



# Cross-cultural adaptation and psychometric validation of the Indonesian version of the Oswestry Disability Index

Phedy Phedy<sup>1</sup> · Yoshi Pratama Djaja<sup>1</sup> · Singkat Dohar Apul Lumban Tobing<sup>2</sup> · Luthfi Gatam<sup>1</sup> · Didik Librianto<sup>1</sup> · Fachrisal<sup>1</sup> · Asrafi Rizki Gatam<sup>1</sup> · Nicko Perdana Hardiansyah<sup>3</sup>

Received: 2 July 2020 / Revised: 12 November 2020 / Accepted: 2 December 2020 / Published online: 2 January 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

## Abstract

**Purpose** The Oswestry Disability Index (ODI) is one of the most common patient-reported outcome measures used for low back pain (LBP). Since it was not available in Indonesian, this study aimed to perform a cross-cultural adaptation of ODI to Indonesian and evaluate its psychometric properties.

**Methods** We performed a cross-cultural adaptation of ODI v2.1a into Indonesian language (ODI-ID) and determined its internal consistency, test–retest reliability, measurement error, confirmatory factor analysis, floor-ceiling effect, and construct validity by hypotheses testing of its correlation with Short Form-36 (SF-36) and visual analog scale (VAS). Adults (> 17 years of age) diagnosed with low back pain were included.

**Results** A total of 96 subjects were included in this study. The original ODI questionnaire was translated into an Indonesian version and showed a good internal consistency (Cronbach's alpha = 0.90) and good reliability (intraclass correlation = 0.97). The standard error of measurement values of 3.35 resulted in a minimal detectable change score of 9. Nine out of nine (100%) a priori hypotheses were met, confirming the construct validity. A strong correlation was found with the physical component of SF-36 (0.77 and 0.76 for pain and physical function, respectively) and VAS (0.79). Confirmatory factor analysis resulted in a poor but significant fit to the original one-factor structure and the static-dynamic two-factor structure. Floor-ceiling effects were not found.

**Conclusions** The Indonesian version of ODI displayed similar reliability, validity, and psychometric characteristics to the original ODI. This questionnaire will be a suitable instrument for assessing LBP-related disability for Indonesian-speaking patients.

**Keywords** Low back pain · Oswestry Disability Index · Cross-cultural adaptation · Indonesian

---

**Supplementary Information** The online version contains supplementary material available at (<https://doi.org/10.1007/s00586-020-06690-3>).

✉ Yoshi Pratama Djaja  
yoshipratamadjaja@yahoo.com

<sup>1</sup> Department of Orthopaedic and Traumatology, Fatmawati Central General Hospital, Jl. Fatmawati No. 1, Jakarta Selatan, DKI Jakarta, Indonesia

<sup>2</sup> Department of Orthopedic, Rumah Sakit Cipto Mangunkusumo, Jakarta, Indonesia

<sup>3</sup> Department of Orthopaedic, Soelianti Saroso General Hospital, Jakarta, Indonesia

## Introduction

Low back pain (LBP) has been recognized as an important health issue worldwide. Despite no published study documenting the exact prevalence of LBP in Indonesia, such condition is considered as the second most frequent reason for pain-related hospital visits [1]. A substantial amount of information regarding LBP prevalence was found, which resulted in the heterogeneity of the data: point prevalence range from 1.0 to 58.1% (mean 18.1%) and 1-year prevalence from 0.8 to 82.5% (mean 38.1%) [2]. In 2010, the Global Burden of Disease Study showed that the global point prevalence of LBP was 9.4% with the global burden increased from 52.8 million in 1990 to 83.0 million in 2010. They also stated that LBP ranked highest in terms of disability and sixth in terms of the overall burden. These data urge

the need for further studies regarding LBP across different settings. [3]

Specific tools have been developed to quantify the functional status, as it cannot be assessed adequately by clinical assessment alone. One of the most commonly used patient-reported outcome measures (PROM) for LBP is the Oswestry Disability Index (ODI) [4–7]. ODI, first established by Fairbank in 1980, consists of ten questions that are categorized into two sections: the intensity of the pain and its disabling effect on personal and social life. The validity, reliability, and responsiveness of the original English version of ODI have also been proven to be satisfactory. A cross-cultural adaptation of ODI is required before using it in different languages or cultures [8]. Currently, the ODI has been culturally adapted, translated, and validated in more than 20 languages. Moreover, the ODI has also been included as a PROM in the International Consortium for Health Outcome Measurement standard set for LBP [9].

The Indonesian language is one of the most commonly spoken languages in the world as it is the national language of Indonesia. Although over 300 different native languages are spoken in Indonesia, the Indonesian language is used by more than 200 million Indonesian people either as their first or second language [10]. However, since no Indonesian version of ODI is available, its use in Indonesia is not possible. The aim of this study was to perform a cross-cultural adaptation of ODI in Indonesian language and to assess its psychometric properties.

## Materials and methods

UK English version of the ODI 2.1a was used in this study. Approval for the adaptation was obtained from the MAPI Research Trust. The Institutional Review Board approved the trial prior to the study (No.47/TU/DM/IX/2016). The adaptation was performed based on MAPI guidelines and Beaton's cross-cultural guidelines [8].

### Translation and cross-cultural adaptation

#### Phase 1: Forward translation

Two native Indonesian speakers independently translated the original questionnaire to Indonesian. The first translator (naïve translator/FT1) was neither familiar with the concept of the ODI questionnaire nor with LBP. On the other hand, the second translator (FT2) had a medical background and was aware of the concepts that are being evaluated. Each forward translation was compared. The different/ambiguous terms were documented and resolved after a discussion between the two translators resulting in combined forward translations (FT12).

