

Differential item functioning analysis using the SF-36 in patients with lumbar disc herniation: Health-related quality of life research

Tania Dehesh ⁽¹⁾, Marzieh Mahmoodi ⁽²⁾, Farhad Iranmanesh ⁽³⁾, Paria Dehesh ⁽⁴⁾, Arash Farvahari ⁽¹⁾

(1) Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran

(2) Department of Biostatistics, School of Health and Nutrition, Bushehr University of Medical Sciences, Bushehr, Iran

(3) Department of Neurology, Neurology Research Center, Kerman University of Medical Sciences, Kerman, Iran

(4) Graduated of veterinary medicine faculty, Shahid Bahonar University of Kerman, Kerman, Iran

CORRESPONDING AUTHOR: Marzieh Mahmoodi, Department of Biostatistics, School of Health and Nutrition, Bushehr University of Medical Sciences, Bushehr, Iran Email: ma.mahmoodi@bpums.ac.ir, Tel: +98-77-33450134

DOI: 10.2427/12839

Accepted on July 2, 2018

ABSTRACT

Background: Differential item functioning (DIF) presents when individuals from different groups perceive the meaning of items differently in health-related quality of life (HRQoL) questionnaires. The aim of this study is to distinguish DIF in the 36-Item Short Form (SF-36) questionnaire and to determine its effect on comparison of HRQoL scores of lumbar disc herniation (LDH) patients and a healthy population.

Methods: A total of 137 patients with LDH and 691 healthy individuals filled out the Persian version of the SF-36 questionnaire. The Rasch model was used to assess DIF for patients with LDH and a healthy population.

Results: The presence of DIF was determined in 6 of 8 (75%) domain scores between patients with LDH and healthy individuals. Although half of the DIF was categorized as either negligible (3 out of 8; 37.5%), high DIF was observed in 3 out of 8 domains (37.5%). Gender was not flagged as important to DIF, with only 3 of 8 (37.5%) categorized as negligible.

Conclusion: Because the use of the SF-36 questionnaire in HRQoL assessment between groups may not be invariant, caution should be used during comparison of HRQoL scores between heterogeneous groups.

Key words: Lumbar disc herniation, Health-related quality of life, Differential item functioning, Sf-36 questionnaire

INTRODUCTION

Lower back pain (LBP) is a major public health problem in individuals aged 45 years or younger that is responsible

for considerable personnel and societal cost [1]. About 12.9% of employed individuals complain of LBP [2]. The national economic burden of LBP surpasses \$100 billion per year, primarily as a decrease in productivity [1]. A global

review on the prevalence of LBP in the adult population has shown that it has an estimated 12% prevalence, a one-month prevalence of 23%, a one-year-old prevalence of 38%, and a lifetime prevalence of approximately 40%[3]. Patients who suffer from sciatica may deal with pain during their lives with conservative treatment or may undergo surgery if the pain worsens [4]. About 90% of sciatica is caused by LDH [5] Aside from clinical diagnostics and treatment, evaluation of health-related quality of life (HRQoL) generates societal awareness that emphasizes the impact of illness on a patient's life [6]{Dehesh, 2014 #32}.

The 36-Item Short Form Survey (SF-36) is one of the most well-known generic HRQoL questionnaires. It has acceptable reliability and validity that has been confirmed in several HRQoL studies [7-9]. Although the SF-36 is one of the most used HRQoL questionnaires for assessing HRQoL between cultures and different chronic illnesses [10-15], it is important to assess whether this questionnaire has measurement equivalence when used in different populations. DIF analysis is used to examine measurement equivalence, which means that respondents from different groups similarly understand the meaning of questionnaire items [16,17] The questionnaire is designed to measure latent traits in the various items. If different groups respond to the items differently, the questionnaire is said to display DIF [8]. If a questionnaire shows DIF for different groups, it is not clear whether or not the observed disagreement in questionnaire scores between groups is an actual difference between groups caused by an underlying interest or different interpretations of the items [16-18]. The presence of DIF requires review and judgment. Although the presence of DIF is a signal that an item or domain may be biased, it does not necessarily indicate the item or domain is unfair [19].

