Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

The Oswestry Disability Index, confirmatory factor analysis in a sample of 35,263 verifies a one-factor structure but practicality issues remain

Charles Philip Gabel[a], Antonio Cuesta-Vargas[b], Meihua Qian[c], Rok Vengust[d], Ulrich Berlemann[e], Emin Aghayev[f], Markus Melloh[g,h,i]

a. Coolum Physiotherapy Sunshine Coast, Coolum Beach, Queensland, Australia
b. Department of Physiotherapy, Faculty of Medicine, Malaga University, Malaga, Spain
c. Department of Education and Human Development, College of Education, Clemson University, 410 Tillman Hall, Clemson, SC 29634, USA
d. Department of Orthopaedic Surgery, University Medical Centre Ljubljana, Ljubljana, Slovenia
e. dasRückenzentrum, Standort Salem-Spital, Bern, Switzerland
f. Swiss RDL, Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
g. Institute for Health Sciences, School of Health Professions, Zurich University of Applied Sciences, Winterthur, Switzerland
h. Faculty of Health Sciences, Curtin University, Perth, Australia
i. Centre for Medical Research, University of Western Australia, Nedlands, Australia

E-mail addresses

- Meihua Qian - mqian@g.clemson.edu
- Charles Philip Gabel - cp.gabel@bigpond.com
- Antonio Cuesta-Vargas - acuesta@uma.es
- Rok Vengust - rok.vengust@kclj.si
- Ulrich Berlemann - uberlemann@hotmail.com
- Emin Aghayev - emin.aghayev@ispm.unibe.ch
- Markus Melloh - markus.melloh@zhaw.ch

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

4 Tables and 1 Figure

Table 1: Percentiles of Oswestry Disability Index references values (ODI-RV) classified into five categories

Table 2: Summary of the one-factor solution with or without error covariance using CFA

Table 3: Sub-group comparisons of CFA outputs—male vs. female participants

Table 4: Factor loadings from sub-group analyses

Figure 1: The second model with correlated errors.

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

ABSTRACT

Purpose

To analyze the factor structure of the Oswestry Disability Index (ODI) in a large symptomatic low back pain (LBP) population using exploratory (EFA) and confirmatory factor analysis (CFA).

Methods

Analysis of pooled baseline ODI LBP patient data from the international Spine Tango registry of EUROSPINE, the Spine Society of Europe. The sample, with *n* = 35,263 (55.2% female; age 15–99, median 59 years), included 76.1% of patients with a degenerative disease, and 23.9% of the patients with various other spinal conditions. The initial EFA provided a hypothetical construct for consideration. Subsequent CFA was considered in three scenarios: the full sample and separate genders. Models were compared empirically for best fit.

Results

The EFA indicated a one-factor solution accounting for 54% of the total variance. The CFA analysis based on the full sample confirmed this one-factor structure. Subgroup analyses by gender achieved good model fit for configural and partial metric invariance, but not scalar invariance. A possible two-construct model solution as outlined by previous researchers: dynamic-activities (personal care, lifting, walking, sex and social) and static-activities (pain, sleep, standing, travelling and sitting) was not preferred.

Conclusions

The ODI demonstrated a one-factor structure in a large LBP sample. A potential two-factor model was considered, but not found appropriate for constructs of dynamic and static activity. The use of the single summary score for the ODI is psychometrically supported. However,. practicality limitations

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

were reported for use in the clinical and research settings. Researchers are encouraged to consider a shift towards newer, more sensitive and robustly developed instruments.

