

A. F. Mannion
A. Junge
J. C. T. Fairbank
J. Dvorak
D. Grob

Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity

Received: 5 May 2004
Accepted: 1 August 2004
Published online: 26 April 2005
© Springer-Verlag 2005

Part 2 of this article can be found at <http://dx.doi.org/10.1007/s00586-004-0816-z>

A. F. Mannion · A. Junge
J. Dvorak · D. Grob
Schulthess Klinik, Lengghalde 2,
8008 Zürich, Switzerland

A. F. Mannion (✉)
Department of Rheumatology
and Institute of Physical Medicine,
University Hospital,
Zürich, Switzerland
E-mail: anne.mannion@kws.ch
Tel.: +41-44-3857584
Fax: +41-44-3857590

J. C. T. Fairbank
Nuffield Orthopaedic Centre, Oxford, UK

Abstract Patient-orientated assessment methods are of paramount importance in the evaluation of treatment outcome. The Oswestry Disability Index (ODI) is one of the condition-specific questionnaires recommended for use with back pain patients. To date, no German version has been published in the peer-reviewed literature. A cross-cultural adaptation of the ODI for the German language was carried out, according to established guidelines. One hundred patients with chronic low-back pain (35 conservative, 65 surgical) completed a questionnaire booklet containing the newly translated ODI, along with a 0–10 pain visual analogue scale (VAS), the Roland Morris Disability Questionnaire, and Likert scales for disability, medication intake and pain frequency [to assess ODI's construct (convergent) validity]. Thirty-nine of these patients completed a second questionnaire within 2 weeks (to assess test–retest reliability). The intraclass correlation coefficient for the test–retest reliability of the questionnaire was 0.96. In test–retest, 74% of the individual questions were answered identically, and 21% just one grade higher or lower. The

standard error of measurement (SEM) was 3.4, giving a “minimum detectable change” ($MDC_{95\%}$) for the ODI of approximately 9 points, i.e. the minimum change in an individual's score required to be considered “real change” (with 95% confidence) over and above measurement error. The ODI scores correlated with VAS pain intensity ($r=0.78$, $P<0.001$) and Roland Morris scores ($r=0.80$, $P<0.001$). The mean baseline ODI scores differed significantly between the surgical and conservative patients ($P<0.001$), and between the different categories of the Likert scales for disability, medication use and pain frequency (in each case $P<0.001$). Our German version of the Oswestry questionnaire is reliable and valid, and shows psychometric characteristics as good as, if not better than, the original English version. It should represent a valuable tool for use in future patient-orientated outcome studies in German-speaking lands.

Keywords Low-back pain · Condition-specific questionnaires · Cross-cultural adaptation · Reliability · Validity

Introduction

Condition-specific, self-administered questionnaires are essential for clinical assessment and research. In the field of spine outcomes research, “back-specific function” is one of the five domains recommended for inclusion in such patient-oriented assessments [8, 14]. Although a number of different questionnaires exist to assess function, most state-of-the-art reviews [9, 14] recommend either the Oswestry Disability Index (ODI [17, 18]) or the Roland Morris Questionnaire (RM [35]). Both of these are widely used, have been extensively tested, show similarly good psychometric properties, and are applicable in a wide variety of settings [34]. The number and nature of both the items and their response categories differ between the two questionnaires, but their purpose is generally the same; namely to indicate the extent to which a person’s activities of daily living are disrupted or restricted by low-back pain (LBP). It has been suggested that the RM may be better suited to settings in which patients have mild to moderate disability and the ODI to situations in which patients may have persistent severe disability [34], although studies in which the two questionnaires have been directly compared in a range of patient types are actually rare [30, 37].

Both the RM and the ODI have been translated into various languages and used in a number of research studies (see Refs. [18, 32, 34] and <http://medweb.bham.ac.uk/roh/odi/index.htm>). The availability and consistent use of established questionnaires in different languages facilitates the collection of reliable data in international multi-centre studies, enables the pooling of data from different studies for use in meta-analyses, allows valid comparison of the attribute in question between different nations, cultures, etc., and reduces the number of individuals who must be excluded from a study because they are not fluent in the native language. However, these benefits can only be enjoyed if the instrument in question has undergone adequate cross-cultural adaptation for the chosen target language prior to its use in a research study. A number of reviews have highlighted the importance of adhering to a standardised procedure for the cross-cultural adaptation of self-assessment questionnaires in order to ensure true equivalence with the original, and have offered useful practical guidelines on the approach that should be taken [4, 24].

A systematic search of the literature and discussion with the originators of the ODI revealed that, although a German version of the ODI has been used in a number of published research studies (see Ref. [18]), there is no version in the peer-reviewed literature that has been through the recommended cross-cultural adaptation and subsequent validation processes (this was further confirmed by personal communication with

various authors who had been cited using the ODI).¹ Thus, there is no guarantee that any of the German versions currently in use demonstrate the necessary equivalence with the original English, or indeed with each other.

The aims of this study were to carry out a cross-cultural adaptation of the ODI version 2.1 [34] for use with German-speaking patients and to investigate the psychometric properties of the German version (test–retest reliability, construct validity) in a large group of patients undergoing conservative or surgical treatment for chronic LBP within the Spine Unit of an orthopaedic hospital in Switzerland. In Part 2 of this series, the responsiveness of the ODI to surgical treatment is examined [31].

Materials and methods

The Oswestry Disability Index

The ODI was originally developed in 1980 [17] and then modified slightly in 1989 [1] to produce the version that is recommended for general use today (ODI version 2.1²) [34]. The ODI version 2.1 is a self-administered questionnaire that consists of ten items to assess the extent of the patient’s back pain and difficulty in carrying out nine different activities of daily life: personal care, lifting, walking, sitting, standing, sleeping, sex life, social life and travelling. The questionnaire is completed in reference to the patient’s functional status “today”. Each item is scored from 0 to 5, with higher values representing greater disability. The total score is multiplied by 2, and normally expressed as a percentage (in the rest of the paper, this percentage will simply be referred to as “the ODI score” and discussed in terms of points (0–100), to avoid confusion when discussing the percentage difference in score (as a mathematical expression) on repeated occasions).