The number of comprehensive HRQoL instruments which has published literature that shows comparison between patients group and general group are limited to 3 instruments. The Sickness Impact profile (SIP), The Medical Outcomes study Short-Form Health Survey (SF-36) and the Quality of Well-Being (QWB)[20]. Despite the relevance of LDH patients and the costly burden of this disability on healthcare budget, the HRQoL of HLD patients are rarely examined (21). We could not find any HRQoL assessment for LDH patients with SF-36 questionnaire that give us this opportunity to compare LDH HRQoL scores with the other chronic diseases scores. This study aimed to assess HRQoL for a population of LDH patients and a generally healthy population and examine DIF between the groups for the SF-36 questionnaire using the Rasch model.

METHODS

Study population

In this cross sectional study, a simple random sample of 137 patients with LDH who referred to the magnetic resonance imaging (MRI) center of Besat Clinic in Kerman

(southern Iran) were recruited from July 2016 to October 2016. Inclusion criteria were a diagnosis of LDH by MRI scan by a neurosurgeon, being 18-60 years of age and having had LBP for at least six months. Exclusion criteria were failure of the MRI scan and specialist to confirm a diagnosis of LDH.

Results from a simulation study observed that if the focal group (patient group) sample size is about 100, the reference group (healthy group) sample size should be greater than 500, in order to limit the acceptable Type I error and power [22], 691 healthy individuals over the age of 18 from a generally healthy hospital population were selected randomly. Healthy people were those who had no chronic diseases. All participants completed informed consent forms before enrollment in the study. This study was approved by the ethical committee of Kerman University of Medical Sciences and it was in accordance with the 1964 Helsinki declaration.

SF-36 questionnaire

The SF-36 questionnaire had previously been translated into Persian and its reliability and validity confirmed in Iran [23,24]. This well-known questionnaire consists of the eight domains of physical functioning (PF: 10 items), role limitation due to physical health (RP: 4 items), bodily pain (BP: 2 items), general health (GH: 5 items), Energy fatigue (EF: 4 items), social functioning (SF: 2 items), role limitations due to emotional problems (RE: 3 items) and Emotional well-being (EWB: 5 items). Each domain has a 0-100 score with 0 denoting the worst HRQoL and 100 the best HRQoL.

Statistical analysis

Differences in the baseline characteristics for patients with LDH and comparison of age and body mass index (BMI) between the LDH and healthy populations were determined using the independent-sample T test in IBM SPSS 23.

Rasch model

The two parameters Rasch model is a logistical model that consists of two parameters for each item, item difficulty and person ability. The partial credit model (PCM) is a type of Rasch model for ordinal items [25], that was used in this study. The two parameters of this model are achieved through joint maximum likelihood estimation [26]. The PCM formula is as follows:

$$P_{ij}(\theta) = \frac{e^{(\theta - b_{ij} - \tau_j)}}{1 + e^{(\theta - b_{ij} - \tau_j)}}$$

where P is the probability of a response to item i for subject j , θ is the ability parameter, b is the item location parameter, τ is the threshold location and i is the item. The most important constraint in this model is that the level of the latent trait in each item must increase monotonically. Winsteps version 3.81.1 software was used for analysis. This software uses two methods to assess DIF: the Mantel-Haenszel and Welch tests. The Welch t-test formula is as follows:

$$t = \frac{d_{i2} - d_{i1}}{\sqrt{s_{i2}^2 - s_{i1}^2}}$$

Where d_{ik} is the difficulty of item i for group k and s_{ik}^2 is the standard error of d_{ik} [27].

DIF was used to distinguish whether or not the items were responded equally between groups and also among males and females.

Common procedures for assessing DIF are Mantel-Haenszel, item response theory based methods, and logistic regression. Simulation studies from educational testing experts have found that Mantel-Haenszel method is better suited for the analysis of uniform DIF [19].

The Mantel-Haenszel (MH) procedure to DIF analysis involves the creation of K contingency tables. The sample is divided into K matched groups based on the total test scores.

Suppose that examining whether the dichotomously scored item i shows DIF for a focal group and a reference group is considered. A contingency table is used for each interval of k comparing both reference and focal groups on an individual item.