Keywords

Oswestry Disability Index, Confirmatory factor analysis, Patient-reported outcome instrument, Validation, Spine Tango, Registry

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

INTRODUCTION

Measuring and monitoring the individual status and functional change in sufferers of low back pain (LBP) is critical for its overall management [1,2]. However, this measurement is not standardized and subsequently cannot systematically reflect the effectiveness of evidence-based interventions. There are over 200 PROs available for LBP measurement with the Oswestry Disability Index (ODI) [3,4] one of the most commonly used and advocated in clinical guidelines [2,4]. First published in 1980 [3], the ODI was developed to guide treatment programmes and ensure critical LBP aspects were recorded and progress monitored through measured changes in functional status. However, its development followed a qualitative item-selection process rather than a scientific clinimetric methodology [3,4,6,7]. Consequently the ODI presents a scale with 'ordinal' or 'preference-based responses' rather than 'interval' or 'precise measurement points', which can affect its validity and capacity for standard statistical analysis [8]. Despite its 40 years of wide use, it has still not been conclusively proven whether the ten ODI items can be summated into a single score [2]. The result is a lack of consensus regarding its factor structure [9-11], an important issue that needs resolution.

Factor structure is critical and demonstrates the underlyingthemes or factors present that must be recognized to indicate a parsimonious structure [12]. Factor structure can be singular, enabling a single-summated score; or two- or multi-factor, which requires separately reported scores [12,13]. The ODI has always been reported as singular [10,14,15]; however, some researchers suggest a two-factor model of: dynamic-activities (personal care, lifting, walking, sex and social) and static-activities (pain, sleep, standing, travelling and sitting) [16,17]. With Rasch analysis, which considers the evenness or interval of the scores, a suboptimal one-factor structure was found along with psychometric concerns of poor coverage, plus a large floor and small ceiling effect [18,19]. If a PRO is to use a single-summated score, a one-factor solution is required to ensure each question reports upon the same underlying construct [11-13] according to COSMIN standards [7]. The gold standard to achieve this is confirmatory factor analysis (CFA) which requires a large dataset for definitive analysis [12,20]. A CFA is validating a preceding exploratory factor analysis (EFA), which expose the underlying traits, and requires 50–100 responses-

per-item and consequently a minimum sample of *n* = 500–1000 for the ODI [12]. There is a gap in the literature as the published studies to date have performed only EFA and only in small samples. In particular, cross-cultural adaptation studies are commonly carried out on samples below n = 100 [10,17]. This is inadequate for EFA as the estimates become unstable [7,12].

Consequently, to address the existing knowledge gaps a single robust study with a large sample size greater than 10,000, or 1000 per item, would be appropriate to resolve the issue conclusively; whether a one- or a multi-factor model has a better fit. The aims of this study were to analyze the ODI factor structure in a LBP population using CFA in an adequately large sample that allows robust testing of competing models, and to determine which model is consistent across genders.

METHODS

Ethical approval was not required for this post hoc analysis of anonymous data.

Participants

This study was carried out using the Spine Tango data pool. Spine Tango, the international spine registry of EUROSPINE [21], the Spine Society of Europe is hosted at the University of Bern's Institute for Social and Preventive Medicine. Completed baseline ODI-PROs (*n* = 35,263, 55.2% female, age = 15–99, median 59-years) were obtained from symptomatic LBP patients included in the registry. The study sample comprised patients with degenerative disease (76.1%), non-generative spondylolisthesis (7.8%), pathological fracture (4.2%), repeat surgery (3.8%), deformity and traumatic fracture (2.7% each), tumour and infection (1% each), and patients with other condition (<0.8%).

Assessment tools

The ODI contains ten pain-related, six responses options questions scored from zero (no pain) to five (most severe pain). Scores are expressed as a percentage of total points, with ≤20% indicating minimal

disability, 21–40% moderate disability, 41–60% severe disability, 61–80% crippled, and 81–100% completely bed-bound [4].

Factor analysis

The EFA considers several statistics including: Eigenvalues, a special set of characteristic values associated with a linear system of equations (generally >1.0 = statistically relevant); percentage of variance explained by a particular factor ([10% = relevant); factor loading, a measure of how well any item is represented by a factor (>0.30 = minimum); and 'Scree Plot', a visual representation chart of Eigenvalues versus items (qualitatively assessed). For PRO's to provide a one-factor solution and single total score [13,15], each criteria must be fulfilled and a single-factor solution needs to be obtained [12]. When a two-factor solution is argued, the second eigenvalue must be >1 and at least 3–4 items load appropriately on the second factor and also be interpretable. An EFA statistically checks an instrument's dimensionality where the factor structure must be theoretically meaningful [12]. Subsequent CFA clarifies and validates the suggested EFA model/s using significantly larger samples [12].