Translation and cross-cultural adaptation

The translation and cross-cultural adaptation of the original English version of the ODI into German was carried out in accordance with previously published guidelines [4, 24]. These guidelines describe the process currently recommended by the American Academy of Orthopaedic Surgeons (AAOS) Outcomes Committee.

¹A recent review [32] erroneously stated that Basler et al. [2] had produced a validated German version of the ODI. That this version was simply a direct translation of the English version, and never subject to formal validation procedures, was confirmed after making contact with the authors.

²The English version of ODI 2.1 is reprinted in full in Ref. [34].

Translation and synthesis

Two native German speakers (T-1, T-2) carried out independent translations of the ODI Version 2.1 from English to German. The translators had different educational and job profiles. T-1 was familiar with the concepts being examined, the clinical content of the questionnaires and with other disability questionnaires for LBP patients. T-2 was a professional translator, who also had an administrative job within the orthopaedic hospital; thus, she was not familiar with the specific concept being investigated, but did have an appreciation of the level of understanding of the typical patient that would ultimately be completing the questionnaire (the “naive translator”; [4, 24]). The different profiles of the two translators assured good agreement and accuracy with the original English version in terms of both the clinical content and the appropriateness of the terminology. The two translations were compared with one another and with the original English version. After discussing any discrepancies that had arisen, a consensus was finally reached and the two versions were synthesised to form one common German version, T-12.

Back-translation

Two native English speakers (one American and one British) with German as their second-language (BT-1, BT-2) carried out a back-translation of the German version (T-12) into English. Both back-translators were considered bilingual, according to the definition of Deyo et al. [13] (people with one language living at least one year in another country with another language). Neither of the back-translators was familiar with the subject matter of the questionnaire; both were blind to the English original; and each carried out their translation independently, working together with a native German-speaking colleague (not one of the original translators T-1 and T-2). A third bilingual person (native English, German as second language) compared the two back-translations with each other and with the original-questionnaire and highlighted any conceptual errors or gross inconsistencies in the content of the translated versions, in preparation for the expert committee meeting.

Expert committee

An expert committee was formed consisting of one of the translators, one of the back-translators, one psychologist/methodologist, one clinician (rheumatologist), one orthopaedic surgeon [the originator of the English version of the ODI (J.C.T.F.)], and one clinical research scientist. The group examined the translations, the back-translations, and the notes made in carrying out/ comparing the translations, and consolidated these to produce a “pre-final” version of the German ODI. The task of this expert committee was to assure semantic and

idiomatic equivalence (i.e. to check for ambiguous words or inappropriately translated colloquialisms) and experiential and conceptual equivalence (i.e. to address any peculiarities specific to the cultures examined) between the German and English versions of the questionnaire. For all parts of the questionnaire (instructions, items, and response options) consensus was always found between the members of the committee. All stages of the translation process, and any discrepancies, problems, or difficulties encountered, were documented in written form.

Test of the pre-final version

A heterogeneous group of approximately 15 people (mostly patients with back problems in the waiting rooms of the hospital, but also some visitors and employees of the hospital) were given the pre-final version of the ODI questionnaire to complete. After completion, they were briefly interviewed in order to check what they thought was meant by each question and the chosen response. They were also asked for their general comments on the questionnaire (layout, wording, ambiguities, ease of understanding, etc.). All the findings from this phase of the adaptation process (face validity of the questionnaire) were evaluated by the work-group before the final German version of the ODI was produced and subject to further psychometric testing.

Assessment of the psychometric properties of the German version of the ODI version 2.1

Questionnaire battery

In each of the sub-studies described below, the patients were asked to complete a small questionnaire booklet, which contained the German version of the ODI and a series of other questions/questionnaires intended to assess the ODI’s construct (convergent) validity. A slightly modified version of the validated German RM disability questionnaire [16] was included in the booklet. The RM enquires as to whether back pain hinders the performance of 24 activities of daily living (today), with possible responses of “yes” and “no” [35]; the score for the RM ranges from 0 to 24 points. We modified the German version of Exner et al. [16] in relation to the formulation “because of/due to my back pain...” found in each item: the original Swiss German version used “wegen meinem Rücken” (i.e. in the dative form, which is the common colloquial form used in Switzerland), and we modified this to use the grammatically correct genitive form “wegen meines Rückens”. The questionnaire booklet also contained a 0–10 visual analogue scale for back/leg pain intensity (VAS_{pain}) in the last week, Likert scale questions about use of pain medication (“never” to “always”), pain frequency (“never” to “always”) and

degree of disability due to the back problem (“none” to “very severe”), and a 0–10 visual analogue scale for overall general health status (VAS_{health}).

Patients

Comparison of ODI baseline values in surgical and conservative patients

One hundred and sixty-six German-speaking patients with chronic LBP (> 3 months) were identified from the hospital computer system. The ‘surgical’ diagnoses included spinal stenosis, herniated disc, failed back/revision surgery, spondylolisthesis, “degenerative disease with chronic LBP”. The conservative patients had non-specific back pain, “degenerative disease with chronic LBP”, or failed-back syndrome. The patients were sent a study information sheet, an informed consent form and a questionnaire booklet (described above) by post. The information sheet explained that their voluntary participation would involve them completing the questionnaire booklet enclosed and another one after approximately 1 week (for all conservative patients and some of the surgical patients) and/or 6 months after their operation (surgical patients only; see Part 2 [31]).

Of the 166 patients that were invited to participate in the study, 105 (63%) returned the first, baseline questionnaire. However, 5/105 questionnaires had been completed *after* an intervention and these were therefore discarded from any further analyses. Thus, 100 baseline questionnaires were available for comparing the ODI scores of conservative patients [$n=32$: 13 men, 19 women; mean (SD) age 49 (17) years] and surgical patients [$n=68$: 34 men, 34 women; mean (SD) age 55 (13) years].