The Mantel-Haenszel estimate of the conditional odds ratio is as follow:

$$\hat{\alpha}_{MH} = \frac{\sum_k N_{R1K} N_{F0K} / N_k}{\sum_k N_{R0K} N_{F1K} / N_k}$$

Where for the k th level, N_{R1K} and N_{F1K} are the numbers of observations in the reference and focal groups, respectively, who answered correctly; N_{R0K} and N_{F0K} denote the numbers of observations in the reference and focal groups who answered incorrectly. N_k is the total number of observations [28].

The obtained α_{MH} is standardized through log transformation, centering the value around zero. The new transformed estimator Δ_{MH} is computed as follows:

$$\Delta_{MH} = -2.35 \ln(\hat{\alpha}_{MH})$$

Thus an obtained value of 0 would indicate no DIF. If items showed DIF, the severity of DIF was calculated using the Educational Testing Service (ETS) classification rules as follows: Δ_{MH} if is greater than 1.5 and P value is less than

0.05, the large DIF has accrued (C); if Δ_{MH} is greater than 1 and P value is less than 0.05 the DIF was considered to be medium (B); if the DIF did not belong to either of these categories, it was considered to be negligible (A)[28].

RESULTS

Table 1 describes the characteristics of patients with LDH in this study. As shown, LDH was reported more frequently by women ($n = 89$; 65%) than men ($n = 48$; 35%; $p = 0.001$). Nearly half of the patients had more than one herniated disc ($p = 0.733$). A substantial proportion of patients were not addicted to drugs ($p < 0.001$) and most did not smoke ($p < 0.001$). Table 2 compares the ages and BMI of patients with LDH and the healthy population. The mean (\pm SD) age of the healthy population was 48.36 ± 13.28 . Of the 691 healthy participants, 480 (69.46%) were female. There was no significant difference in BMI between the two groups ($p = 0.116$). The mean (\pm SD) BMI in the LDH group (27.45 ± 11) showed they were overweight with respect to their age range (18-60 years).

Health-related quality of life measures

The SF-36 contain eight domains for assessing Physical functioning (PF), Role limitations due to physical health (RP), Role limitations due to emotional problems (RE), Energy fatigue (EF), Emotional wellbeing (EWB), Social functioning (SF), Body Pain (BP) and General health (GH). Table 3 presents the HRQoL scores (mean \pm SD) of the LDH and healthy groups for each domain. All domains showed significantly lower scores in the LDH group than the healthy group, except for EWB and GH. Table 4 compares the domain scores between females and males overall. As seen, there were no statistically significant differences between males and females, except for the PF and EF domains.

Impact of DIF

The results of DIF analysis for the LDH and healthy groups are shown in Table 5. Out of eight SF-36 domains, six showed a degree of DIF. The EF and EWB domains showed no DIF. The RE, SF and GH showed medium DIF. The PF and RP and BP domains displayed large DIF values. The results of DIF analysis between males and females are shown in Table 6. Out of eight domains, only three revealed DIF and the other domains displayed no DIF.

Fig 1 shows the results of the DIF for domain scores between groups more clearly. Domains 4 (EF) and 5 (EWB) are within the DIF confidence boundaries; therefore, these domains showed no DIF. Domains 3 (RE), 6 (SF) and 68 (GH)

TABLE 1. Demographic characteristics of patients on LDH

Variable	N (%)	P value
Gender		
Male	48 (35)	0.001
Female	89 (65)	
Smoking		
Yes	14 (10.2)	<0.001
No	123 (89.8)	
Drug Addiction		
Yes	16 (11.7)	0.015
NO	121 (88.3)	
N of herniated Disc		
1	66 (48.2)	0.733
>1	71 (51.8)	

TABLE 2. Demographic characteristics of LDH patients and Healthy population

	Mean±SD	P value
Age		
Healthy population	48.36±13.28	0.116
LDH patient	46.37±14.83	
BMI		
Healthy population	26.01±12.81	0.220
LDH patient	27.45±11.00	
Sex (Male)		
Healthy population	211(30.5)	0.299
LDH patient	48(35)	