Hence this study investigated the ODI factor structure through EFA from a randomly selected 10% sub-group (*n* = 3526) using SPSS 22. Then CFA was conducted on the remaining 90% (*n* = 31,736, 90%) using Mplus 7.11 [20].

In CFA, model parameters were estimated using the maximum likelihood method which is robust to non-normality [20]. The model fit was assessed using the Root Mean Square Error of Approximation (RMSEA) and the Comparative Fit Index (CFI). A RMSEA value of 0.05 or lower suggests excellent fit, and values B0.08 indicate acceptable fit [22]. For the CFI, 0.90 is considered acceptable and 0.95 or above reflects excellent model fit [23]. Additionally, modification indices (MI) were analysed to determine if allowing error terms to co-vary would significantly improve the model fit, and during the CFA, errors with MI exceeding 4.00 were allowed to correlate [20].

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

ODI references values (ODI_RV)

To fully describe the level of severity of participants' disability, an ODI-RV was created.

Sub-group analyses

Multi-group analyses were conducted to examine whether the identified model through EFA and CFA fits the data equally well for male and female participants. Namely, the degree to which a confirmatory factor model measuring LBP with ten items per six-point response scale exhibited measurement and structural invariance between male and female participants was assessed using Mplus 7.11 [20].

The original CFA model was first analyzed using the remaining 90% sample. Then the initial configural invariance model was compared with a series of models with increasing invariance constraints. Specifically: (1) the first configural invariance model constrained the pattern of fixed and free parameters to be equivalent across groups; (2) the second metric invariance model constrained factor loadings to be equal across groups; (3) the scalar invariance model constrained all factor loadings and intercepts to be equal across groups; (4) the residual variance invariance model constrained error variance to be equal across groups; (5) the residual covariance invariance model constrained error covariance to be equal across groups; (6) the factor variance invariance model constrained factor variance to be equal across groups; and (7) the factor mean invariance model constrained factor mean to be equal across groups.

Invariance between groups on a particular parameter is achieved when non-significant statistical difference is found between a model without a parameter constrained to be equal across groups and the model with the parameter constrained. Then the more parsimonious model is retained and compared to the subsequent model with additional constraints.

Assessing competing models

The most common method to assess model equivalence is a Chi-square based Likelihood ratio test, which compares the overall goodness of fit Chi-square values between the two models. However,

given Chi-square tests are highly sensitive to trivial differences in large samples [24], other measures, including the Akaike Information Criterion (AIC) and ΔCFI, were also used [25]. The ΔCFI was obtained by subtracting the CFI of compared models, where 0.01 indicates a lack of invariance [25]. The AIC measures the parsimony of two competing models, where lower values suggest better model fit [26].

If a significant, meaningful difference between two compared models exists, then fewer constraints are selected. This indicates a lack of invariance of the parameters in question across groups. The measurement variance across male and female sub-groups was evaluated through multigroup analyses.

RESULTS

Odi_rv

The ODI_RV was calculated from standardized scores classified into five categories: 'minimal', 'moderate', 'severe', 'crippling' and 'bed-bound/exaggerated' (Table 1).

Explanatory factor analysis

The initial EFA showed a one-factor structure which explained 54% of the total variance. The first eigenvalue was 5.49 and all others were <1.0. Factor loading ranged from 0.58–0.81.

Confirmatory factor analysis

The CFA confirmed a one-factor structure. Factor loadings ranged 0.53–0.81. The CFI = 0.945 and RMSEA = 0.075, suggesting adequate model fit. However, further examination of modification indices indicated that allowing some error terms to co-vary would significantly improve model fit (Fig. 1). Hence, the model was re-run to depict the second model with correlated errors (Fig. 1; Table 2). The AIC and RMSEA values of the second model decreased,^CFI increased (~0.04) and the difference in Chi-square values between the two models was significant (Table 2). Consequently the second model, with correlated errors, fit the data significantly better than the first model.