#### Phase 2: Back translation

Two translators separately translated back the FT12 version of ODI into English (BT1 and BT2). Both translators had no medical background and were unaware of the original version. Both back-translations (BT1 and BT2) then were compared with the original version of the questionnaire to validate whether the translated version reflects the same content as the original version.

#### Phase 3: The expert committee

The back-translation was reviewed by the MAPI project manager, four translators, and the principal investigators. This review process aimed to highlight any discrepancies in meaning or terminology used and to obtain the best possible translation, which was the pre-final version. Each issue, rationale, and decisions during the discussions were documented.

#### Phase 4: Pilot Test of the Pre-Final Version

The comprehensiveness of the pre-final questionnaire was tested in 30 LBP patients to ensure the adapted version was understandable. After completing the questionnaire, the subjects were interviewed to explore their understanding of each question and response. This result of this test was then re-evaluated by the committee, and the final form of the questionnaire (ODI-ID) was then established (Supplemental Data File 1).

#### Phase 5: Test of the Final Version

The questionnaire (ODI-ID) was field-tested to assure the validity, and other psychometric properties remained intact. Consecutive sampling was conducted in an outpatient spine clinic in a tertiary referral general hospital from November 2016 to February 2017. The sample size (115 patients) was predetermined based on the subject to item ratio  $\geq 10$  [11]. The inclusion criteria were: LBP  $\geq 6$  weeks; adult (older than 17 years); able to read and write in Indonesian fluently. The exclusion criteria were acute LBP; the presence of neurological deficit; had an incidental event (including surgery) during two periods observation that might increase/reduce back pain significantly. Each respondent filled the booklet (consisting of the ODI-ID, Short Form-36 questionnaire and visual analog scale) twice with a one-week interval. Both baseline and follow-up assessments were performed in the clinic. Afterward, a short-interview was performed to detect whether there were any changes in their clinical condition during this period. Additionally, we measured the changes in VAS and SF-36 (bodily pain and physical function subscale). VAS  $\leq 1$ , bodily pain (BP)  $\leq 3$ , and physical function

(PF)  $\leq 3$  were regarded as clinically stable [12]. If there was a missing answer, the total score of ODI-ID was adjusted and calibrated on a scale 0 to 100, in accordance with the original version of ODI [5].

The assessment of psychometric properties (internal consistency, test–retest reliability, measurement errors, and construct validity) was performed and presented based on COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) guidelines [13].

*Statistical Analysis* was performed using the IBM SPSS software version 22.0 (SPSS Inc., Chicago, Illinois).

## Factor analysis

*Confirmatory factor analysis* (CFA) was performed using AMOS software version 26.0 (SPSS Inc., Chicago, Illinois). Two models were evaluated from the baseline data: (1) one-dimensional and (2) two-dimensional: static activities (pain, sleep, standing, traveling, sitting) and dynamic activities (personal care, lifting, walking, sex, social), as proposed in previous studies [14–16]. Indicators determining the model fit such as: standardized root-mean-square (SRMR), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), comparative fit index (CFI), and root-mean-square error of approximation (RMSEA) were calculated. We used the following cut-off values to indicate a good fit: SRMR value  $< 0.08$  [17]; GFI and AGFI  $> 0.9$  [18]; CFI  $> 0.9$  [19]; and RMSEA  $< 0.05$  [20].

## Reliability

*Internal consistency* of ODI-ID was evaluated by calculating Cronbach's alpha (CA) of the baseline questionnaires [21, 22]. Correlation between each item and the whole instrument (ODI-ID) was calculated using Pearson correlation. A significant item-total correlation was determined if the correlation coefficient exceeded 0.576 (95% critical value of the sample correlation coefficient for 10 items) [23].

*Test–retest reliability* was determined using intraclass correlation (ICC) between the first and second tests. The ICC used was a single measurement, absolute agreement, 2-way mixed-effects model with 95% confidence intervals. An ICC of  $> 0.70$  was regarded as good reliability [24].

## Measurement error

*Standard error of measurement (SEM)* was determined from the error variance of the ANOVA associated with determination of the ICC [25].

*Minimum detectable change (MDC<sub>95%</sub>)* was calculated by multiplying SEM by 2.77 where 2.77 was obtained from Z value for the 95% CI (1.96) times the variance of two measurements ( $\sqrt{2}$ ) [25, 26].

*The distribution of floor-and-ceiling effect* (percentage of sample achieving the worst and best possible scores, respectively) was also determined for both the baseline test and the follow-up test. Test instruments should exhibit minimal floor and ceiling effect (less than 15% of the respondents) to be considered reliable [27].

## Construct validity

### Hypothesis testing

A total of nine hypotheses (Table 1) were tested to evaluate the construct validity of ODI-ID, using the standard hypothesis testing methodology [28]. Pearson correlation coefficient was used to assess the association of baseline ODI-ID with all the subscales of the SF-36 questionnaire and visual analog scale for pain. Pearson correlation ( $r$ ) of  $> 0.50$ ,  $0.36$ – $0.49$ , and  $< 0.35$  was considered strong, moderate, and weak correlations, respectively. The results are reported as hypothesis confirmed or not confirmed. The total of met hypotheses was reported as percentages. If it was more than 75%, we confirmed the construct validity of the ODI-ID [28].

**Table 1** Hypotheses testing for construct validity of ODI-ID

| Scale                                    | Hypothesis                              |
|--|---|
| ODI-ID score versus VAS                  | Positive strong association             |
| <i>ODI-ID score versus SF-36 domains</i> |   |
| SF-36 physical functioning (PF)          | Negative strong association             |
| SF-36 role limitation physical (RP)      | Negative moderate to strong association |
| SF-36 role limitation emotional (RE)     | Negative moderate to strong association |
| SF-36 mental health (MH)                 | Negative moderate to strong association |
| SF-36 bodily pain (BP)                   | Negative strong association             |
| SF-36 general health perceptions (GH)    | Negative moderate to strong association |
| SF-36 vitality (energy/fatigue) (VT)     | Negative moderate to strong association |
| SF-36 social functioning (SF)            | Negative moderate to strong association |

## Results

### Translation and cross-cultural adaptation

During the cross-cultural adaptation process, several noteworthy issues arose during the translation phase and were solved in an expert meeting.