TABLE 3. Comparison of the SF-36 questionnaire's domains between HDL patients and healthy general population

	Healthy Mean±SD	Patient Mean±SD	t value (p value)
Scale			
Physical functioning	56.57±1.09	33.81±2.07	9.73 (<0.001)
Role limitations due to physical health	34.90±1.53	49.21±4.01	-3.33 (0.001)
Role limitations due to emotional problems	27.78±1.46	39.29±3.96	-2.73 (0.007)
Energy fatigue	50.02±0.66	44.89±1.41	3.19 (0.002)
Emotional well being	49.81±0.58	52.13±1.52	-1.57 (0.116)
Social functioning	58.20±0.85	43.94±1.89	6.83 (<0.001)
Body pain	55.74±1.00	38.30±2.12	7.16 (<0.001)
General health	44.54±0.73	46.21±1.95	-0.80 (0.423)

fell near the DIF confidence boundaries and showed medium DIF. Domains 1 (PF), 2 (RP) and 7 (BP) showed large DIF.

Fig 2 illustrates the difficulty of each domain score between groups. As shown, domains 2, 3 and 8 were more difficult for the LDH group than the healthy group and other domains were more difficult for the healthy group than the LDH group. The greatest differences were between the two groups in the domains 1,2 and 7.

DISCUSSION

The aim of this study was to determine whether or not SF-36 performed equally between groups. To our knowledge, this is the first study to assess DIF on SF-36 domains between patients with LDH and a healthy population. The results showed that 6 out of 8 domains displayed DIF, suggesting that patients with LDH and healthy individuals interpreted the meanings of these domain items differently. The largest DIF values were associated with the PF, RP and BP domains. All these domains are related and reflect a reduction in mobility and physical activity. The different perception of domain items by the two groups can be explained by the restrictions in the daily activities of the LDH group because of continuous pain and weakness which could be aggravated by position and drive [29]. LDH symptoms deteriorate upon regular flexion; thus, any physical activity that requires extension of the back will worsen the pain [Haddadi, 2016 #40] [30]. In the RE, SF and GH domains, the LDH and healthy groups revealed medium DIF. This moderate disagreement could be due to exhaustion stemming from continuous pain, which is a frequent complaint by patients with LDH [31]. Surprisingly, EF and EVB recorded no DIF and the LDH and healthy groups understood the meaning of the items in these domains similarly. This may relate to the adaptation of patients with LDH to problems with mobility and chronic back pain [32]. Clinical experience has confirmed that individuals adapt to ill health over time [33]. In the EVB

TABLE 4. Comparison of the SF-36 questionnaire’s domains between female and male

Scale	Female Mean±SD	Male Mean±SD	t value (p value)
Physical functioning	50.07±1.25	58.12±1.68	-3.77 (<0.001)
Role limitations due to physical health	38.12±1.79	34.1±2.45	1.31 (0.189)
Role limitations due to emotional problems	30.60±1.72	27.13±2.31	1.18 (0.238)
Energy fatigue	47.81±0.74	52.67±1.01	-3.85 (<0.001)
Emotional well being	48.69±0.65	50.23±0.98	-1.34 (0.182)
Social functioning	55.55±0.99	56.49±1.33	-0.56 (0.574)
Body pain	51.84±1.15	54.69±1.57	-1.44 (0.149)
General health	43.16±0.83	43.65±1.14	-1.75 (0.080)

FIGURE 1. Confidence band for the difficulty of each eight scales of SF-36 questionnaire across LDH patients and healthy population

