RESPONSIVENESS OF ARABIC INSTRUMENTS FOR PAIN AND DISABILITY IN PATIENTS WITH LOW BACK PAIN

1Fahad Alanazi
2Hamad S. Al Amer

ABSTRACT

Background: Fear-Avoidance Beliefs Questionnaire (FABQ), Quebec Back Pain Disability Scale (QDS), and Roland-Morris Disability Questionnaire (RMDQ) is widely used in patients with low back pain (LBP) to assess the level of disability. Nonetheless, there are limited data about the responsiveness properties of the Arabic versions of these scales. This study was conducted to assess the responsiveness of the Arabic versions of the FABQ, QDS, and RMDQ compared to that of the Visual Analog Scale (VAS).

Methods: A sample of 68 patients with LBP completed FABQ, QDS, RMDQ, and VAS at baseline and after 14 days. Responsiveness was evaluated by calculating the standard error of measurement (SEM), the minimal detectable difference at 95% confidence level (MDD95%), standardized response mean (SRM), Cohen’s effect size (ES), Guyatt’s responsiveness index (GRI), area under the curve (AUC), and minimal clinically significant difference (MCID).

Results: The SEM, MDD95%, SRM, ES, GRI, AUC, and MCID for FABQ, QDS, RMDQ, and VAS were 2.54, 2.83, 0.77, and 0.82; 7.05, 7.85, 2.14, and 2.28; 0.67, 0.96, 0.74, and 1.04; 0.39, 0.39, 0.36, and 0.79; 0.76, 1.34, 1.26, and 1.66; 0.49, 0.63, 0.57, and 0.70; and 3.5, 4.5, 2.5, and 1.5; respectively.

Conclusion: Although the responsiveness of the Arabic versions of FABQ, QDS, and RMDQ was below the recommended standards and less than the responsiveness calculated for the VAS, it was comparable with previously published versions in other languages. Additional studies are necessary to examine the three scales’ responsiveness with a more extended follow-up period.

Keywords: Fear-Avoidance Beliefs Questionnaire, Quebec Back Pain Disability Scale, Roland-Morris Disability Questionnaire, Arabic, responsiveness, low back pain.

Received 18th August 2020, accepted 28th November 2020, published 09th December 2020

CORRESPONDING AUTHOR

1Fahad Alanazi
Department of Physical Therapy, Faculty of Applied Medical Sciences, University of Tabuk, Tabuk, Saudi Arabia.
e-Mail: Fsalenezi@ju.edu.sa
INTRODUCTION

Low back pain (LBP) is a widely distributed condition that affects around 80% of the world's general population [1]. Such ailing may arise due to several pathological conditions. Apart from the pain, this condition, if not treated, might eventually lead to substantial functional disability, which affects the daily activities of the affected person. In the long term, LBP might also affect the psychology of the affected individual [2]. In the absence of appropriate diagnostic and management techniques, chronic and recurrent LBP is often considered a challenging disorder [3]. In addition to the physiological and psychological issues, LBP is also responsible for economic stress for affected persons regarding health care and work absenteeism [4].

Several epidemiological studies and clinical trials have started employing the self-report outcome measures to. Physicians and clinicians extensively use such measures to evaluate patients' recovery status from various types of ailments. These tools' responsiveness is known as the capacity to measure the clinically significant differences between two different occasions. Essentially, it indicates the clinical relevance of the data produced by such tools. These tools generally comprise of questionnaires and surveys that pertain to the well-being of the patient. Examples of such instruments include the Fear-Avoidance Beliefs Questionnaire (FABQ) [7], Quebec Back Pain Disability Scale (QDS) [8], Roland-Morris Disability Questionnaire (RMDQ) [9], Modified Oswestry Disability Questionnaire (MODQ) [10]. Each instrument serves a specific purpose, and most of these instruments have been extensively studied and considered reliable. Figure 1 shows the major pain assessment tools used today.

However, the major challenge lies in the absence of a culturally adapted, reliable instrument. The transcultural adaptation involves translating the questionnaire into the desired language, retaining the content validity, and making it semantically equivalent to the original questionnaire. The items are required to be culturally relevant to the target population[5]. Also, these instruments' psychometric/measurement properties must also be assessed while developing and validating their adapted versions. Some of the parameters used for psychometric validation include internal consistency, factor analysis, reliability, and responsiveness.