- (1) **Translation of “personal care”.** The phrase can be translated into two similar words: “*asuhan pribadi*” or “*perawatan pribadi*”. However, none of them is clear enough to describe personal care activities in Indonesian language. During the pretesting phase, some respondents were confused with this phrase. Thus, we decided to add a description to explain which activities are included such as “wash, dressed, etc.”
- (2) **Translation of walking distance.** There was a difference in terms of metric used in distance between the original version and Indonesian culture. The metric units such as miles and yards are rarely used in Indonesia. The original English version used 1 mile, quarter-mile, and 100 yards. The direct conversions of these metrics are 1.6 km, 400 m, and 91 m, respectively. After discussion, we decided to round these units to “1.5 km”, “500 m”, and “100 m”. We expected that the rounding would not affect the validity or reliability.
- (3) **Translation of “prevents”.** The direct translation of the word “prevents” in Indonesian is “*mencegah*” which means avoiding something negative to occur. Thus, after discussion, we felt another word (“*menghalangi*”), whose direct translation was “hinder/limit”, is more suitable in our language to describe the disabilities that occurred.

### Patient demographics and scores distribution

The study initially included 115 respondents; 4 were dropped out due to inability to return for the second test; meanwhile, 15 were considered unstable as they underwent changes in their treatment course and/or clinical condition. This left 96 final respondents to be further analyzed. Their average age

was  $40.1 \pm 12.6$  years, and 48 respondents (50%) were male. The average time to complete ODI-ID was  $4.4 \pm 1.5$  min (2.5–6.5 min), and the average interval between the first and second questionnaires was  $6.0 \pm 1.3$  days.

The average ODI-ID score on the first and second administrations was  $44.7 \pm 13.8$  and  $44.5 \pm 12.6$ , respectively. Twenty-eight respondents did not fill out an answer to the question about sexual life. There were no floor or ceiling effects observed for the total ODI-ID. The average score of VAS was  $3.8 \pm 1.5$  (moderate pain). The average SF-36 score was  $50.4 \pm 23$ .

### Confirmatory factor analysis

The baseline ODI-ID data had a significant yet poor fit with both 1-factor and 2-factor (static-dynamic) models as shown in Table 2. The Chi-square difference between these two models was not significant ( $p > 0.05$ ); therefore, we used the 1-factor model to calculate the Cronbach’s alpha.

### Internal consistency

CA index for ODI-ID was 0.90, suggesting good internal consistency. A significant correlation with total ODI-ID was found in every item, except for social life and travelling. Correlation between social life and traveling with the whole scale was 0.55 and 0.52, respectively. The CA for the rest of the scale did not exceed 0.90 if any of the items were removed.

### Test-retest reliability and measurement error

The intraclass correlation coefficient for ODI-ID was 0.97 (95% CI 0.96–0.99) showing good reliability. Standard error of measurement was 3.35. The minimum detectable change was 9.

### Construct validity

ODI-ID has statistically significant correlations with VAS and all SF-36 subscales (Table 3). We found a total of 9 out of 9 a priori hypotheses were met, confirming the construct

**Table 2** Confirmatory factor analysis of the ODI-ID

| Models                 | $\chi^2$ | d.f. | Sig.   | SRMR | GFI  | AGFI | CFI  | RMSEA | Sig. of $\Delta\chi^2$ |
|------------------------|----------|------|--------|------|------|------|------|-------|------------------------|
| One-dimensional        | 159.68   | 35   | <0.001 | .202 | .701 | .532 | .692 | .222  |                        |
| Two-dimensional        | 153.28   | 34   | <0.001 | .252 | .721 | .561 | .705 | .214  | $P > 0.05$             |
| Static: 1, 5, 6, 7, 10 |          |      |        |      |      |      |      |       |                        |
| Dynamic: 2, 3, 4, 8, 9 |          |      |        |      |      |      |      |       |                        |

$\chi^2$ : chi square statistics, *d.f.* degree of freedom, *Sig.* significance, *SRMR* standardized root mean square residual, *GFI* goodness of fit index, *AGFI* adjusted goodness of fit index, *CFI* comparative fit index, *RMSEA* root-mean-square error of approximation

**Table 3** Construct validity: correlation coefficient between ODI-ID, VAS and SF-36

| Comparison                            | <i>r</i> | <i>p</i> value | Hypothesis confirmed |
|---------------------------------------|----------|----------------|----------------------|
| ODI score vs. VAS                     | 0.849    | 0.002          | Yes                  |
| <i>ODI score vs. SF-36 domains</i>    |          |                |                      |
| SF-36 physical functioning (PF)       | – 0.731  | <0.001         | Yes                  |
| SF-36 role limitation physical (RP)   | – 0.405  | <0.001         | Yes                  |
| SF-36 role limitation emotional (RE)  | – 0.432  | <0.001         | Yes                  |
| SF-36 mental health (MH)              | – 0.369  | 0.002          | Yes                  |
| SF-36 bodily pain (BP)                | – 0.755  | <0.001         | Yes                  |
| SF-36 general health perceptions (GH) | – 0.463  | <0.05          | Yes                  |
| SF-36 vitality (energy/fatigue) (VT)  | – 0.562  | <0.001         | Yes                  |
| SF-36 social functioning (SF)         | – 0.644  | <0.001         | Yes                  |

*ODI* Oswestry Disability Index, *VAS* visual analog scale for assessment of chronic back pain, *SF-36* Short Form 36-item survey, *ρ*: correlation coefficient

validity of ODI-ID. It correlated strongly with VAS, SF-36 physical functioning (PF), and bodily pain (BP), moderately with the other SF-36 subscales.