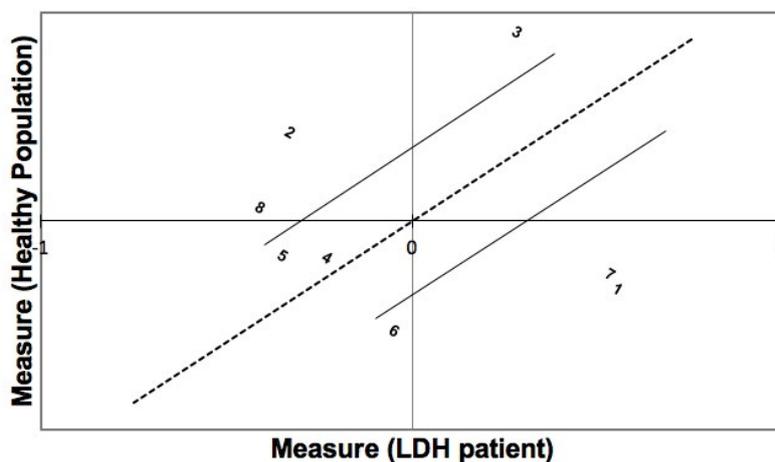
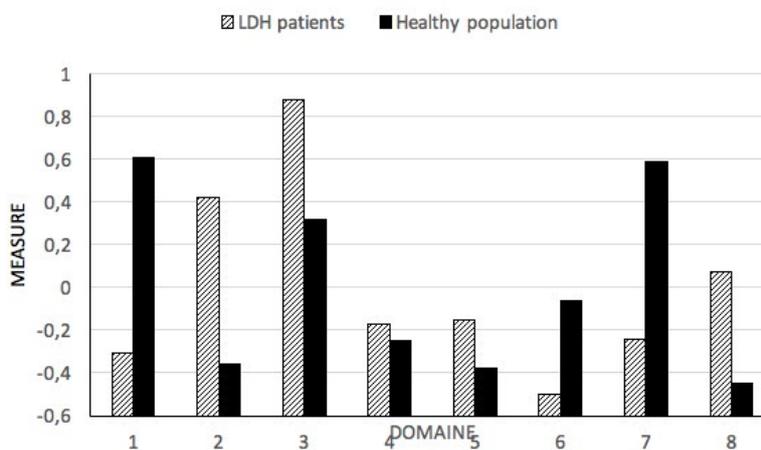


FIGURE 2. DIF plot for the SF-36 eight-dimensional questionnaire across LDH patients and healthy population



domain, the findings differed from those of other studies as being vitality related to physical health more than mental health [34], with the LBP group suffering more from mental health problems than mobility problems [35].

Because this is the first study to display DIF between LDH and healthy groups, a comparable study could not be found. Comparison of the mean HRQoL scores between LDH and healthy groups showed a lower mean score for the LBP group, except for the EWVB and GH domains. This finding contrasts with those of several previous studies that had shown that mean physical functioning in LBP patients does not differ from that of healthy individuals [36-38].

The results revealed that gender did not cause important DIF. Only three domains displayed negligible DIF. This result reveals that gender had no effect on the perceived domain items. A possible explanation for the discrepancy in PF, RP and RE could be differences in the daily activities of men and women. It is true that men were more affected by LBP than women, especially in those domains that are related to physical activity [39].

The SF-36 domain scores reveal statistically significant differences between the LDH and healthy groups HRQoL in all domains, except for EWVB and GH. In almost all domains, the scores for men show lower HRQoL scores than those for women. This finding is similar to previous findings that similar amounts of pain affect the HRQoL of men more than women [40]. Note that EWVB displayed no DIF and GH showed negligible DIF; therefore, the results of DIF analysis support the differences between HRQoL scores. The similarity of these two domain scores is more reliable than the significant differences in other domain HRQoL scores because the two groups understood the meaning of the items in these domain similarly.

The significant differences in other HRQoL domain scores may be due to different perceptions of items and do not reflect actual diverse HRQoL scores. The evidence presented in this study shows that the HRQoL domain scores deriving from LDH patients could be biased,

because the LBP and healthy groups may differently understand meanings, depending in their characteristics. Hence, caution is warranted in usage of the SF-36 when comparing HRQoL scores across LDH patient and healthy populations. The results that indicate that SF-36 operates differently across the LDH and healthy groups could be generalized to other heterogeneous group comparisons {Dimitrov, 2008 #51}[41]. Nevertheless, SF-36 is a well-known HRQoL questionnaire for assessing HRQoL in populations with varying languages and cultures [23,24], like GHQ and WHOQOL-BREF questionnaires [42]. Thus, displaying DIF does not impair the validity of the SF-36 for assessing HRQoL in a specific group [43].