Sub-group analyses

Multi-group analyses comparing males (*n* = 14,173) and females (*n* = 17,507) demonstrated configural invariance and partial metric invariance. The configural invariance model had good fit (CFI = 0.983, RMSEA = 0.046), and partial metric invariance was achieved ($\Delta Chisquare_{configural\ vs.\ partial\ metric}$ (2) = 14.022, $p>0.05$; $\Delta CFI<0.001$; Table 3). Table 4 shows the unstandardized and standardized factor loadings that are statistically similar between male and female (see ODI 2, 4, and 8). However, scalar invariance was not achieved ($\Delta Chisquare_{partial\ metric\ vs.\ scalar}$ (2) = 101.005, $p<0.001$), although the DCFI was<0.01.

DISCUSSION

The findings from both the EFA and CFA confirmed that the ODI's one-factor structure was preferable from both the statistical perspective and parsimony. This is critical as it ensures a valid, single-summated score can be used. No appropriate two-factor model was found that is preferred to the one-factor model, but ambiguity is present. Specifically, the two-factor solution, proposed recently of dynamic and static-activities, was not preferred in the total population or either gender sub-group. This study's findings support previous research for EFA in several samples [10,15,16]. It also supports the Rasch analysis that found a one-factor structure, but it was suboptimal [18]. In our study, while the Chi-square test of the model fit was significant ($p<0.001$), it is heavily impacted by large sample size and further investigations may be optimal. The gender sub-group analysis indicated both configural invariance and partial metric invariance were obtained between men and women specifying the relationships of some items to the latent factor of disability were equivalent in both groups. However, the scalar invariance was not observed. It suggests women tend to have a slightly higher item response than men at the same absolute trait level of disability. The concerns with the ODI's practicality and consequential clinimetric performance aspects affect both the limitations and implications from clinical and research perspectives [2,7]. The influence of pain on response options is overwhelming with the iteration of similar optional answers in different sections limiting the patients'ability to express

their perceptions of their condition [9,11]. This is reflected in the large minimum detectable change (MDC) and minimum clinically relevant difference (MCID), which determine responsiveness and error [7,11]. These have been demonstrated in previous studies to be around 20–25% of baseline level [1,9,11]. This is insufficient in comparison to several other regional PROs for which the MDC is in the order of 10% or lower, and numerical rating scales have errors of around 15% in the same sample and require only a single question[14].

Consequently, the ODI as a modern viable PRO is less practical than simpler PROs that are easier to use and have smaller error scores that reduce the 'number needed to treat' (NTT). This, consequently, determines a smaller sample size and shorter time to provide meaningful results that verifies if true change has occurred and ensures statistically significant outcomes for both the individual and investigative research. The ODI is also unable to include objective parameters which limit post-operative evaluation [1,11]. By comparison, recent computer based PROs have such values represented or transferred into response options and algorithms that calculate a final single outcome score [8]. The practicality aspect of 'patient demand' to complete a PRO, expound the potential for completion errors and inconsistency [10,11]. These include excessive completion time and scoring inaccuracies, a consequence of a large number of response options and increased cognitive demand, that leads to respondent uncertainty and reduced precision [1,9,11]. Solutions to overcome these issues include shortening the PRO, modifications to improve practicality, modern scientific development methodology [10,11] and a shift toward digital software systems such as computerized adaptive testing (CAT) or computerized decision support systems (CDSS) [27] in future randomized controlled trials that incorporate objective and individual response options [1,11].

LIMITATIONS AND STRENGTHS

This study's limitations are several. As a secondary analysis, diagnostic sub-groups (e.g., spinal stenosis, radiculopathy or disc degeneration) could not be considered due to limited diagnostic codes within the data set. The implications of potential constructs of 'dynamic' and 'static' function, as

suggested by some researchers [17], could potentially have been present within the participants' occupational, social, sporting or daily routine. However, this could not be ascertained from the available data set. It is highly unlikely, from the statistical findings, that such considerations potentially influenced the analysis. If so then this would affect the overall validity of the ODI in terms of the capability of providing a single-summated score.