Test–retest reliability of the ODI

For the test–retest reliability analysis, 54 patients [all conservative ($n=32$) and the first 22 surgical patients] from whom a valid first questionnaire was received were immediately sent a second questionnaire and asked to complete and return it as soon as possible. Of these patients, 49/52 (94%) returned the second questionnaire. Five questionnaires were not valid for the test–retest reliability analyses: four patients had undergone treatment between the two questionnaires and one had given his replies for the two occasions by means of different colours on the same questionnaire. Five questionnaires that had been completed more than 2 weeks apart were excluded from the analysis. Thus, the data from 39 patients (54% conservative, 46% surgical) were available for the test–retest reliability analyses [25 women, 14 men; mean (SD) age 42 (16) years]. The study was approved by the local Ethics Committee.

Statistical analysis

Floor and ceiling effects were determined by calculating the number of individuals obtaining the lowest (0) or highest (100) ODI scores possible in the baseline questionnaire. This indicates the proportion of patients for whom it would not be possible to measure a meaningful deterioration or improvement of their condition, respectively (as they are already at the extreme of the range). A perhaps more relevant indication of these effects was also given by the proportion of individuals obtaining a score within the limits of the “minimum detectable change” MDC (see below) at the two ends of the scale.

Internal consistency was assessed with Cronbach’s alpha, using the data from the baseline questionnaires ($n=100$). Cronbach’s alpha indicates the strength of the relationship between all the items within the test instrument; i.e. it examines the extent to which the instrument measures a single trait or characteristic [5].

Construct validity indicates the extent to which the instrument’s scores relate to those of other instruments in the manner expected, i.e. whether the instrument really measures the intended construct. Tests of “correlational” and “contrasted measures” convergent validity were used for this purpose. The relationships between the ODI and the 0–10 VAS_{pain} , and between ODI and RM were examined with regression analysis (using all baseline questionnaires, $n=100$). Analysis of variance was used to examine the significance of differences in mean ODI scores between the five individual categories of the Likert scale disability questionnaire. The Spearman Rank correlation coefficient was used to determine the correlation (non-parametric) between the ODI scores and the Likert scale data for disability, with the latter treated as ordinal data. Differences between the mean scores of the conservative and surgical patients were examined using independent *t*-tests.

Test–retest reliability indicates the extent to which the same results are obtained on repeated administrations of a given instrument when no change is expected. The differences in mean values for the repeated trials were examined using paired *t*-tests. The intraclass correlation coefficient (ICC) and the standard error of measurement (SEM) (or “typical error of measurement”) for the repeated trials, each with their 95% confidence intervals, were also determined [27]. The SEM was used to indicate the $MDC_{95\%}$ for the ODI; i.e. the degree of change required in an individual’s score, in order to establish it (with a given level of confidence) as being a “real change”, over and above measurement error [3]. At the 95% confidence level, this is defined as $1.96 \times \sqrt{2} \times SEM$ which is equivalent to $2.77 \times SEM$. Statistical significance was accepted at the $P < 0.05$ level. No adjustments were made for multiple testing [33].

Results

Cross-cultural adaptation of the ODI

The German version of the ODI version 2.1 is shown in the Appendix. A few noteworthy difficulties arose during its development. (1) Translation of “travel”. The strict equivalent of the word in German (“reisen”) would not be used to describe short journeys, e.g. “travelling to receive treatment”. As such, we used the word “reisen” for the “longer journeys”, “unterwegs sein” for the “somewhat shorter journeys”, and “Fahrten machen” for the “minimal duration journeys”. (2) Translation of the distances able to be walked. The original English version used 1 mile, quarter of a mile, 100 yards. The strict metric equivalents of these would be 1.6 km, 400 m, 91 m, and in our first version we rounded these down/up to 1.5, 0.5 km, 100 m. However, during the pre-test phase it was considered that stating such a relatively long distance as accurately as “1.5 km” might cause uncertainty (people may feel they have to measure it out exactly, in order to know whether they can do it). As such, after discussion with other European colleagues regarding their perceptions of walking distances, and examination of other similar questionnaires, we opted to use “1–2 km”, “0.5 km” and “100 m”, respectively. (3) The “personal care” statements in the original English ODI read (in increasing order of disability) “I can look after myself but it is very painful” and then “It is painful to look after myself and I am slow and careful”. It seemed illogical to have “very painful” in the statement that corresponded to “less disability”, and just “painful” in the statement reflecting “greater disability”. Thus, in the German version, we omitted the word “very” from the first of these two statements (to yield English equivalents of “I can look after myself but it is painful” and “It is painful to look after myself and I am slow and careful”). (4) During the pre-testing phase, some of the volunteers were confused by the phrase “...but it gives/causes extra pain” in questions 3, 6, 8–10 (translated by us for the pre-testing phase version as “...aber es verursacht zusätzliche Schmerzen”). In the original, this phraseology was supposed to address the notion of activities that cause an increase of the “baseline level of pain”, and as such we modified the German version to “aber die Schmerzen werden dadurch stärker” for the final version.

Missing data, normality of score distribution at baseline

Nineteen patients (out of 100) failed to answer the “sex life” question and one patient didn’t answer the “social life” question. The scores for the group were normally distributed and no individual scored the worst or best possible score (no floor/ceiling effects). However, one patient had a baseline score of just 6, and three further

patients a score of 8, each of which would lie within the range of the $MDC_{95\%}$ (nine points; see below).

Internal consistency of the ODI at baseline

Cronbach’s alpha (internal consistency) for the ODI, determined from the data for the whole group of patients with at least one questionnaire at baseline ($n=100$) was 0.90. The item-total scale correlations ranged from 0.55 (for “walking”, question 4) to 0.82 (for “travelling”, question 10). When the conservative and surgical patients were analysed separately, the Cronbach’s alpha values were similar to those for the whole group (0.90 and 0.86, respectively) as were all the item-scale correlations, with the exception of the item “walking” for the conservative patients (correlation 0.09).