## Discussion

The aim of this study was to adapt the ODI questionnaire to Indonesian-speaking patients and evaluate the psychometric properties of the Indonesian version. The translation process was carried out following the established guideline for cross-cultural adaptations [8] to obtain a reliable and valid adaptation of the questionnaire.

The ODI-ID was easily administered and understood by the patients. This was shown by the average time to complete the questionnaire of 4.5 min, which is comparable to other ODI versions (range: 3.4–6.6 min) [29–32]. Twenty-eight patients failed to answer to “sex life” question. This finding was similar to other studies [30, 31, 33], which most likely attributed to cultural issues. Additionally, since ODI version 2.1a was used as a reference, the question number 8 comes with an additional statement: “if applicable” which rendered the patients for not answering this question. There was also some concern that the sex life question was not answered accurately. Some believed that patients did not answer this question due to psychosocial factor (do not have a partner, other condition that does not allow them to have sex, etc.) rather than a pain-related factor [34]. However, we decided to keep this question in the ODI-ID since the content of the question was considered to be important. Furthermore, as stated in the original version, the scoring method will be adjusted if there is a missing/inapplicable question [4].

The result of confirmatory factor analysis showed a significant fit for both tested models; however, the fit was not brilliant in our series. There was also no statistically significant difference in the chi-square between models in

our study, indicating that the 2-factor model did not provide a better fit than the original ODI 1-factor model. Previous studies indicated uncertainty about the factor analysis for ODI. While Italian [31], Slovenian [35], Polish [36], Hungarian [37] and Dutch [38] version of ODI revealed 1-factor structure, 2-factor structures were shown by German [33], Spanish-Colombia [39], Finnish [40], and Arabic [29] version of ODI. The number of samples might play a role in their result, which was proven by Gabel et al. They performed a factor analysis in 35,623 LBP patients and verified the one-factor structure that was proposed by the original author [14].

The Indonesian version of ODI demonstrated good internal consistency (CA = 0.9) that even exceeded the original English version, which ranged from 0.71 to 0.87 [4, 41–43]. CA values are considered high if it ranged from 0.70 to 0.90. If CA is too high (> 0.90), it may suggest that some items are redundant as they are testing the same question [22]. Similar internal consistencies were found with the most translated ODI versions, especially Danish (CA = 0.88) [44], Turkish (CA = 0.9) [45], German-Swiss (CA = 0.9) [46], German (CA = 0.89) [33], Slovenian (CA = 0.9) [35], French Canada (CA = 0.88) [47], Polish (CA = 0.9) [36], Hungarian (CA = 0.89) [37], India-Marathi (CA = 0.88) [48] and Arabic (CA = 0.89) [29]. The other versions were ranged from 0.75 to 0.99 (Table 4). Furthermore, the alpha for each item in ODI-ID did not exceed the alpha for the total item, which indicates the homogeneity of the questionnaire.

Test–retest reliability shows the stability of a questionnaire within a certain timeline, with ICC values illustrating the inter-rater reliability within the two time intervals. A value of 0.97 showed that this version has excellent test–retest reliability. One of the limitations of this study is that we did not include a transition question to confirm whether the patients reported any change in their condition during this period. However, by conducting an interview and

**Table 4** Characteristics of the published cross-cultural adaptation studies of the Oswestry Disability Index (updated from Domazet et al. [30])

| First author, year              | Adapted language     | n <sup>c</sup> | %F <sup>d</sup> | Ag <sup>e</sup> | q8 (%n) <sup>f</sup>  | Factor analysis <sup>g</sup> |      |
|---------------------------------|----------------------|----------------|-----------------|-----------------|-----------------------|------------------------------|------|
|                                 |                      |                |                 |                 |                       | n                            | R2   |
| Boscainos (2003) [54]           | Greek                | 697            | 66.3            | 43.2±14.2       | –                     | –                            | –    |
| Fujiwara (2003) [55]            | Japanese             | 53             | 44.3            | 51 (16–85)      | 40                    | –                            | –    |
| Grotle (2003) [50]              | Norwegian            | 105            | 67.6            | 38.9±9.5        | –                     | –                            | –    |
| Yakut (2004) [45]               | Turkish              | 95             | 75              | 65.3±11.6       | –                     | –                            | –    |
| Kim (2005) [56]                 | Korean               | 206            | 36              | 46±14.8         | Omitted               | –                            | –    |
| Lauridsen (2006) [44]           | Danish               | 233            | 50.6            | 43.8±15.4       | 11                    | –                            | –    |
| Mannion (2006) [46]             | German (Swiss)       | 100            | 53              | 53.1±14.6       | 19                    | –                            | –    |
| Osthus (2006) [33]              | German               | 160            | 27              | 47.5±9.0        | 12.5                  | 2                            | 30.2 |
| Mousavi (2006) [32]             | Persian (Iran)       | 100            | 55              | 40.1±11.6       | –                     | –                            | –    |
| Hashimoto (2006) [57]           | Japanese             | 167            | 50              | 57±15           | 27                    | –                            | –    |
| Sanjaroensuttikul, (–2007) [58] | Thai                 | 40             | –               | 40.1±10.7       | –                     | –                            | –    |
| Vigatto (2007) [52]             | Brasilian (Portugal) | 120            | 66.7            | 37.9            | –                     | –                            | –    |
| Ramzy (2008) [59]               | Arabic (UAE)         | 108            | 25.9            | 40.2±11         | 25                    | –                            | –    |
| Liu (2009) [51]                 | Simple Chinese       | 179            | 58.1            | 44.5±14.7       | 30.7                  | –                            | –    |
| Monticone (2009) [31]           | Italian              | 126            | 58              | 47±14           | 11.9                  | 1                            | 45   |
| Klemenc-Ketiš (2011) [35]       | Slovenian            | 129            | 52.7            | 50.1±10.2       | –                     | 1                            | 54.1 |
| Payares, 2011 [39]              | Spanish (Colombia)   | 111            | 67.6            | 44.9±16.4       | 23                    | 2                            | 55.6 |
| Pekkanen (2011) [40]            | Finnish              | 115            | 44              | 49±13           | 14.8                  | 2                            | 51   |
| Denis (2012) [47]               | French (Canada)      | 72             | 51.4            | –               | 17; 19 <sup>a</sup>   | –                            | –    |
| Miekisiak (2013) [36]           | Polish               | 164            | 57.3            | 49.7±11.8       | 22                    | 1                            | 54.6 |
| Valasek (2013) [37]             | Hungarian            | 133            | 58              | 48.1±15.3       | 33.1                  | 1                            | 51.3 |
| Joshi (2013) [48]               | Marathi (India)      | 100            | 56              | 41.7±15.5       | –                     | –                            | –    |
| Algarni (2014) [29]             | Arabic (SA)          | 100            | 45              | 40.4±12.7       | –                     | 2                            | 63.5 |
| van Hooff (2015) [38]           | Dutch                | 244            | 57.4            | 45.6±10.8       | –                     | 1                            | 36   |
| Yu (2016) [60]                  | Russian              | 97             | –               | 74 (70–80)      | 82.5; 90 <sup>a</sup> | –                            | –    |
| Lue (2008) [61]                 | Chinese              | 177            | 55.9            | 42.9±15.5       | 29.4                  | –                            | –    |
| Domazet (2018) [30]             | Croatian             | 114            | 52.6            | 52.0±15.2       | 28.9                  | 2                            | 82.7 |
| Current Study                   | Indonesian           | 96             | 50              | 40.1±12.6       | 29.1                  | –                            | –    |