This study had some limitations. The effects of demographic variables such as education, income, job and severity of disease were not assessed in DIF analysis due to lack of data. Further studies should evaluate the effect of these variables on DIF analysis for LDH patients with the SF-36 questionnaire. This study may have detected an explanation for significant HRQoL scores between populations in other studies. It is suggested that the DIF of the questionnaire can be examined between any two groups. If the equivalence is acceptable, the questionnaire can be used to assess the HRQoL between those groups.

Acknowledgments

The authors are grateful to Radan English institute for English editing of this paper.

Compliance with ethical standards

Conflict of interest

The authors declares that he has no conflict of interest.

TABLE 5. Differential Item Functioning and the SF-36 eight-dimensional questionnaire across LDH patients and healthy general population

Domain	Group	Measure 1	SE (Measure 1)	Group	Measure 2	SE (Measure 2)	Difference	SE (Difference)	Welch t	P-Value	Mantel-Haenszel	P-Value MH	ETS
Physical functioning	Healthy	-0.31	0.07	LDH	0.66	0.15	-0.92	0.17	-5.55	<0.0001	-2.162	<0.0001	C
Role limitations due to physical health	Healthy	0.42	0.07	LDH	-0.36	0.15	0.78	0.16	4.85	<0.0001	1.833	0.0006	C
Role limitations due to emotional problems	Healthy	0.88	0.07	LDH	0.32	0.15	0.56	0.16	3.42	0.0007	1.316	0.0089	B
Energy fatigue	Healthy	-0.17	0.07	LDH	-0.25	0.15	0.08	0.16	0.49	0.6211	0.188	0.6888	N
Emotional well being	Healthy	-0.15	0.07	LDH	-0.38	0.15	0.23	0.16	1.46	0.1466	0.540	0.1730	N
Social functioning	Healthy	-0.50	0.07	LDH	-0.06	0.15	-0.44	0.16	-2.73	0.0068	-1.034	0.0001	B
Body pain	Healthy	-0.24	0.07	LDH	0.59	0.15	-0.83	0.16	-5.03	<0.0001	-1.950	<0.0001	C
General health	Healthy	0.07	0.07	LDH	-0.45	0.15	0.52	0.16	3.21	0.0015	1.222	0.0001	B

ETS: Educational Testing Service criteria for DIF (A is negligible; B is medium; C is large and N is No DIF); .: Mantel-Haenszel delta; SE: standard error

TABLE 6. Differential Item Functioning and the SF-36 eight-dimensional questionnaire by sex

Domain	Group	Measure 1	SE (Measure 1)	Group	Measure 2	SE (Measure 2)	Difference	SE (Difference)	Welch t	P-Value		P-Value MH	ETS
Physical functioning	Female	-0.02	0.07	Male	-0.5	0.11	0.48	0.13	3.65	0.0003	1.128	0.0046	B
Role limitations due to physical health	Female	0.15	0.07	Male	0.63	0.11	-0.48	0.13	-3.62	0.0003	-1.128	0.0288	B
Role limitations due to emotional problems	Female	0.64	0.07	Male	1.13	0.11	-0.48	0.14	-3.56	0.0004	-1.128	0.0494	B
Energy fatigue	Female	-0.14	0.07	Male	-0.26	0.11	0.12	0.13	0.90	0.3710	0.282	0.0808	N
Emotional well being	Female	-0.19	0.07	Male	-0.19	0.11	0.00	0.13	0.00	0.9999	0.000	0.2819	N
Social functioning	Female	-0.40	0.07	Male	-0.48	0.11	0.08	0.13	0.63	0.5259	0.188	0.6610	N
Pain	Female	-0.05	0.07	Male	-0.24	0.11	0.18	0.13	1.45	0.1462	0.423	0.4879	N
General health	Female	-0.02	0.07	Male	-0.05	0.11	0.03	0.13	0.26	0.7956	0.070	0.5650	N

ETS: Educational Testing Service criteria for DIF (A negligible; B medium; C large and N is No DIF); MH α : Mantel-Haenszel delta; SE: standard error