Here, we assessed the responsiveness of the Arabic versions of FABQ, QDS, and RMDQ and compared them with this questionnaire was developed to assess the unfavorable experiences and behaviors associated with the pain [7]. This scale consists of 16 items. Each item is scored from 0 (do not agree at all) to 6 (completely agree), with a score extending from 0 (no fear avoidance behavior) to 96 (extreme fear avoidance behavior) [7]. QDS is primarily used to measure the disability of LBP patients. It assesses the difficulty levels of the patients to perform even elementary daily activities. It consists of 24 items divided into six domains based on the functional aspect that they evaluate. Items are scored as 0 (no difficulty) to 5 (extreme disability) [8]. Consisting of 24 items, RMDQ is also used to assess the physical disability that arises due to LBP. This questionnaire is based on the Sickness Impact Profile. Each item in the questionnaire is scored either 0 (not applicable) or 1 (applicable), making a grand score that ranges between 0 and 24, representing no disability and severe disability, respectively [9].

VAS is used to assess the patient’s condition based on the perception of pain. Thus, it directly evaluated the pain intensity in LBP patients. It is a simple instrument in which the patient self-assessed and reports the intensity of pain as 0 (no pain) to 10 (worst possible pain) [10].

**METHODS**

**Participants**

Sixty-eight Arab patients with LBP were recruited using consecutive sampling from different hospitals in Tabuk, Saudi Arabia, based on the following criteria: diagnosed with acute or chronic LBP, between the ages of 18 and 65, and can read Arabic. Pregnant female patients and patients with neurological conditions, psychiatric illnesses, or malignant diseases were excluded. Approvals from the appropriate ethical committee were obtained to conduct this study.

**PROCEDURE**

A booklet containing the Arabic versions of the FABQ [12], QDS [13], RMDQ [14], and VAS was completed by the patients at baseline. Two weeks later, the participants filled the same booklet again in addition to a 7-level global rating of change (GRC) scale to identify the patients who experienced variations in their conditions compared to baseline. The patients whose answers were “about the same” or “a little better” or “a little worse” were considered as stable cases. From baseline to the second assessment, the patients continued their assigned medical interventions by the healthcare providers.

**Data analysis**

Responsiveness of the Arabic FABQ, QDS, RMDQ, and VAS was assessed using distribution and anchor-based approaches. The distribution-based approach included the standard error of measurement (SEM), the minimal detectable difference at 95% confidence level (MDD), standard response mean (SRM), Cohen’s effect size (ES), and Guyatt’s responsiveness index (GRI). The formulas for calculating each statistical parameter are listed in Table 1. The intraclass correlation coefficient (ICC) values were also calculated.

**Table 1:** Formulas used for calculating statistical parameters used in a distribution-based approach for responsiveness assessment.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Formula</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEM</td>
<td>$SD \sqrt{(1-ICC)}$ [15]</td>
<td>-</td>
</tr>
<tr>
<td>MDD$_{95%}$</td>
<td>$1.96 \times SD \times SEM$ [15]</td>
<td>-</td>
</tr>
<tr>
<td>SRM</td>
<td>$M_{post} - M_{pre}/SD_{change}$ [15]</td>
<td>$\leq 0.20$: small effect, $0.50$: moderate effect, $0.80$: large effect [15]</td>
</tr>
<tr>
<td>ES</td>
<td>$M_{post} - M_{pre}/SD_{pre}$ [15]</td>
<td>-</td>
</tr>
<tr>
<td>GRI</td>
<td>$M_{improved} / SD_{stable}$ group [10]</td>
<td>-</td>
</tr>
</tbody>
</table>


Anchor-based approach comprised a receiver operating characteristic (ROC) curve from which the area under the curve (AUC) for each scale was calculated. AUC ranges between 0.5 (no discriminating accuracy) and 1 (optimal discriminating accuracy) [16]. The minimal clinically important difference (MCID) was the point linked with the highest sensitivity and specificity in the ROC curve. Finally, responsiveness was assessed by analyzing relationships between the changes in scores of FABQ, QDS, and RMDQ with the GRC using Spearman’s rho and with the changes in scores of the VAS utilizing Pearson’s r. Furthermore, the relationships between the changes in scores of FABQ, QDS, and RMDQ were also calculated. Both Spearman’s rho and Pearson’s r were interpreted as described by Portney and Watkins [15]. All analyses were performed using SPSS 25.0 at a 0.05 alpha level of significance.

**RESULTS**

**Participants**

The mean age ± standard deviation of the patients was 37.01 ± 7.57 years. Out of the 68 patients, 55 were males, and 13 were females. Sixteen participants had a high school education, 12 had a postsecondary diploma, and the rest had a university education. Fifty-nine patients were employed at the time of the study, while four were students, and the rest were unemployed. LBP duration was between 3 weeks and 3 months for 37 patients, and more than 3 months for 31 patients. Fifty-three out of 68 patients were classified as stable based on their GRC scores after 14 days. The means and standard deviations of the participants’ scores of the Arabic FABQ, QDS, RMDQ, and VAS at baseline and follow-up assessments, in addition to the GRC scores after 2 weeks, are shown in Table 2.

