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 treatments, 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 that 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 HLH 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:

$$P_i j (\theta) = \frac{e^{\theta - c_{ij} - r_j}}{1 + e^{\theta - c_{ij} - r_j}}$$
where $P$ is the probability of a response to item $i$ for subject $j$, $\theta$ is the ability parameter, $b$ is the item location parameter, $t$ 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 test formula is as follows:

$$t = \frac{d_{ij} - d_{ik}}{\sqrt{s^2_{d_{ij}} - s^2_{d_{ik}}}}$$

Where $d_{ij}$ is the difficulty of item $i$ for group $k$ and $s^2_{d_{ik}}$ 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 follows:

$$\Delta_{MH} = \frac{\sum_k n_{11k} n_{00k}/n_k}{\sum_k n_{01k} n_{10k}/n_k}$$

Where for the $k$th level, $n_{11k}$ and $n_{01k}$ are the numbers of observations in the reference and focal groups, respectively, who answered correctly; $n_{00k}$ and $n_{10k}$ 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 $\alpha_{\text{MH}}$ is standardized through log transformation, centering the value around zero. The new transformed estimator $\Delta_{\text{MH}}$ is computed as follows:

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

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: $\Delta_{\text{MH}}$ if is greater than 1.5 and $P$ value is less than 0.05, the large DIF has accrued (C); if $\Delta_{\text{MH}}$ 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 (± 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 (± 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 ± 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)
Differential item functioning analysis using the SF-36 in patients with lumbar disc herniation: Health-related quality of life research

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

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

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 EVWB 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 EVWB 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 effect the HRQoL of men more than women [40]. Note that EVWB 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 derived 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 WHOQOLBREF 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>
<thead>
<tr>
<th>Table 5. Differential Item Functioning and the SF-36 eight-dimensional questionnaire across LDH patients and healthy general population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>Physical functioning</td>
</tr>
<tr>
<td>Role limitations due to physical health</td>
</tr>
<tr>
<td>Role limitations due to emotional problems</td>
</tr>
<tr>
<td>Energy fatigue</td>
</tr>
<tr>
<td>Emotional well being</td>
</tr>
<tr>
<td>Social functioning</td>
</tr>
<tr>
<td>Body pain</td>
</tr>
<tr>
<td>General health</td>
</tr>
</tbody>
</table>

ETS: Educational Testing Service criteria for DIF (A is negligible; B is medium; C is large and N is No DIF); \( \Delta \): Mantel-Haenszel delta; SE: standard error
TABLE 6. Differential Item Functioning and the SF-36 eight-dimensional questionnaire by sex

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

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

References


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.


*