References

- Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *JBS*. 2006;88[suppl_2]:21-4.
- Bošković K, Cigi T, Graji M, Todorović-Tomašević S, Knežević A. The quality of life of patients after a lumbar microdiscectomy: A four-year monitoring study. *Clinical neurology and neurosurgery*. 2010;112(7):557-62.
- Manchikanti L, Singh V, Falco FJ, Benyamin RM, Hirsch JA. Epidemiology of low back pain in adults. *Neuromodulation: Technology at the Neural Interface*. 2014;17(S2):3-10.
- Kamper SJ, Ostelo RW, Rubinstein SM, Nellensteijn JM, Peul WC, Arts MP, et al. Minimally invasive surgery for lumbar disc herniation: a systematic review and meta-analysis. *European Spine Journal*. 2014;23(5):1021-43.
- Valat J-P, Genevay S, Marty M, Rozenberg S, Koes B. Sciatica. *Best Practice & Research Clinical Rheumatology*. 2010;24(2):241-52.
- Dehesh T, Zare N, Jafari P, Sagheb MM. Psychometric assessment of the Persian version of the Ferrans and Powers 3.0 index in hemodialysis patients. *International urology and nephrology*. 2014;46(6):1183-9.
- Bagheri Z, Jafari P, Faghih M, Allahyari E, Dehesh T. Testing measurement equivalence of the SF-36 questionnaire across patients on hemodialysis and healthy people. *International urology and nephrology*. 2015;47(12):2013-21.
- Bagheri Z, Jafari P, Mahmoodi M, Dabbaghmanesh MH. Testing whether patients with diabetes and healthy people perceive the meaning of the items in the Persian version of the SF-36 questionnaire similarly: a differential item functioning analysis. *Quality of Life Research*. 2017;26(4):835-45.
- Diaz-Buxo JA, Lowrie EG, Lew NL, Zhang H, Lazarus JM. Quality-of-life evaluation using Short Form 36: comparison in hemodialysis and peritoneal dialysis patients. *American Journal of Kidney Diseases*. 2000;35(2):293-300.
- Pollard B, Johnston M, Dixon D. Exploring differential item functioning in the SF-36 by demographic, clinical, psychological and social factors in an osteoarthritis population. *BMC musculoskeletal disorders*. 2013;14(1):346.
- Dallmeijer AJ, de Groot V, Roorda LD, Schepers VP, Lindeman E, van den Berg LH, et al. Cross-diagnostic validity of the SF-36 physical functioning scale in patients with stroke, multiple sclerosis and amyotrophic lateral sclerosis: a study using Rasch analysis. *Journal of rehabilitation medicine*. 2007;39(2):163-9.
- Horner-Johnson W, Krahn GL, Suzuki R, Peterson JJ, Roid G, Hall T, et al. Differential performance of SF-36 items in healthy adults with and without functional limitations. *Archives of physical medicine and rehabilitation*. 2010;91(4):570-5.
- Taylor WJ, McPherson KM. Using Rasch analysis to compare the psychometric properties of the Short Form 36 physical function score and the Health Assessment Questionnaire disability index in patients with psoriatic arthritis and rheumatoid arthritis. *Arthritis Care & Research*. 2007;57(5):723-9.
- Yu YF, Yu AP, Ahn J. Investigating differential item functioning by chronic diseases in the SF-36 health survey: a latent trait analysis using MIMIC models. *Medical care*. 2007;45(9):851-9.
- Perkins AJ, Stump TE, Monahan PO, McHorney CA. Assessment of differential item functioning for demographic comparisons in the MOS SF-36 health survey. *Quality of Life Research*. 2006;15(3):331-48.
- Kim ES, Yoon M. Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*. 2011;18(2):212-28.
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*. 2000;3(1):4-70.
- Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Quality of Life Research*. 2007;16(1):33-42.
- Abedalaziz N. Detecting gender related DIF using logistic regression and Mantel-Haenszel approaches. *Procedia-Social and Behavioral Sciences*. 2010;7:406-13.
- Schaller A, Dejonghe L, Haastert B, Froboese I. Physical activity and health-related quality of life in chronic low back pain patients: a cross-sectional study. *BMC musculoskeletal disorders*. 2015;16(1):62.
- Andresen EM, Meyers AR. Health-related quality of life outcomes measures. *Archives of physical medicine and rehabilitation*. 2000;81:S30-S45.
- Woods CM. Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*. 2009;44(1):1-27.
- Montazeri A, Goshtasebi A, Vahdaninia M, Gandek B. The Short Form Health Survey (SF-36): translation and validation study of the