**Table 2:** Means and standard deviations for scores of FABQ, QDS, RMDQ, VAS, and GRC.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Baseline</th>
<th>Follow-up</th>
<th>Change score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>FABQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (n = 68)</td>
<td>25.5</td>
<td>7.2</td>
<td>22.7</td>
</tr>
<tr>
<td>Stable (n = 53)</td>
<td>26.5</td>
<td>7.2</td>
<td>23.8</td>
</tr>
<tr>
<td>Improved (n = 15)</td>
<td>22.3</td>
<td>6.5</td>
<td>19.0</td>
</tr>
<tr>
<td>QDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (n = 68)</td>
<td>31.8</td>
<td>11.1</td>
<td>27.5</td>
</tr>
<tr>
<td>Stable (n = 53)</td>
<td>32.9</td>
<td>10.9</td>
<td>29.1</td>
</tr>
<tr>
<td>Improved (n = 15)</td>
<td>28.0</td>
<td>11.1</td>
<td>22.1</td>
</tr>
<tr>
<td>RMDQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (n = 68)</td>
<td>7.3</td>
<td>3.1</td>
<td>6.2</td>
</tr>
<tr>
<td>Stable (n = 53)</td>
<td>7.4</td>
<td>3.0</td>
<td>6.4</td>
</tr>
<tr>
<td>Improved (n = 15)</td>
<td>7.1</td>
<td>3.3</td>
<td>5.5</td>
</tr>
<tr>
<td>VAS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (n = 68)</td>
<td>4.8</td>
<td>1.6</td>
<td>3.5</td>
</tr>
<tr>
<td>Stable (n = 53)</td>
<td>4.9</td>
<td>1.7</td>
<td>3.8</td>
</tr>
<tr>
<td>Improved (n = 15)</td>
<td>4.5</td>
<td>1.2</td>
<td>2.5</td>
</tr>
<tr>
<td>GRC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (n = 68)</td>
<td>-</td>
<td>-</td>
<td>4.2</td>
</tr>
<tr>
<td>Stable (n = 53)</td>
<td>-</td>
<td>-</td>
<td>4.5</td>
</tr>
<tr>
<td>Improved (n = 15)</td>
<td>-</td>
<td>-</td>
<td>2.8</td>
</tr>
</tbody>
</table>
Responsiveness

Table 3 shows the responsiveness values of FABQ, QDS, RMDQ, and VAS. The SEM and MDD\textsubscript{95\%} for FABQ, QDS, RMDQ, and VAS were 2.54 and 7.05, 2.83 and 7.85, 0.77 and 2.14, and 0.82, 2.28, respectively. The responsiveness of FABQ, QDS, and RMDQ ranged from 0.67 to 0.96 and 0.76 to 1.34 as indicated by the SRM and GRI, respectively, representing moderate to large effect size. These values are comparable to the SRM and GRI values obtained for VAS. However, using Cohen’s method, the responsiveness values decreased to 0.39 for FABQ and QDS, and 0.36 for RMDQ, which indicated a small effect size.

Graphs 1-4 illustrate the ROC curves for the FABQ, QDS, RMDQ, and VAS. The highest AUC value was found for QDS (0.63), followed by RMDQ (0.57). Although not statistically significant, these values were just above the minimum level cut-off value of 0.5 for discrimination accuracy. They were still significantly lower than the AUC value obtained for the VAS (0.70) (Table 3). FABQ had the least AUC value of 0.49. The MCID scores for FABQ, QDS, RMDQ, and VAS were 3.5, 4.5, 2.5, and 1.5, respectively. The MCID value for RMDQ was comparable to the MDD\textsubscript{95\%} value (2.5 vs. 2.14). The MCID values of the rest of the parameters were less than their respective MDD\textsubscript{95\%} values (Table 3).
The correlation coefficients for changes in the scores of various parameters are summarized in Table 4. The correlation values for changes in the scores of three scales, FABQ, QDS, and RMDQ, were statistically significant, and ranged from 0.26 to 0.44, which indicated fair relationships. However, the relationships between FABQ, QDS, and RMDQ with GRC and changes in the scores of VAS were weak.

Table 4: Correlation between the changes in scores of FABQ, QDS, and RMDQ with GRC and VAS.