The dominant strength of this study is the very large sample size. The 10% EFA sample alone was over tenfold larger than all previous factor analysis studies. This is certainly one of the important benefits of registries besides implant tracking, detection of rare adverse events, early warning, benchmarking, real-life perspective and so forth [21]. Furthermore, a statistician independent of the data collectors is responsible for the data analysis.

CONCLUSION

The findings are conclusive that the one-factor solution is preferable from the perspectives of both the statistical analysis and parsimony. Consequently, the ongoing use of the ODI summary score is psychometrically supported. However, the ODI, as an outcome instrument, continues to have prominent limitations that include practicality and measurement error. Clinicians must be aware of the completion burden for patients, and that a minimum detectable change is around 20–25% of the baseline level. This may have consequences on the research. Researchers are encouraged to consider a shift towards newer, more sensitive and robustly developed instruments.

ACKNOWLEDGMENTS

The participants of the Spine Tango Register are acknowledged for their continuous contribution that makes possible such studies reflecting the daily practise of spine surgeons. The data of the following centres were used (in alphabetic order of country, city, hospital and department): Dept. of Spinal Surgery in Royal Adelaide Hospital (Australia); Dept. of Spinal Surgery in St. Andrew's Hospital in Adelaide (Australia); Dept. of Orthopaedic Surgery in Landesklinikum Krems (Austria); Dept. of

Orthopaedic Surgery in Orthopaedic Hospital Speising in Vienna (Austria); Dept. of Orthopaedic Surgery in University Hospital St. Luc in Brussels (Belgium); Dept. of Orthopaedic Surgery in Grand Hôpital de Charleroi (Belgium); Dept. of Neurosurgery in University Hospital Cologne (Germany); Dept. of Orthopaedic Surgery in University Hospital of Cologne (Germany); Dept. of Spine Surgery in Hospital Dortmund (Germany); Dept. of Orthopaedic Surgery in University Hospital of Greifswald (Germany); Dept. of Spine Surgery and Neurotraumatology in St. Nobifacius Hospital Lingen (Germany); Dept. of Neurosurgery in Orthopädisches Klinikum Markgröningen (Germany); Dept. of Neurosurgery in Klinikum Offenbach (Germany); Dept. of Orthopaedic Surgery in Asklepios Klinikum Uckermark in Schwedt (Germany); Dept. of Special Spine Surgery in Leopoldina Hospital of Schweinfurt (Germany); Dept. of Spine Surgery in Krankenhaus der Barmherzigen Brüder in Trier (Germany); Dept. of Spine Surgery in Clinica Cellini (Italy); Dept. of Spine Surgery in IRCCS Galeazzi in Milan (Italy); Dept. of Neurosurgery in Sapienza University of Rome (Italy); Dept. of Spine Surgery in Centro Medico Puerta de Hierro (Mexico); Dept. of Spine Surgery in SCTO in Chisinau (Moldova); Dept. of Neurosurgery in Wojewódzki Szpital Specjalistyczny nr 2 in Jastrzębie-Zdrój (Poland); Dept. of Neurosurgery in Specialized Medical Center S.A. Polanica (Poland); Dept. of Neurosurgery in Medical University Silesia (Poland); Dept. of Neurosurgery in General Hospital Torun (Poland); Dept. of Orthopaedic Surgery and Traumatology in Kliniczny in Wroclaw (Poland); Dept. of Orthopaedic Surgery in Tan Tock Seng Hospital (Singapore); Dept. of Orthopaedic Surgery in University Hospital of Ljubljana (Slovenia); Dept. of Neurosurgery in Bethesda Hospital of Basel (Switzerland); Dept. of Spine Surgery in Bethesda Hospital of Basel (Switzerland); Dept. of Orthopaedic Surgery in Salem Hospital of Bern (Switzerland); Dept. of Spine Surgery in University Hospital in Lausanne (Switzerland); Dept. of Spine Surgery in Hirslandenklinik Birshof in Münchenstein (Switzerland); Dept. of Spine Surgery in The Spine Center Thun (Switzerland); Dept. of Orthopaedic Surgery in Hospital Schwyz (Switzerland); Dept. of Spine Surgery in Nottingham University Hospitals NHS Trust (UK); Spine Unit of Nuffield Oxford Centre (UK); Division of Orthopaedic Surgery in SUNY Downstate Medical Center in New York (USA); Division of Spine Surgery in NYU Hospital of New York (USA).