- Iranian version. *Quality of life research*. 2005;14(3):875-82.
24. Jafari H, Lahaesaeizadeh S, Jafari P, Karimi M. Quality of life in thalassemia major: reliability and validity of the Persian version of the SF-36 questionnaire. *Journal of postgraduate medicine*. 2008;54(4):273.
 25. Masters GN. Partial credit model. *Handbook of Item Response Theory, Volume One: Chapman and Hall/CRC*; 2016. p. 137-54.
 26. Linacre JM. A user's guide to WINSTEPS MINISTEP Rasch-model computer programs. Chicago IL: Winsteps.com. 2006.
 27. Lord F. *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ, Lawrence Erlbaum Ass. 1980.
 28. Zwick R. A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*. 2012;2012(1).
 29. García Cosamalón J, Del Valle ME, Calavia MG, García Suárez O, López Muñoz A, Otero J, et al. Intervertebral disc, sensory nerves and neurotrophins: who is who in discogenic pain? *Journal of anatomy*. 2010;217(1):1-15.
 30. Haddadi K. Pediatric lumbar disc herniation: a review of manifestations, diagnosis and management. *Journal of Pediatrics Review*. 2016;4(1).
 31. Teli M, Lovi A, Brayda-Bruno M, Zagra A, Corriero A, Giudici F, et al. Higher risk of dural tears and recurrent herniation with lumbar micro-endoscopic discectomy. *European spine journal*. 2010;19(3):443-50.
 32. Mann R, Brazier J, Tsuchiya A. A comparison of patient and general population weightings of EQ 5D dimensions. *Health Economics*. 2009;18(3):363-72.
 33. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes: Oxford university press*; 2015.
 34. Bingefors K, Isacson D. Epidemiology, co-morbidity, and impact on health-related quality of life of self-reported headache and musculoskeletal pain—a gender perspective. *European journal of pain*. 2004;8(5):435-50.
 35. Rampersaud YR, Ravi B, Lewis SJ, Stas V, Barron R, Davey R, et al. Assessment of health-related quality of life after surgical treatment of focal symptomatic spinal stenosis compared with osteoarthritis of the hip or knee. *The Spine Journal*. 2008;8(2):296-304.
 36. Bousema EJ, Verbunt JA, Seelen HA, Vlaeyen JW, Knottnerus JA. Disuse and physical deconditioning in the first year after the onset of back pain. *Pain*. 2007;130(3):279-86.
 37. Weering M, Vollenbroek-Hutten M, Tönis T, Hermens H. Daily physical activities in chronic lower back pain patients assessed with accelerometry. *European journal of pain*. 2009;13(6):649-54.
 38. Verbunt JA, Westerterp KR, van der Heijden GJ, Seelen HA, Vlaeyen JW, Knottnerus JA. Physical activity in daily life in patients with chronic low back pain. *Archives of physical medicine and rehabilitation*. 2001;82(6):726-30.
 39. Ono R, Higashi T, Takahashi O, Tokuda Y, Shimbo T, Endo H, et al. Sex differences in the change in health-related quality of life associated with low back pain. *Quality of Life Research*. 2012;21(10):1705-11.
 40. Affleck G, Tennen H, Keefe FJ, Lefebvre JC, Kashikar-Zuck S, Wright K, et al. Everyday life with osteoarthritis or rheumatoid arthritis: independent effects of disease and gender on daily pain, mood, and coping. *Pain*. 1999;83(3):601-9.
 41. Dimitrov DM. *Quantitative research in education*. New York, NY: Whittier. 2008.
 42. Zare N, Sharif F, Dehesh T, Moradi F. General health in the elderly and younger adults of rural areas in Fars province, Iran. *International journal of community based nursing and midwifery*. 2015;3(1):60.
 43. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. *Journal of Educational measurement*. 1990;27(4):361-70.