| First author, year    | Adapted language | Test–retest reliability <sup>h</sup> |      | Int. cons | Measurement error |     |        | Criterion validity (r) <sup>j</sup> |                    |
|-----------------------|------------------|--------------------------------------|------|-----------|-------------------|-----|--------|-------------------------------------|--------------------|
|                       |                  | Days                                 | ICC  |           | α                 | SEM | MDC    | SF-36                               | VAS <sup>k</sup>   |
| Boscainos (2003) [54] | Greek            | –                                    | –    | 0.83      | –                 | –   | –      | –                                   | RMDQ, SPPS         |
| Fujiwara (2003) [55]  | Japanese         | 1                                    | 0.93 | 0.83      | –                 | –   | All    | –                                   | –                  |
| Grotle (2003) [50]    | Norwegian        | 2                                    | 0.88 | 0.94      | –                 | 11  | All    | 0.39 (ac.), 0.52 (chr.)             | –                  |
| Yakut (2004) [45]     | Turkish          | 7                                    | 0.94 | 0.9       | –                 | –   | –      | 0.37; 0.39 <sup>a</sup>             | Schober test, RMDQ |
| Kim (2005) [56]       | Korean           | 2                                    | 0.92 | 0.84      | –                 | –   | –      | 0.63                                | WHOQoL-BREF        |
| Lauridsen (2006) [44] | Danish           | 1, 7                                 | 0.91 | 0.88      | –                 | 13  | PF, BP | 0.61 (NRS)                          | LBP-RS, RMDQ       |
| Mannion (2006) [46]   | German (Swiss)   | 2–14                                 | 0.96 | 0.9       | 3.4               | 9   | –      | 0.78                                | RMDQ               |
| Osthus (2006) [33]    | German           | 1, 21                                | 0.91 | 0.89      | –                 | –   | All    | –                                   | HFAQ               |
| Mousavi (2006) [32]   | Persian (Iran)   | 1                                    | 0.91 | 0.75      | –                 | –   | All    | 0.54                                | –                  |
| Hashimoto (2006) [57] | Japanese         | –                                    | –    | 0.94      | –                 | –   | PF, BP | –                                   | –                  |