Compliance with ethical statement.

Conflict of interest

None of the authors has any potential conflict of interest.

REFERENCES

1. Cleland JA, Gillani R, Bienen EJ, Sadosky A (2011) Assessing dimensionality and responsiveness of outcomes measures for patients with low back pain. Pain Pract 11(1):57–69
2. Chiarotto A, Maxwell LJ, Terwee CB, Wells G, Tugwell P, Ostelo R (2016) Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. Phys Ther 96(10):1620–1637
3. Fairbank JCT, Couper J, Davies JB, O'Brien JP (1980) The oswestry low back pain disability questionnaire. Physiotherapy 66(8):271–273
4. Fairbank JCT, Pynsent PB (2000) The Oswestry Disability Index. Spine 25(22):2940–2952
5. Guzman JZ, Cutler HS, Connolly J et al (2016) Patient-reported outcome instruments in spine surgery. Spine (Phila Pa 1976) 41(5):429–437
6. Ostelo RW, Deyo R, Stratford P et al (2008) Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 33(1):90–94
7. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 21(4):651–657
8. Gabel CP, Cuesta-Vargas AI, Osborne JO, Burkett B, Melloh M (2014) Confirmatory factory analysis of the neck disability index indicates a one-factor model. Spine J 14(8):1410–1416
9. Mehra A, Baker D, Disney S, Pynsent PB (2008) Oswestry Disability Index scoring made easy. Ann R Coll Surg Engl 90(6):497–499
10. Eranki V, Koul K, Fagan A (2013) Rationalization of outcome scores for low back pain: the Oswestry disability index and the low back outcome score. ANZ J Surg 83(11):871–877
11. Gabel CP, Melloh M, Burkett B, Michener LA (2013) The Spine Functional Index: development and clinimetric validation of a new whole-spine functional outcome measure. Spine J. doi:10.1016/j.spinee.2013.09.055
12. 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
13. Doward LC, McKenna SP (2004) Defining patient-reported outcomes. Value Health 7(S1):S4–S8
14. Hägg O, Fritzell P, Nordwall A, SLSS Group (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 12(1):12–20
15. van Hooff 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–E90

16. Guermazi M, Mezghani M, Ghroubi S et al (2005) The Oswestry index for low back pain translated into Arabic and validated in a Arab population. [Article in French]. Ann Readapt Med Phys 48(1):1–10
17. Tan K, Zheng M, Yang BX et al (2009) Validating the Oswestry Disability Index in patients with low back pain in Sichuan. [Article in Chinese]. Sichuan Da Xue Xue Bao Yi Xue Ban 40(3):559–561
18. Brodke DS, Goz V, Lawrence BD, Spiker WR, Neese A, Hung M (2016) Oswestry Disability Index: a psychometric analysis with 1,610 patients. Spine J 17(3):321–327
19. 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
20. Muthe´n LK, Muthe´n BO (1998–2015) Mplus user's guide, 7th edn. Muthe´n & Muthe´n, Los Angeles
21. Staub LP, Ryser C, Ro¨der C et al (2016) Total disc arthroplasty versus anterior cervical interbody fusion: use of the Spine Tango registry to supplement the evidence from randomized control trials. Spine J 16(2):136–145
22. Schumacher RE, Lomax RGA (1996) A beginner's guide to structural equation modeling. Lawrence Erlbaum, Mahwah
23. McDonald RP, Marsh HW (1990) Choosing a multivariate model: noncentrality and goodness of fit. Psychol Bull 107:247–255
24. Cheung GW, Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model 9(2):233–255
25. Byrne BM (2010) Structural equation modeling with AMOS: basic concepts, applications, and programming, 2nd edn. Routledge, New York
26. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723
27. Moja L, Passardi A, Capobussi M et al (2016) Implementing an evidence-based computerized decision support system linked to electronic health records to improve care for cancer patients: the ONCO-CODES study protocol for a randomized controlled trial. Implement Sci 11(1):153