Construct validity: relationship between ODI values and other parameters at baseline

Using the data from the whole group of patients at baseline ($n=100$), the two disability scores, ODI and RM, showed a highly significant correlation with one another ($r=0.80$, $P<0.001$) (Fig. 1). However, using the regression equation of ODI on RM ($Y = 7.157 + 2.503 \times X$), extrapolation to the minimum RM score (=score of 0 points) and the maximum RM score (=score of 24 points) yielded ODI scores of 7 and 70%, respectively. Thus, whilst the minimum disability was quite similar with each instrument, maximum disability as judged by the RM did not equate to maximum disability on the ODI. Similar correlation coefficients and regression equations were obtained when the conservative and surgical groups were analysed separately [correlation between ODI and RM for the surgical patients $r=0.77$, conservative $r=0.74$ (each $P<0.001$)].

VAS_{pain} showed a highly significant correlation with both the ODI (all patients $r=0.78$, surgical $r=0.72$, conservative $r=0.79$; each $P<0.001$) and the RM (all patients $r=0.72$, surgical $r=0.67$, conservative $r=0.67$; each $P<0.001$).

The mean ODI and RM scores for each of the Likert categories of “pain-related disability in everyday activities” (1, none; 2, minimal; 3, moderate; 4, considerable; 5, very severe) are shown in Fig 2. No patients reported being in category 1 (no disability). Whilst the mean RM scores increased consistently by 3–4 points for each increasing category of disability (from 2 through to 5), those for the ODI were similar for categories 2 and 3, and then showed a steady increase in going from category 3 to 5.

When the Likert scale data for disability was treated as an ordinal variable (from 1 to 5) it showed a Spearman Rank correlation coefficient of 0.70 with ODI and 0.60 with RM (each $P<0.001$).

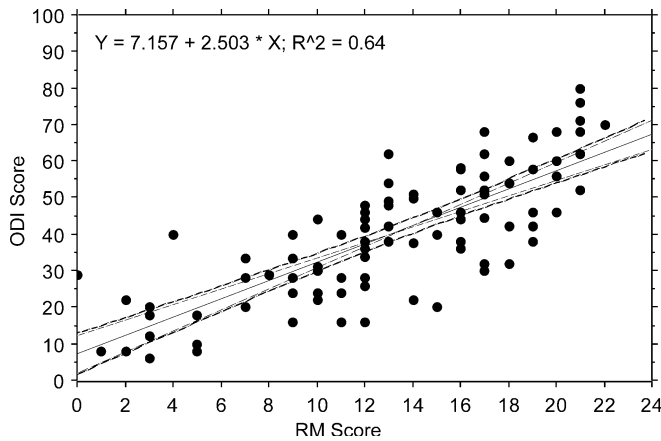


Fig. 1 Relationship between ODI and RM scores for all patients at baseline. (Regression plot, with 95% CI for the mean and the slope)

In advancing through the categories for “pain frequency” and “frequency of pain medication use”, there was generally a corresponding increase in the mean RM and ODI scores, though not always by a systematic amount (Fig. 2).

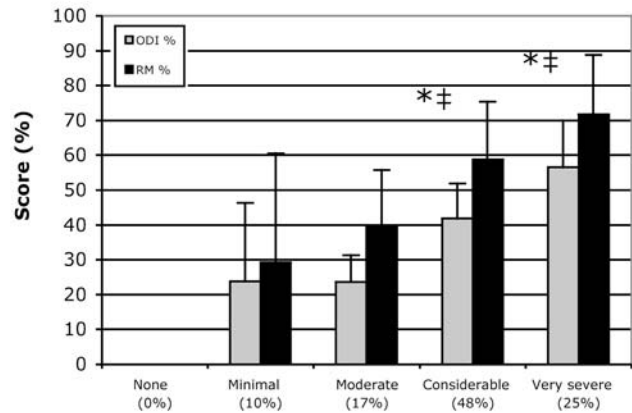
Construct validity: comparison of ODI baseline values in surgical and conservative patients

The mean values for disability (ODI and RM), pain intensity and general health are shown separately for the surgical and conservative patients in Table 1. The disability scores for the surgical patients were approximately 150% those of the conservative, for both the ODI and RM ($P < 0.001$). The VAS_{pain} scores for the surgical patients were approximately 130% those of the conservative patients ($P < 0.001$). The VAS_{health} did not differ significantly between the two groups. The conservative and surgical patients also differed significantly in relation to their distributions of answers to the Likert scale questions regarding pain frequency and overall disability ($P = 0.001$), with the surgical patients reporting a worse state of affairs in each case; medication use showed the same tendency, but did not differ significantly between the groups ($P = 0.19$) (exact data not shown).

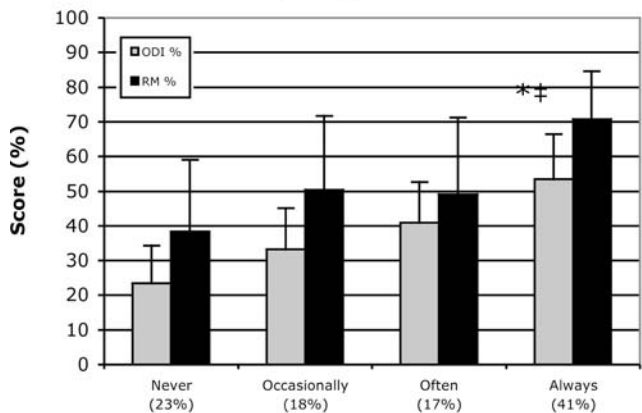
Test–retest reliability of the ODI

The mean (SD) time between completions of the two questionnaires was 6.0 (3.0) days (range 2–14 days). With 39 patients each answering ten questions in the ODI, the maximum number of questions that the group could have answered on each occasion was 390. Ten people failed to answer one of the questions on each occasion (nine the “sex life” question, one the “social life” question), which meant that a total of 380 questions were answered twice. Of these 380 questions, 73.7% were answered identically on the two occasions; 21.5% were answered within one