**Table 4** (continued)

| First author, year            | Adapted language     | Test–retest reliability <sup>h</sup> |      | Int. cons<br>$\alpha$ | Measurement error |      |            | Criterion validity ( $r$ ) <sup>j</sup> |                    |
|-------------------------------|----------------------|--------------------------------------|------|-----------------------|-------------------|------|------------|---|--------------------|
|                               |                      | Days                                 | ICC  |                       | SEM               | MDC  | SF-36      | VAS <sup>k</sup>                        | Other <sup>l</sup> |
| Sanjaroensuttikul (2007) [58] | Thai                 | 20–30 min                            | 0.98 | –                     | –                 | –    | –          | –                                       | –                  |
| Vigatto (2007) [52]           | Brasilian (Portugal) | 1                                    | 0.99 | 0.87                  | –                 | –    | All        | 0.66 (NRS)                              | RMDQ               |
| Ramzy (2008) [59]             | Arabic (UAE)         | 2, 28                                | 0.99 | 0.99                  | –                 | –    | –          | 0.49                                    | Squat test         |
| Liu (2009) [51]               | Simple Chinese       | 1                                    | 0.99 | 0.93                  | –                 | –    | All        | 0.69                                    | –                  |
| Monticone (2009) [31]         | Italian              | 7                                    | 0.96 | 0.86                  | –                 | –    | Not MH     | 0.73                                    | RMDQ               |
| Klemenc-Ketiš (2011) [35]     | Slovenian            | 10                                   | –    | 0.9                   | –                 | –    | –          | 0.49                                    | EuroQol            |
| Payares (2011) [39]           | Spanish (Colombia)   | 7                                    | 0.94 | 0.86                  | –                 | –    | All        | > 0.48                                  | RMDQ               |
| Pekkanen (2011) [40]          | Finnish              | 1                                    | 0.9  | 0.86                  | –                 | –    | –          | –                                       | Million-VAS        |
| Denis (2012) [47]             | French (Canada)      | 2                                    | 0.92 | 0.88                  | 4.38              | –    | –          | –                                       | QBPDS, RMDQ        |
| Miekisiak (2013) [36]         | Polish               | 2–14                                 | 0.97 | 0.9                   | 3.54              | 10   | –          | 0.37 (back); 0.56 (leg)                 | RMDQ               |
| Valasek (2013) [37]           | Hungarian            | 14                                   | 0.93 | 0.89                  | 4.8               | 13   | –          | 0.68                                    | WHOQoL-BREF        |
| Joshi (2013) [48]             | Marathi (India)      | 2–14                                 | 0.88 | 0.88                  | –                 | –    | –          | 0.67                                    | RMDQ               |
| Algarni (2014) [29]           | Arabic (SA)          | 2                                    | 0.99 | 0.89                  | –                 | –    | –          | 0.71                                    | QBPDS, RMDQ        |
| van Hooff (2015) [38]         | Dutch                | –                                    | –    | 0.79                  | –                 | –    | PC, MC     | 0.40                                    | RMDQ, HADS-D/A     |
| Yu (2016) [60]                | Russian              | 7–15                                 | 0.7  | 0.82                  | –                 | –    | –          | 0.46 (back); 0.56 (leg)                 | –                  |
| Lue (2008) [61]               | Chinese              | 7                                    | 0.89 | 0.90                  | 4.6               | 12.8 | Not GH, VT | 0.68                                    | RMDQ               |
| Domazet (2018) [30]           | Croatian             | 7                                    | 0.94 | 0.84                  | 2.15              | 6    | Not PRL    | 0.54                                    | –                  |
| Current Study                 | Indonesian           | 7                                    | 0.97 | 0.90                  | 3.35              | 11   | All        | 0.79                                    | –                  |

PF physical function SF-36 subscale, BP bodily pain SF-36 subscale, PRL physical role limitation SF-36 subscale, MH mental health SF-36 subscale, PC physical component, MC mental component,  $\alpha$ : Cronbach’s alpha, NRS numeric rating scale, RMDQ Roland–Morris Disability Questionnaire, HFAQ Hannover Functional Ability Questionnaire, QBPDS Quebec Back Pain Disability Scale, WHOQoL BREF short form of the World Health Organization Quality of Life Scale, LBP-RS Low Back Pain Rating Scales (pain and disability); EuroQol European Quality of Life questionnaire; SPPS six-point pain rating scale, HADS-A: Hospital Anxiety and Depression Scale for Anxiety, HADS-D Hospital Anxiety and Depression Scale for Depression, UAE United Arab Emirates; SA Saudi Arabia,

<sup>a</sup>Test and retest values

<sup>b</sup>Section 8 was omitted from the Korean ODI

<sup>c</sup>Construct validity test sample size

<sup>d</sup>Proportion of female patients

<sup>e</sup>Numbers are mean  $\pm$  SD, or median (range)

<sup>f</sup>Proportion of patients who did not fulfill Sect. 8 (sex life)

<sup>g</sup>First column is the number of identified factors, and the second explained variance

<sup>h</sup>First column is the test–retest interval, and the second intraclass correlation coefficient (ICC)

<sup>i</sup>Internal consistency

<sup>j</sup>First column is the standard error of measurement (SEM), and second minimum detectable change with Z value for 95% confidence interval (MDC<sub>95%</sub>)

<sup>k</sup>Criterion validity—“all” means that all Short Form 36-item survey (SF-36) subscales were significantly and negatively correlated with the ODI

<sup>l</sup>First column represents the correlation coefficients for pain VAS (visual analog scale). All P values are < 0.001

<sup>m</sup> “other” represents measures other than VAS used to assess construct validity

evaluating their VAS and SF-36 (bodily pain and physical function subscale) on the second visit, we confirmed that there were no changes regarding their condition. The 7-day interval was also chosen since the selected time intervals should be short to avoid changes due to nature, but not too short to allow recalling of previous answers. Interval of 1 to 2 weeks is reported as the best time to measure the reproducibility of the functional status questionnaire [49]. In the original ODI study, the ICC value reached 0.99; however, recall bias might exist since an interval of 1 day between measurements was used [4].

As for measurement error, only a few ODI-validation-studies published the standard error of measurement (SEM) and minimum detectable change (MDC), which are an important feature of a questionnaire. For the Indonesian version, the MDC<sub>95%</sub> was 9 which is similar to the German (Swiss) version [46]. The range of SEM in the previously validated version was 3.4–4.8 which resulted in MDC<sub>95%</sub> around 9–13 (Table 4).

To assess construct validity, we performed hypotheses testing as recommended by the COSMIN guideline [28]. All a priori hypotheses were met, which support the construct validity of the ODI-ID. We selected the SF-36 questionnaire and visual analog scale as our instruments for the hypotheses, due to their availability in Indonesian language. Besides the Roland Morris Disability Questionnaire (RMDQ), which is unavailable in Indonesian, SF-36 and VAS were commonly used to assess the validity of other cross-cultural adaptation of ODI [31, 32, 39, 44, 50–52].

The hypotheses were developed based on the direction and magnitude of the correlations obtained from previous cross-culturally adapted versions. Due to the wide margin of correlations in the previous studies, the magnitude of the correlations was made wider in our hypotheses. Although the hypotheses on the association of ODI-ID with all subscales of SF-36 and VAS were met, the strength of correlations was varied. The correlation of ODI-ID with BP subscales ( $r = -0.76$ ) were higher than the Italian ( $r = -0.69$ ) [31], Iranian ( $r = -0.68$ ) [32], Norwegian ( $r = -0.64$ ) [50] and Brazilian-Portuguese version ( $r = -0.58$ ) [52]. The correlation of ODI-ID with PF subscales ( $r = -0.72$ ) was similar to the Italian ( $r = -0.75$ ), Norwegian ( $r = -0.77$ ), and Iranian version ( $r = -0.68$ ) but lower than the Brazilian Portuguese version ( $r = -0.83$ ). Similar discrepancies were found for the correlation with other SF-36 subscales. Meanwhile, ODI-ID showed stronger VAS correlation ( $r = 0.85$ ) when compared with the other adapted versions ( $r = 0.37-0.78$ ; as shown in Table 4).

