possible sum is 32 (iDI) instead of 50 (ODI). This distortion is only relevant when calculating percentages instead of using absolute values and can be overcome by using a correction factor. When a comparison to the ODI is intended, the iDI percentage value can be divided by 1.56. Using this correction factor, the differences in percentages between the 2 scores are substantially less than the minimal clinically important relevant difference for lumbar spine surgery patients.³⁸

Comparing the electronic iDI with the paper-and-pencil ODI version is a point of methodological criticism. The ongoing discussion about comparability of online versus paper assessments are mixed with studies reporting favorable results in both directions.³⁹ Yet most of the studies included in the systematic review and meta-analysis have found no differences.³⁹

Furthermore, participation rates differed in the different phases of the project. The data of step 1 (participation rate: 46%) and step 3 (participation rate: 43%) were collected in a different way than in step 2 (international multicenter outcome study, participation rate of 70%). In steps 1 and 3, patients were approached by a research assistant during their waiting time for a medical consultation in a spine clinic. In step 2, patients were recruited by the treating surgeons. This highlights that a good physician-patient relationship enhances the willingness to respond to outcome instruments.

Finally, in order to clear up possible cultural barriers with regard to the original version of the iDI in German and this article in English, a cross-cultural research design should be included in a further study to evaluate possible differences between the iDI in German and the iDI in English.

Overall, this study demonstrates that the iDI compares very well to the “gold standard” ODI regarding item quality, reliability, and validity measures in patients with spinal disorders. The comparability is demonstrated in 3 different longitudinal German-speaking samples using a paper-and-pencil version as well as an electronic version. Three results highlight its strengths. First, although the iDI has less items than the ODI, reliability as well as validity measures are comparable. Second, factor analysis repeatedly revealed higher item loadings, as well as higher percentages of explained variance for iDI. Third, we demonstrated the simple application and programming of the iDI on a tablet computer (iPad) in a way that missing data are omitted improving overall data quality. In this context, the outcome tool iDI exhibits advantageous features and can be seen as an alternative for the assessment of self-reported disability.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this

article: Michael Ogon reports other from AOSpine, other from ISASS, personal fees from Medtronic, other from DePuySynthes, other from Globus Medical, other from EuroSpine, other from Vivantes, outside the submitted work; Norbert Boos reports grants from AOSpine, during the conduct of the study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by a grant from AO Spine Europe (Grant # ORD 08-03).

References

1. Boos N. Outcome assessment and documentation: a friend or foe? *Eur Spine J.* 2006;15(suppl 1):S1-S3.
2. Deyo RA, Mirza SK. The case for restraint in spinal surgery: does quality management have a role to play? *Eur Spine J.* 2009;18:331-337.
3. Baxendale S. When less is more. Data reduction in the prediction of postoperative outcome. *Epilepsy Behav.* 2014;31:219.
4. Sahlqvist S, Song Y, Bull F, Adams E, Preston J, Ogilvie D. Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised controlled trial. *BMC Med Res Methodol.* 2011;11:62.
5. Galesic M, Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opin Q.* 2009;73:349-360.
6. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976).* 2000;25:2940-2953.
7. Mannion A, Junge A, Fairbank JC, Dvorak J, Grob D. Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity. *Eur Spine J.* 2006;15:55-65.
8. 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;25:3186-3191.
9. Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ.* 2010;15:625-632.
10. Mannion A, Junge A, Grob D, Dvorak J, Fairbank JC. Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. *Eur Spine J.* 2006;15:66-73.
11. 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 (Phila Pa 1976).* 1983;8:141-144.
12. Wiesinger GF, Nuhr M, Quittan M, Ebenbichler G, Wöfl G, Fialka-Moser V. Cross-cultural adaptation of the Roland-Morris Questionnaire for German-speaking patients with low back pain. *Spine (Phila Pa 1976).* 1999;24:1099-1103.
13. Kind P. The EuroQoL instrument: an index of health-related quality of life. *Qual Life Pharmacoecon Clin Trials.* 1996;2:191-201.
14. Von der Schulenburg J, Claes C, Greiner W, Uber A. Die deutsche Version des EuroQoL-Fragebogens. *Z Gesundheitswissenschaften.* 1998;6:3-20.
15. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess

- comorbidity for clinical and health service research. *Arthritis Rheum.* 2003;49:156-163.
16. Fairbank J, Davies J, Couper J, O'Brien J. The Oswestry Disability Low Back Pain Questionnaire. *Physiotherapy.* 1980;66:271-273.
 17. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34-42.
 18. Bortz J, Döring N. *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler.* Heidelberg, Germany: Springer Verlag; 2006.
 19. Kline P. *An Easy Guide to Factor Analysis.* New York, NY: Routledge; 2014.
 20. Porter ME, Teisberg EO. Redefining competition in health care. *Harv Bus Rev.* 2004;82:64-76.
 21. Porter ME. A strategy for health care reform—toward a value-based system. *N Engl J Med.* 2009;361:109-112.
 22. Porter ME. Value-based health care delivery. *Ann Surg.* 2008; 248:503-509.
 23. Porter ME, Teisberg EO. How physicians can change the future of health care. *JAMA.* 2007;297:1103-1111.
 24. Haefeli M, Elfering A, Aebi M, et al. What comprises a good outcome in spinal surgery? A preliminary survey among spine surgeons of the SSE and European spine patients. *Eur Spine J.* 2008;17:104-116.
 25. Mannion AF, Elfering A, Staerke R, et al. Outcome assessment in low back pain: how low can you go? *Eur Spine J.* 2005;14: 1014-1026.
 26. Haefeli M, Elfering A. Pain assessment. *Eur Spine J.* 2006; 15(suppl 1):S17-S24.
 27. Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American Spine Society lumbar spine outcome assessment Instrument: reliability and validity tests. *Spine (Phila Pa 1976).* 1996;21:741-748.
 28. Katz JN. Measures of adult back and neck function: The North American Spine Society (NASS) Lumbar Spine Outcome Assessment Instrument, Neck Disability Index, Oswestry Low Back Pain Disability Index, Quebec Back Pain Disability Scale, and Roland-Morris Low Back Pain Measure. *Arthritis Care Res.* 2003;49(5 suppl):S43-S49.
 29. World Health Organization. *World Report on Disability 2011.* http://www.who.int/disabilities/world_report/2011/en/. Accessed October 19, 2015.
 30. Longo UG, Loppini M, Denaro L, Maffulli N, Denaro V. Rating scales for low back pain. *Br Med Bull.* 2010;94:81-144.
 31. Patrician PA. Multiple imputation for missing data. *Res Nurs Health.* 2002;25:76-84.
 32. Vernon H, Lawson G. Development of the Headache Activities of Daily Living Index: initial validity study. *J Manipulative Physiol Ther.* 2015;38:102-111.
 33. Boone HN, Boone DA. Analyzing Likert data. *J Extension.* 2012; 50:1-5.
 34. Jäger R, Bortz J. Rating scales with smilies as symbolic labels—determined and checked by methods of psychophysics. *CiteSeer.* <http://psychologie.biphaps.uni-leipzig.de/fechner/generalinfo/PDFs/RJaeger.pdf>. Accessed February 4, 2017.
 35. Free C, Phillips G, Watson L, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS Med.* 2013; 10:e1001363.
 36. Reijula J, Rosendahl T, Reijula K, Roilas P, Roilas H, Sepponen R. A new method to assess perceived well-being among elderly people—a feasibility study. *BMC Geriatr.* 2009;9:55.
 37. Bruder C, Blessing L, Wandke H. Adaptive training interfaces for less-experienced, elderly users of electronic devices. *Behav Inform Technol.* 2014;33:4-15.
 38. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *Spine J.* 2008;8:968-974.
 39. Muehlhausen W, Doll H, Quadri N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health Qual Life Outcomes.* 2015;13:167.