LBP-related disability in everyday activities



Pain medication: frequency of use



Frequency of pain

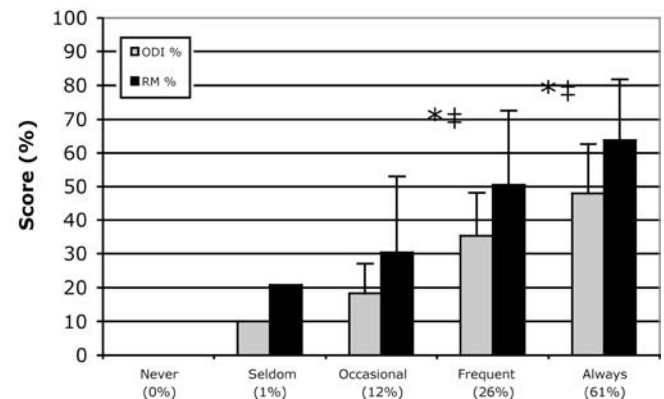


Fig. 2 Mean (SD) ODI and RM scores for each of the Likert scale categories for disability, pain frequency and pain medication. The percentage value given under each category on the x-axis shows the proportion of patients in each category. RM scores are expressed as a percentage value, for direct comparison with the ODI scores (RM, 24 points = 100%). The symbols indicate significant difference in the mean scores (* ODI, ‡ RM) between the two x-axis categories on either side of the symbol

Table 1 Questionnaire scores for the conservative and surgical patients

Variable	Conservative (<i>n</i> = 32) mean (SD)	Surgical (<i>n</i> = 68) mean (SD)	Comparison conservative versus surgical (<i>P</i> value)
ODI	30.5 (17.0)	45.4 (14.9)	0.001
Roland Morris	10.1 (5.5)	15.0 (4.7)	0.001
VAS _{pain}	5.2 (2.1)	6.8 (1.9)	0.001
VAS _{general health}	5.5 (2.4)	4.6 (2.6)	0.113

category higher or lower; 3.4% within two categories; and just 1.6% within three categories.

The κ values for the individual questions ranged from 0.48 (pain intensity) to 0.73 (walking); the weighted κ values ranged from 0.59 (lifting) to 0.85 (walking).

There was no significant difference between the mean ODI scores on the two test occasions (Table 2). The ODI showed excellent test–retest reliability, as evidenced by the high ICC for the two test occasions (0.96) and the low SEM (3.4) (Table 2). The reliability was comparable to that of the RM, and both the disability questionnaires were somewhat more reliable than VAS_{pain} or VAS_{health} (Table 2).

When the SEMs for the two disability questionnaires were expressed in relation to their corresponding mean values for the group, the percentage error was almost identical for each (ODI 9.2%, RM 10.5%) (Table 2).

With an SEM of 3.4, the MDC_{95%} for the ODI would be approximately nine points. This represents the minimum difference in an individual's score required to state with 95% confidence that “real change” is responsible for the difference, as opposed to just measurement error (“noise” in the system). The corresponding MDC_{95%} values for the RM and the 0–10 VAS_{pain} would be approximately four points and two points, respectively.

In response to a Likert scale question asking the patients to rate their LBP-related disability (from 0, “none”, to 5, “very severe”), 27/39 (69%) patients remained in the same category on the two occasions, 4/39 (10%) rated themselves as one grade more disabled and 8/39 (21%) as one grade less disabled. No patients rated themselves more than one grade apart on the two occasions. When the test–retest reliability for the various instruments was analysed for just the 27/39 patients who

rated themselves in the same global category for disability on the two occasions, the ICCs were almost identical to those for the whole group (ODI = 0.94; RM = 0.90; VAS_{pain} = 0.88; VAS_{general} = 0.85).

Discussion

Translation/cross-cultural adaptation of the ODI

The aim of the present study was to cross-culturally adapt the ODI, for use with German-speaking patients in Switzerland, and to examine the psychometric properties of the German version produced. The process of translating and back-translating the English ODI was carried out strictly in accordance with established guidelines [4, 24], in an attempt to produce a reliable and valid adaptation of the questionnaire that would show a high degree of agreement with the original English version. The English and German languages are closely related and the translation could be carried out with almost literal equivalence for many of the terms used. The only real problems encountered were in describing the appropriate wording for the “travel” items, and in finding the most appropriate metric equivalent for the walking distances (imperial units are used in the English original).

Only minor modifications to the initial translation were required after the back-translations, and after testing the pre-final version.

The adaptation of the ODI was carried out and tested in a group of patients living in the German-speaking part of Switzerland, and thus the psychometric characteristics of the questionnaire in other German-speaking lands (Austria and Germany), or lands with German-speaking immigrants, is not necessarily guaranteed. However, there are few grammatical or semantic differences in the use of the written language amongst the German-speaking countries; the main linguistic difference between these German-speaking lands/regions concerns the different dialects of the spoken language. Further, in putting together this German version of the ODI, we paid special attention to choosing words that were in common everyday use in both Germany and Switzerland; two of the people involved in the adaptation (one translator, and one back-translator's assistant) were of German nationality. Thus, we believe that the current version can most likely

Table 2 Questionnaire scores for the two repeated assessments < 2 weeks apart (*n* = 39)

Parameter/questionnaire	Trial 1 mean (SD)	Trial 2 mean (SD)	Comparison trial 1 versus trial 2 <i>P</i> value	ICC (95% CI)	SEM (95% CI)	SEM as percentage of mean score ^a
Oswestry (score range, 0–100)	36.2 (15.9)	37.4 (16.2)	0.12	0.96 (0.93–0.98)	3.4 (2.8–4.4)	9.2
Roland Morris (score range, 0–24)	12.4 (5.3)	12.3 (5.6)	0.86	0.94 (0.90–0.97)	1.3 (1.1–1.7)	10.5
VAS pain intensity (score range, 0–10)	6.0 (2.1)	6.1 (2.1)	0.47	0.86 (0.75–0.92)	0.8 (0.6–1.0)	13.2
VAS general health (score range, 0–10)	4.9 (2.5)	5.2 (2.1)	0.17	0.85 (0.73–0.92)	0.9 (0.7–1.1)	17.8

^aMean of trials 1 and 2

be used without difficulty in other German-speaking European countries. However, according to the guidelines of Guillemin et al. [24], this would require verification from other research groups working in these lands.