Our study has several limitations. First, regarding the transition question that was brought up earlier. We also did not measure the responsiveness of ODI-ID, which is necessary to evaluate its ability of the questionnaire to detect small important clinical changes. And lastly, exploratory

factor analysis (EFA) was not performed since the number of samples was inadequate. The required number to perform an EFA is around 500–1000 respondents to obtain a good cumulative explained variance [53].

## Conclusion

Translation and cultural adaptation of ODI in Indonesian was successful. The Indonesian version of ODI maintained the reliability, validity, and psychometric characteristics of the original ODI. This questionnaire will be a suitable instrument for assessing LBP-related disability for Indonesian-speaking patients.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest related with this study.

## References

- Darmawan J, Valkenburg HA, Muirden KD, Wigley RD (1992) Epidemiology of rheumatic diseases in rural and urban populations in Indonesia: a World Health Organisation International League Against Rheumatism COPCORD study, stage I, phase 2. *Ann Rheum Dis* 51(4):525–528
- Hoy D, Brooks P, Blyth F, Buchbinder R (2010) The Epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 24(6):769–781
- Hoy D, March L, Brooks P et al (2014) The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 73(6):968–974
- Fairbank JC, Couper J, Davies JB, O'Brien JP (1980) The Oswestry low back pain disability questionnaire. *Physiotherapy* 66(8):271–273
- Fairbank JC, Pynsent PB (2000) The Oswestry Disability Index. *Spine (Phila Pa 1976)* 25(22):2940–2952 (**discussion** 52)
- Herdman M, Fox-Rushby J, Badia X (1998) A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res* 7(4):323–335
- Vavken P, Ganai-Antonio AK, Quidde J, Shen FH, Chapman JR, Samartzis D (2015) Fundamentals of clinical outcomes assessment for spinal disorders: clinical outcome instruments and applications. *Global Spine J* 5(4):329–338
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB (2000) Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 25(24):3186–3191
- Standard Set for Low Back Pain [Internet] (2014) [cited August 2020]. <https://www.ichom.org/portfolio/low-back-pain/>.
- Indonesia S (2013) Indonesian population projection 2010–2035. Statistics indonesia, Jakarta
- Anthoine E, Moret L, Regnault A, Sebillé V, Hardouin JB (2014) Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 12:176
- Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N (2006) Responsiveness and minimal clinically