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

365 **TABLES**

366 **Table 1**

367 Percentiles of Oswestry Disability Index references values (ODI-RV) classified into five categories

	ODI-RV			
	Total (n = 35,249)	Male (n = 15,801)	Female (n = 19,448)	Disability categories
Percentile				
<20th	−0.902	−1.004	−0.799	Minimal
<40th	−0.284	−0.387	−0.284	Moderate
<60th	0.230	0.126	0.229	Severe
<80th	0.847	0.847	0.949	Crippling
<99th	2.390	2.390	2.287	Bed-bound/exaggerated

368

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

Table 2

Summary of the one-factor solution with or without error covariance using CFA.

χ^2 value indicates the difference between observed variance–covariance matrix and the model-implied variance–covariance matrix; *p* value indicates probability of the difference; and *df* stands for the degrees of freedom. *RMSEA* the root mean square error of approximation, is a measure of model fit, with a value of 0.05 or lower suggesting excellent fit, and values <0.08 indicating reasonable fit [24]; CFI stands for the Comparative Fit Index, with 0.90 being considered acceptable, and 0.95 or above reflecting excellent model fit [24]. AIC, the Akaike Information Criterion, is a comparative measure of fit, with lower values indicating a better model fit[25]. $\Delta\chi^2$ is the difference in Chi-square values between the first model and the second model with correlated errors. Correlated errors in the second model represent that the unique variances of the associated indicators such as pain intensity and sleeping overlap (see Fig. 1 for details)

Model	χ^2	*df*	*p*	CFI	RMSEA	AIC	Significance of $\Delta\chi^2$
First model	6942.724	35	<0.001	0.945	0.075	1019349.168	
Second model with correlated errors	2083.422	29	<0.001	0.983	0.045	1013579.776	$P < 0.001$

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

Table 3

Sub-group comparisons of CFA outputs—male vs. female participants

χ2 value indicates the difference between observed variance–covariance matrix and the model implied variance–covariance matrix; p value indicates probability of the difference; and df stands for the degrees of freedom. RMSEA the root mean square error of approximation, is a measure of model fit, with a value of 0.05 or lower suggesting excellent fit, and values <0.08 indicating reasonable fit 24. CFI stands for the comparative fit index, with 0.90 being considered acceptable, and 0.95 or above reflecting excellent model fit 24. Δχ2 (14.022) is the difference in Chi-square values between the configural model and partial metric model, and Δχ2 (101.005) is the difference in Chi-square values between the partial metric model and scalar model. Partial metric invariance was achieved (p>0.05), whereas scalar invariance was not achieved (p>0.001)

Model	χ^2	*df*	*p*	$\Delta\chi^2$	CFI	*p*	RMSEA
Configural model	2186.526	58	<0.001		0.983		0.046
Partial metric model (item 2, 4 and 8)	2200.548	60	<0.001	14.022	0.983	>0.05	0.045
Scalar model	2301.553	62	<0.001	101.005	0.982	<0.001	0.045

Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

388 **Table 4**

389 Factor loadings from sub-group analyses. * Factor loadings held equal across groups

Item	Unstandardized factor loading		Standardized factor loading	
	Males	Females	Males	Females
ODI 1	0.687	0.661	0.620	0.587
ODI 2*	0.796	0.796	0.720	0.694
ODI 3	0.936	1.005	0.694	0.702
ODI 4*	0.972	0.972	0.690	0.700
ODI 5	0.743	0.676	0.598	0.552
ODI 6	0.906	0.963	0.652	0.674
ODI 7	0.642	0.595	0.565	0.505
ODI 8*	1.054	1.054	0.788	0.780
ODI 9	1.195	1.283	0.813	0.813
ODI 10	1.310	1.412	0.764	0.762