Internal consistency of the ODI

The internal consistency of the German ODI was examined using Cronbach's alpha, an item correlation test that reflects the homogeneity of all the items. For the whole scale the Cronbach's alpha was 0.90, which is higher than the majority of coefficients previously reported (0.76 [28]; 0.77 [20]; 0.83 [10]; 0.94 [23]). Cronbach's alphas greater than 0.8 are generally recommended for psychometric scales [36], although for individual patient assessments in the clinical situation, an alpha coefficient of at least 0.9 is recommended [7]. Thus, in this sense, the German ODI should be suitable not only for group analyses but also for the interpretation of individual scores. It was interesting to note that, in the sub-group of conservative patients, the item "walking" showed no correlation with the whole scale score and the mean value was considerably lower for this item compared with the others; this suggests that either walking does not represent such a typical LBP-associated impediment in this particular group of patients, or that our relative scaling for this item was not appropriate in relation to that of the other items in the questionnaire. This may need to be looked at in further detail.

Construct validity of the ODI

The German ODI showed good construct validity, as assessed with a variety of different techniques. Convergent validity was examined by investigating the strength of the relationship between the ODI and various other indices of disability (RM score and Likert scale ratings) and LBP intensity. The ODI showed a somewhat higher correlation with the RM ($r=0.80$) than has been previously reported ($r=0.66$ (LBP patients with radicular pain), $r=0.72$ (low back sprain) [30]; $r=0.73$ [10]). However, it was noticed that extrapolation of the regression equation for the relationship between ODI and RM, to a maximum RM score (24 points), yielded an ODI score of just 70 points. This tends to confirm previous suggestions that ODI may be more appropriate than RM for use in patients with a greater degree of disability: at high levels of disability the ODI may still show change when RM scores are maximal [34]. The mean ODI score for each of the Likert scale categories of disability (none, minimal, moderate, considerable, very severe) generally increased with increasing disability, although there was no difference in the mean ODI scores of the patients with "minimal disability" and those "moderate disability". Nonetheless, this may have been the result of having fewer patients in these two categories

(27% of the whole group) than in the higher disability categories, generating less accurate mean values; there was also no significant difference between the mean RM scores for these two lowest disability categories.

The German ODI correlated somewhat better with VAS_{pain} ($r=0.78$) than it has been shown to in previous studies {e.g. 0.62 (chronic LBP) [22]; 0.37–0.39 [15]; 0.39 (acute LBP), 0.52 (chronic LBP) [23]}, although in the latter study, VAS_{pain} itself was not particularly reliable, especially for the acute patients [23]. The correlation between LBP-related disability and pain ratings is expected to be good, but it should not be *extremely* high, otherwise it would suggest that the two instruments are carrying identical information. The validity of the ODI was also assessed using "contrasted measures" [36], and the ODI was shown to be capable of discriminating between groups of conservative and surgical patients in terms of the severity of their LBP disability.

Test-retest reliability of the ODI

A questionnaire's test-retest reliability indicates how stable it is over a predefined time interval, with the ICC indicating the strength of agreement between measurements recorded on two occasions. ICCs greater than 0.7 are generally considered acceptable [19]. In the present study, the ODI showed excellent reliability, with an ICC of 0.96 for a maximum time between the repeated measures of 2 weeks [mean (SD) time interval, 6 (3) days]. This time interval was chosen because, compared with shorter time intervals, it minimises the possible memory effect and provides a more realistic view of the degree of score change that may occur for non-specific reasons (random error) [6]. Although five individuals did ultimately return their second questionnaire, which they had completed from 17 to 25 days after the first, in the interests of homogeneity we chose not to include these data in the final analysis (though the data from these five patients was actually remarkably reliable: ICC for ODI 0.93; RM 0.91). The ICC reported in the present study for the German ODI was higher than those reported in previous studies, even though many of the latter used shorter test-retest intervals (median 4 days, ICC=0.91 [29]; 1 week, 0.83 [22], 0.94 [15]; 2 days, 0.88 [23], 2 weeks, 0.94 [26]; 4 weeks, 0.90 [21], 6 weeks, 0.84 [11]). The weighted κ values for the individual questions were moderate to very good, ranging from 0.59 (for the "lifting" question) to 0.85 (for "walking"); no κ values for the ODI have previously been reported in the literature for comparison [32]. Although the κ values reported in the present study indicate that some of the individual questions could be used reliably in sub-analyses, as done in previous studies [12], the questionnaire was originally intended only to provide a sum-score for disability.

The SEM is another expression of the error associated with repeated measurements, and for the ODI the SEM

was 3.4 points. The SEM is used to indicate the MDC for the scale, i.e. the degree of change required in a given individual's measures, in order to establish it (with a given level of confidence) as being a "real change", over and above measurement error. At the 95% confidence level, the $MDC_{95\%}$ for the ODI was 9 points—slightly lower than the values previously reported in the literature (13 points [21], 11 points [23], 10 points [25], 17 points [11]). An $MDC_{95\%}$ of 9 points indicates that, if an individual were to record a change of more than 9 points after a given intervention, then the odds would be 19 to 1 (i.e. 95% confidence level) that this represented a "real change". Some authors argue that the 95% confidence limits are too stringent to use as a threshold for deciding that a real change has occurred, and recommend using $1.5 \times SEM$ or $2.0 \times SEM$ (rather than $2.77 \times SEM$) [27]. In this case, the corresponding odds of measuring a real change are still 5 to 1 (83% confidence level), and 11 to 1 (92% confidence level), respectively (as opposed to 19 to 1 for the 95% confidence level). Using the SEM for the ODI, and the formulae given above, readers can decide for themselves the odds they wish to accept in making decisions regarding "real" individual change, and can calculate the MDC accordingly.

Conclusion

We have produced a German version of the Oswestry questionnaire that is reliable and valid, and that shows

psychometric characteristics as good as, if not better than, the original English version. The questionnaire represents a valuable tool for use in future patient-orientated, spine outcome studies in German-speaking lands. It will be used as the official German version for the Spine Society of Europe's "Spine Tango" Spine Surgery Registry, and is available from their website (<http://www.spinnetango.com>) or from the ODI website (<http://medweb.bham.ac.uk/roh/odi/index.htm>).