- important difference for pain and disability instruments in low back pain patients. *BMC Musculoskeletal Disord.* <https://doi.org/10.1186/1471-2474-7-82>
13. Mokkink LB, Terwee CB, Patrick DL et al (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19:539–549
  14. Gabel CP, Cuesta-Vargas A, Qian M et al (2017) The Oswestry Disability Index, confirmatory factor analysis in a sample of 35,263 verifies a one-factor structure but practicality issues remain. *Eur Spine J* 26(8):2007–2013
  15. Guerhazi M, Mezghani M, Ghroubi S et al (2005) The Oswestry index for low back pain translated into Arabic and validated in a Arab population. *Ann Readapt Med Phys* 48(1):1–10
  16. Tan K, Zheng M, Yang BX et al (2009) Validating the Oswestry disability index in patients with low back pain in Sichuan. *Sichuan Da Xue Xue Bao Yi Xue Ban* 40(3):559–561
  17. Hooper D, Coughlan J, Mullen MR (2008) Structural equation modelling: guidelines for determining model fit. *Electron J Bus Res Methods* 6(1):53–60
  18. Mulaik SA, James LR, Van Alstine J, Bennett N, Lind S, Stilwell CD (1989) Evaluation of goodness-of-fit indices for structural equation models. *Psychol Bull* 105(3):430–445
  19. Bentler PM (1990) Comparative fit indexes in structural models. *Psychol Bull* 107(2):238–246
  20. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods* 4(3):272–299
  21. Cronbach LJ, Warrington WG (1951) Time-limit tests: estimating their reliability and degree of speeding. *Psychometrika* 16(2):167–188
  22. Tavakol M, Dennick R (2011) Making sense of Cronbach's alpha. *Int J Med Educ* 2:53–55
  23. Sokal RR, Rohlf JJ (1995) *Biometry. The principles and practice of statistics in biological research*, 3rd edn. Freeman, New York
  24. Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 26(4):217–238
  25. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC (2007) Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 7(5):541–546
  26. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8(2):135–160
  27. McHorney CA, Tarlov AR (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 4(4):293–307
  28. Terwee CB, Bot SD, de Boer MR et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
  29. Algarni AS, Ghorbel S, Jones JG, Guerhazi M (2014) Validation of an Arabic version of the Oswestry index in Saudi Arabia. *Ann Phys Rehabil Med* 57(9–10):653–663
  30. Domazet I, Nemir J, Barl P et al (2018) Validation of the Croatian version of the Oswestry Disability Index. *Eur Spine J* 27(11):2814–2822
  31. Monticone M, Baiardi P, Ferrari S et al (2009) Development of the Italian version of the Oswestry Disability Index (ODI-I): A cross-cultural adaptation, reliability, and validity study. *Spine (Phila Pa 1976)* 34(19):2090–2095
  32. Mousavi SJ, Parnianpour M, Mehdian H, Montazeri A, Mobini B (2006) The Oswestry disability index, the Roland-Morris disability questionnaire, and the Quebec back pain disability scale: translation and validation studies of the Iranian versions. *Spine (Phila Pa 1976)* 31(14):E454–E459
  33. Osthus H, Cziske R, Jacobi E (2006) Cross-cultural adaptation of a German version of the Oswestry Disability Index and evaluation of its measurement properties. *Spine (Phila Pa 1976)* 31(14):E448–E453
  34. Aithala JP (2015) Difficulties in using Oswestry Disability Index in Indian patients and validity and reliability of translator-assisted Oswestry Disability Index. *J Orthop Surg Res* 10:90
  35. Klemenc-Ketiš Z (2011) Disability in patients with chronic non-specific low back pain: validation of the Slovene version of the Oswestry Disability Index. *Slov Public Health.* <https://doi.org/10.2478/v10152-010-0031-8>
  36. Miekisiak G, Kollataj M, Dobrogowski J et al (2013) Validation and cross-cultural adaptation of the Polish version of the Oswestry Disability Index. *Spine (Phila Pa 1976)* 38(4):E237–E243
  37. Valasek T, Varga PP, Szoverfi Z, Kumin M, Fairbank J, Lazary A (2013) Reliability and validity study on the Hungarian versions of the Oswestry disability index and the Quebec back pain disability scale. *Eur Spine J* 22(5):1010–1018
  38. van Hooft ML, Spruit M, Fairbank JC, van Limbeek J, Jacobs WC (2015) The Oswestry disability index (version 2.1a): validation of a Dutch language version. *Spine (Phila Pa 1976)* 40(2):E83–90
  39. Payares K, Lugo LH, Morales V, Londono A (2011) Validation in Colombia of the Oswestry disability questionnaire in patients with low back pain. *Spine (Phila Pa 1976)* 36(26):E1730–E1735
  40. Pekkanen L, Kautiainen H, Ylinen J, Salo P, Hakkinen A (2011) Reliability and validity study of the Finnish version 2.0 of the Oswestry disability index. *Spine (Phila Pa 1976)* 36(4):332–338
  41. Fisher K, Johnston M (2009) 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(1):67–80
  42. Bland JM, Altman DG (1997) Cronbach's alpha. *BMJ* 314(7080):572
  43. Vianin M (2008) Psychometric properties and clinical usefulness of the Oswestry Disability Index. *J Chiropr Med* 7(4):161–163
  44. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N (2006) Danish version of the Oswestry Disability Index for patients with low back pain. Part 1: Cross-cultural adaptation reliability and validity in two different populations. *Eur Spine J* 15(11):1705–1716
  45. Yakut E, Duger T, Oksuz C et al (2004) Validation of the Turkish version of the Oswestry disability index for patients with low back pain. *Spine (Phila Pa 1976)* 29(5):581–585 (**discussion 5**)
  46. Mannion AF, Junge A, Fairbank JC, Dvorak J, Grob D (2006) Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity. *Eur Spine J* 15(1):55–65
  47. Denis I, Fortin L (2012) Development of a French-Canadian Version of the Oswestry Disability Index. *Spine* 37(7):E439–E444
  48. Joshi VD, Raiturker PP, Kulkarni AA (2013) Validity and reliability of English and Marathi Oswestry Disability Index (version 2.1a) in Indian population. *Spine (Phila Pa 1976)* 38(11):E662–E668
  49. Beurskens AJ, de Vet HC, Koke 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 (Phila Pa 1976)* 20(9):1017–1028
  50. 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(5):241–247
  51. Liu H, Tao H, Luo Z (2009) Validation of the simplified Chinese version of the Oswestry Disability Index. *Spine (Phila Pa 1976)* 34(11):1211–1216 (**discussion 7**)
  52. Vigatto R, Alexandre NM, Correa Filho HR (2007) Development of a Brazilian Portuguese version of the Oswestry Disability Index: cross-cultural adaptation, reliability, and validity. *Spine (Phila Pa 1976)* 32(4):481–486

53. Costello AB, Osborne J (2005) Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract Assess Res Eval* 10(7):1–9
54. Boscainos PJ, Sapkas G, Stilianessi E, Prouskas K, Papadakis SA (2003) Greek versions of the Oswestry and Roland-Morris Disability Questionnaires. *Clin Orthop Relat Res* 411:40–53
55. Fujiwara A, Kobayashi N, Saiki K, Kitagawa T, Tamai K, Saotome K (2003) Association of the Japanese Orthopaedic Association score with the Oswestry Disability Index, Roland-Morris Disability Questionnaire, and short-form 36. *Spine (Phila Pa 1976)* 28(14):1601–1607
56. Kim DY, Lee SH, Lee HY et al (2005) Validation of the Korean version of the Oswestry disability index. *Spine (Phila Pa 1976)* 30(5):E123–E127
57. Hashimoto H, Komagata M, Nakai O et al (2006) Discriminative validity and responsiveness of the Oswestry Disability Index among Japanese outpatients with lumbar conditions. *Eur Spine J* 15(11):1645–1650
58. Sanjaroensuttikul N (2007) The Oswestry low back pain disability questionnaire (version 1.0) Thai version. *J Med Assoc Thai* 90(7):1417–1422
59. Ramzy R (2008) Validation of the Arabic version of the Oswestry Disability Index developed in Tunisia for low back pain patients in the UAE. Stellenbosch University, Stellenbosch
60. Yu EM, Nosova EV, Falkenstein Y, Prasad P, Leasure JM, Kondrashov DG (2016) Validation of a Russian Language Oswestry Disability Index Questionnaire. *Global Spine J* 6(7):636–639
61. Lue YJ, Hsieh CL, Huang MH, Lin GT, Lu YM (2008) Development of a Chinese version of the Oswestry Disability Index version. *Spine (Phila Pa 1976)* 33(21):2354–2360

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.