390

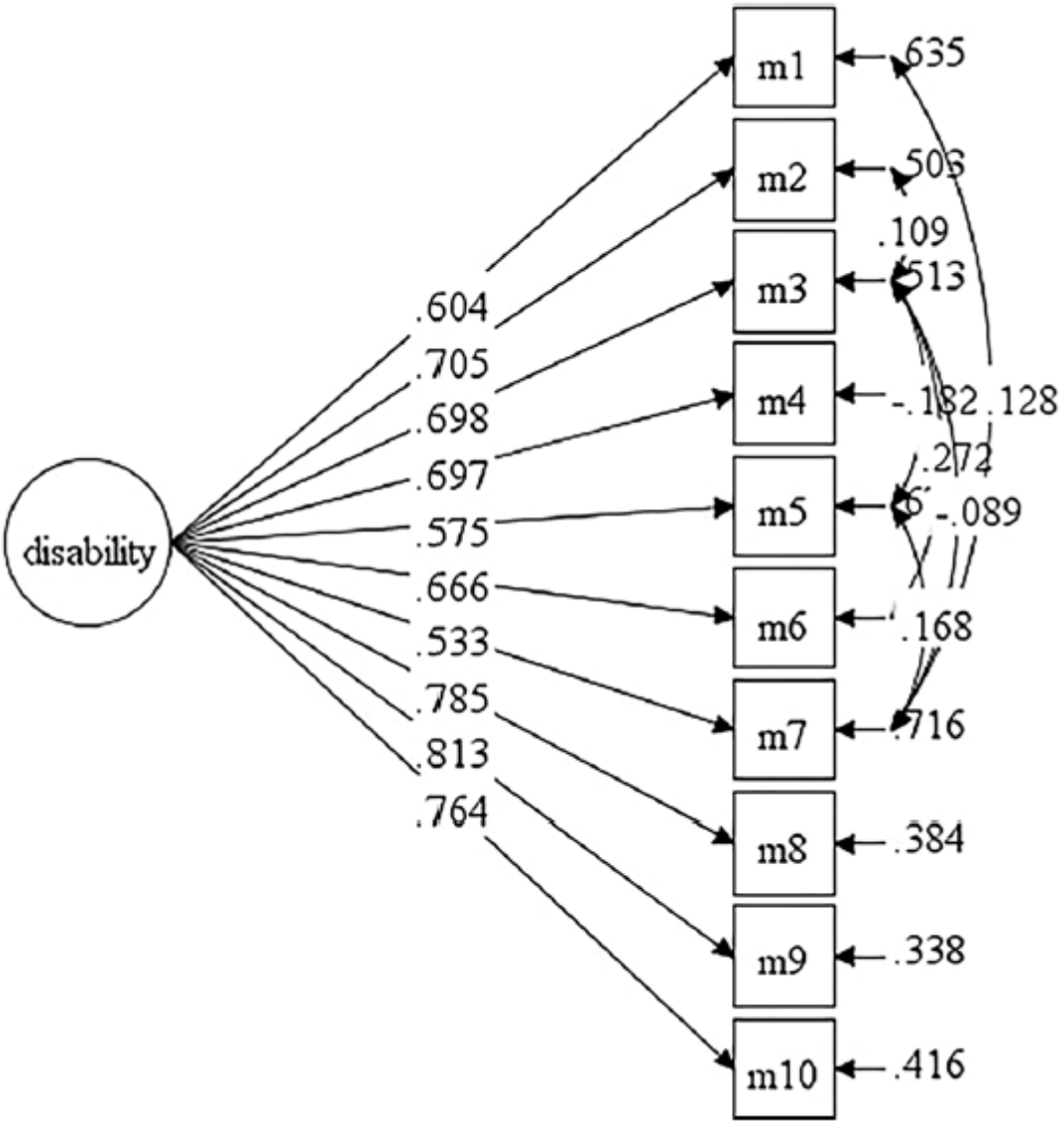
Published in final form edited form as: *Eur Spine J.* 2017 Aug;26(8):2007-2013. doi: 10.1007/s00586-017-5179-3

391 **FIGURES**

392 **Figure 1**

393 The second model with correlated errors. Disability represents the ODI, and m1–10 stand for pain

394 intensity, personal care, walking, lifting, sitting, standing, sleeping, social life, travelling, and sex life



395