Acknowledgements The authors would like to thank Gordana Balaban, Simon Smit and Katrin Knecht for the administration of the questionnaires; Gabi Umbricht for the translation; and Geoff Klein and Sue Huber for the back-translations. The study was funded by the Schulthess Klinik Research Funds.

Appendix

At the start of the whole questionnaire booklet the patients are reminded that back problems can lead to back pain and/or leg pain, and that the questions should be answered in relation to these symptoms.

Bitte füllen Sie diesen Fragebogen aus. Er soll uns darüber informieren, wie Ihre Rücken- (oder Bein-) Probleme Ihre Fähigkeit beeinflussen, den Alltag zu bewältigen. Wir bitten Sie, jeden Abschnitt zu beantworten. Kreuzen Sie in jedem Abschnitt nur die Aussage an, die Sie heute am besten beschreibt.

Abschnitt 1: Schmerzstärke

- ₀ Ich habe momentan keine Schmerzen
- ₁ Die Schmerzen sind momentan sehr schwach
- ₂ Die Schmerzen sind momentan mässig
- ₃ Die Schmerzen sind momentan ziemlich stark
- ₄ Die Schmerzen sind momentan sehr stark
- ₅ Die Schmerzen sind momentan so schlimm wie nur vorstellbar

Abschnitt 2: Körperpflege (Waschen, Anziehen etc.)

- ₀ Ich kann meine Körperpflege normal durchführen, ohne dass die Schmerzen dadurch stärker werden
- ₁ Ich kann meine Körperpflege normal durchführen, aber es ist schmerzhaft
- ₂ Meine Körperpflege durchzuführen ist schmerzhaft, und ich bin langsam und vorsichtig
- ₃ Ich brauche bei der Körperpflege etwas Hilfe, bewältige das meiste aber selbst
- ₄ Ich brauche täglich Hilfe bei den meisten Aspekten der Körperpflege
- ₅ Ich kann mich nicht selbst anziehen, wasche mich mit Mühe und bleibe im Bett

Abschnitt 3: Heben

- ₀ Ich kann schwere Gegenstände heben, ohne dass die Schmerzen dadurch stärker werden
- ₁ Ich kann schwere Gegenstände heben, aber die Schmerzen werden dadurch stärker
- ₂ Schmerzen hindern mich daran, schwere Gegenstände vom Boden zu heben, aber es geht, wenn sie geeignet stehen (z.B. auf einem Tisch)
- ₃ Schmerzen hindern mich daran, schwere Gegenstände zu heben, aber ich kann leichte bis mittelschwere Gegenstände heben, wenn sie geeignet stehen
- ₄ Ich kann nur sehr leichte Gegenstände heben
- ₅ Ich kann überhaupt nichts heben oder tragen

Abschnitt 4: Gehen

- ₀ Schmerzen hindern mich nicht daran, so weit zu gehen, wie ich möchte
- ₁ Schmerzen hindern mich daran, mehr als 1–2 km zu gehen
- ₂ Schmerzen hindern mich daran, mehr als 0.5 km zu gehen
- ₃ Schmerzen hindern mich daran, mehr als 100 m zu gehen

- ₄ Ich kann nur mit einem Stock oder Krücken gehen
₅ Ich bin die meiste Zeit im Bett und muss mich zur Toilette schleppen

Abschnitt 5: Sitzen

- ₀ Ich kann auf jedem Stuhl so lange sitzen wie ich möchte
₁ Ich kann auf meinem Lieblingsstuhl so lange sitzen wie ich möchte
₂ Schmerzen hindern mich daran, länger als 1 Stunde zu sitzen
₃ Schmerzen hindern mich daran, länger als eine halbe Stunde zu sitzen
₄ Schmerzen hindern mich daran, länger als 10 Minuten zu sitzen
₅ Schmerzen hindern mich daran, überhaupt zu sitzen

Abschnitt 6: Stehen

- ₀ Ich kann so lange stehen wie ich möchte, ohne dass die Schmerzen dadurch stärker werden
₁ Ich kann so lange stehen wie ich möchte, aber die Schmerzen werden dadurch stärker
₂ Schmerzen hindern mich daran, länger als 1 Stunde zu stehen
₃ Schmerzen hindern mich daran, länger als eine halbe Stunde zu stehen
₄ Schmerzen hindern mich daran, länger als 10 Minuten zu stehen
₅ Schmerzen hindern mich daran, überhaupt zu stehen

Abschnitt 7: Schlafen

- ₀ Mein Schlaf ist nie durch Schmerzen gestört
₁ Mein Schlaf ist gelegentlich durch Schmerzen gestört
₂ Ich schlafe auf Grund von Schmerzen weniger als 6 Stunden
₃ Ich schlafe auf Grund von Schmerzen weniger als 4 Stunden
₄ Ich schlafe auf Grund von Schmerzen weniger als 2 Stunden
₅ Schmerzen hindern mich daran, überhaupt zu schlafen

Abschnitt 8: Sexualeben (falls zutreffend)

- ₀ Mein Sexualeben ist normal, und die Schmerzen werden dadurch nicht stärker
₁ Mein Sexualeben ist normal, aber die Schmerzen werden dadurch stärker
₂ Mein Sexualeben ist nahezu normal, aber sehr schmerzhaft
₃ Mein Sexualeben ist durch Schmerzen stark eingeschränkt
₄ Ich habe auf Grund von Schmerzen fast kein Sexualeben
₅ Schmerzen verhindern jegliches Sexualeben

Abschnitt 9: Sozialleben

- ₀ Mein Sozialleben ist normal, und die Schmerzen werden dadurch nicht stärker
₁ Mein Sozialleben ist normal, aber die Schmerzen werden dadurch stärker
₂ Schmerzen haben keinen wesentlichen Einfluss auf mein Sozialleben, ausser dass sie meine eher aktiven Interessen, z.B. Sport einschränken
₃ Schmerzen schränken mein Sozialleben ein, und ich gehe nicht mehr so oft aus
₄ Schmerzen schränken mein Sozialleben auf mein Zuhause ein
₅ Ich habe auf Grund von Schmerzen kein Sozialleben

Abschnitt 10: Reisen

- ₀ Ich kann überallhin reisen, und die Schmerzen werden dadurch nicht stärker
₁ Ich kann überallhin reisen, aber die Schmerzen werden dadurch stärker
₂ Trotz starker Schmerzen kann ich länger als 2 Stunden unterwegs sein
₃ Ich kann auf Grund von Schmerzen höchstens 1 Stunde unterwegs sein
₄ Ich kann auf Grund von Schmerzen nur kurze notwendige Fahrten unter 30 Minuten machen
₅ Schmerzen hindern mich daran, Fahrten zu machen, ausser zur medizinischen Behandlung

References

- Baker D, Pynsent P, Fairbank J (1989) The Oswestry disability index revisited. In: Back pain: new approaches to rehabilitation and education. Manchester University Press, Manchester, pp 174–186
- Basler HD, Jakle C, Kroner-Herwig B (1997) Incorporation of cognitive-behavioral treatment into the medical care of chronic low back patients: a controlled randomized study in German pain treatment centers. *Patient Educ Couns* 31:113–124
- Beaton DE (2000) Understanding the relevance of measured change through studies of responsiveness. *Spine* 25:3192–3199
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25:3186–3191
- Bergner M, Rothman ML (1987) Health status measures: an overview and guide for selection. *Ann Rev Public Health* 8:191–210
- Beurskens AJ, de Vet HC, Kōke AJ, van der Heijden GJ, Knipschild PG (1995) Measuring the functional status of patients with low back pain. Assessment of the quality of four disease-specific questionnaires. *Spine* 20:1017–1028
- Bland JM, Altman DG (1997) Cronbach's alpha. *BMJ* 314:572
- Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders. Summary and general recommendations. *Spine* 25:3100–3103

9. Bombardier C (2000) Spine focus issue introduction. Outcome assessments in the evaluation of treatment of spinal disorders. *Spine* 25:3097–3099
10. Boscainos PJ, Sapkas G, Stilianessi E, Prouskas K, Papadakis SA (2003) Greek versions of the Oswestry and Roland–Morris Disability Questionnaires. *Clin Orthop* 411:40–53
11. Davidson M, Keating JL (2002) A comparison of five low back disability questionnaires: reliability and responsiveness. *Phys Ther* 82:8–24
12. Dederling A, Elfving B, Nemeth G (2004) EMG recovery and ratings after back extensor fatigue in patients with lumbar disc herniation and healthy subjects. *Eur J Appl Physiol* 92:150–159
13. Deyo RA (1984) Pitfalls in measuring the health status of Mexican Americans: comparative validity of the English and Spanish sickness impact profile. *Am J Public Health* 74:569–573
14. Deyo RA, Battie M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* 23:2003–2013
15. Edibe Y, Tulin D, Oksuz C, Yorukan S, Ureten K, Turan D, Frant T, Kiraz S, Krd M, Kayhan H, Yakut Y, Guler C (2004) Validation of the Turkish version of the Oswestry Disability Index for patients with low back pain. *Spine* 29:581–585
16. Exner V, Keel P (2000) Erfassung der Behinderung bei Patienten mit chronischen Rückenschmerzen. *Schmerz* 14:392–400
17. Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain questionnaire. *Physiotherapy* 66:271–273
18. Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. *Spine* 25:2940–2952
19. Fayers PM, Machin D (2000) Quality of life: assessment, analysis and interpretation. Wiley, Chichester
20. Fisher K, Johnson M (1997) Validation of the Oswestry low back pain disability questionnaire, its sensitivity as a measure of change following treatment and its relationship with other aspects of the chronic pain experience. *Physiother Theory Pract* 13:67–80
21. Fritz JM, Irrgang JJ (2001) A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther* 81:776–788
22. Gronblad M, Hupli M, Wennerstrand P, Jarvinen E, Lukinmaa A, Kouri JP, Karaharju EO (1993) Intercorrelation and test–retest reliability of the Pain Disability Index (PDI) and the Oswestry disability questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 9:189–195
23. Grotle M, Brox JI, Vollestad NK (2003) Cross-cultural adaptation of the Norwegian versions of the Roland–Morris Disability Questionnaire and the Oswestry Disability Index. *J Rehabil Med* 35:241–247
24. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 46:1417–1432
25. Hagg O, Fritzell P, Nordwall A (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 12:12–20
26. Holm I, Friis A, Storheim K, Brox JI (2003) Measuring self-reported functional status and pain in patients with chronic low back pain by postal questionnaires. *Spine* 28:828–833
27. Hopkins WG (2000) Measures of reliability in sports medicine and science. *Sports Med* 30:1–15
28. Hsieh C-YJ, Phillips RB, Adams AH, Pope MH (1992) Functional outcomes of low back pain. Comparison of four treatment groups in a randomised controlled trial. *J Manipulative Physiol Ther* 15:4–9
29. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI (1995) The Quebec Back Pain Disability Scale. Measurement properties *Spine* 20:341–352
30. Leclaire R, Blier F, Fortin L, Proulx R (1997) A cross-sectional study comparing the Oswestry and Roland–Morris functional disability scales in two populations of patients with low back pain of different levels of severity. *Spine* 22:68–71
31. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank J (2005) Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. *Eur Spine J*, DOI 10.1007/s00586-004-0816-z
32. Muller U, Duetz MS, Roeder C, Greenough CG (2004) Condition-specific outcome measures for low back pain. Part 1: validation. *Eur Spine J* 13:301–313
33. Perneger TV (1998) What's wrong with Bonferroni adjustments. *BMJ* 316:1236–1238
34. Roland M, Fairbank J (2000) The Roland–Morris disability questionnaire and the Oswestry disability questionnaire. *Spine* 25:3115–3124
35. Roland M, Morris R (1983) A study of the natural history of back pain. Part 1: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8:141–144
36. Streiner DL, Norman GR (1995) Health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford
37. Yang Y, Eaton S, Maxwell MW (1983) The relationship between the St Thomas and Oswestry Disability Scores and the severity of low back pain. *J Manipulative Physiol Ther* 16:14–18