<table>
<thead>
<tr>
<th>Scale</th>
<th>FABQ</th>
<th>QDS</th>
<th>RMDQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>FABQ</td>
<td>-</td>
<td>r = 0.43**</td>
<td>r = 0.26*</td>
</tr>
<tr>
<td>QDS</td>
<td>r = 0.19</td>
<td>r = 0.03</td>
<td>r = 0.13</td>
</tr>
<tr>
<td>VAS</td>
<td>r = 0.19</td>
<td>r = 0.25</td>
<td>r = 0.19</td>
</tr>
<tr>
<td>GRC</td>
<td>rho = 0.06</td>
<td>rho = 0.25</td>
<td>rho = 0.19</td>
</tr>
</tbody>
</table>


*Significant at α = 0.05.
**Significant at α = 0.01.

DISCUSSION

LBP strongly affects the functional status of an individual [2]. The pain induced by this condition also afflicts pain avoidance behaviors among the patients. Here, we assessed the pain intensity, fear, functional disability, related behavioral changes among the patients using Arabic versions of the self-assessment tools. Among the total of 68 patients recruited in this study, most of the LBP patients were adults (mean age: 37.01 ± 7.57 years), male (80.9%), and working (86.8%). However, we did not assess the variation in the instruments’ scores based on the demographic characteristics of the individuals. However, we used GRC to classify the patients based on whether their condition remained stable or improved due to intervention. As mentioned above, several approaches are used to measure the responsiveness of FABQ, RMDQ, and QDS; however, none of these approaches have been deemed to be the best. The utilization of different approaches may result in variation in the measurement in the same study.

On the contrary, using the same method in different studies also gives rise to inconsistency in the measurements [17]. In this study, we used the distribution-based methods, which help assess the variability of sample and data precision and anchor-based methods, which help assess the longitudinal change in the health status [18]. Parameters such as MDD and MCID were used to evaluate the sensitivity of each instrument.

Overall, among the three instruments, the maximum change was observed in the scores of QDS, while the minimum change was observed in the RMDQ scores. The change in the VAS scores was comparable to that observed for RMDQ. Similar trends in terms of change in scores were observed for the patients in the 'Improved' group and the 'Stable' group.

The ICC values of all three scales and VAS were significantly high, showing good reliability. This result corroborated the findings of the previous studies. In 2016, Terho et al. [19] validated the reliability of FABQ for the Finnish population. They confirmed good reliability (0.91) with a high Cronbach’s α score, which indicated the adapted version’s high internal consistency. The ICC value obtained for QDS was similar to those obtained from previous studies employing other adapted versions of this instrument, including the original version [8,20].

Furthermore, we observed MCID and MDD comparable values for RMDQ, which indicated high sensitivity and specificity. However, we observed a higher AUC, SRM, and ES for QDS, which indicated better performance of QDS for the assessment of the pain-induced patients’ disability. However, the maximum values of AUC, SRM, and ES were observed for VAS. Only the AUC of VAS was significant, which indicated a higher accuracy in predicting an individual’s condition. Our findings concerning the effect size did not corroborate the previous findings. For instance, Monticone et al. [21] reported a higher MDD (4.87) and ES (0.68) for RMDQ than that observed in this study. However, they reported a comparable MDIC and AUC values for RMDQ.

None of the measures showed any significant correlation with either VAS or GRC. These results were in disagreement with the findings of the previous studies. In 2015, Al-Abbad and Al-Howimel [22] assessed the reliability of the Arabic version of RMDQ for chronic LBP patients. They reported a positive association between VAS and RMDQ scores (r = 0.299; P < 0.06). In another study on the Finnish population, Terho et al. [19] reported a positive correlation between VAS and FABQ (p-value = 0.02). In a recent Egyptian study, Salama et al. [23] also reported a positive correlation between VAS and FABQ. In a recent study on the Hindi-speaking population, Zaidi et al. [24] described a significant relationship between the VAS and QDS (P < 0.001). Chung et al. [25] reported a significant positive association between VAS and FABQ and RMDQ. These findings might be primarily attributed to the small sample size and a small follow-up period. However, we observed a positive correlation of FABQ with QDS and RMDQ and between QDS and RMDQ, in agreement with the earlier findings. Zaidi et al. [24] showed a significant association between RMDQ and QDS (p<0.001) in the Hindi-speaking population. However, the correlation coefficient values were lower than those obtained in other studies on validation of other adapted versions of these scales, such as the English version (QDS vs. RMDQ r = 0.77) [8], the Turkish version (QDS vs. RMDQ r = 0.68) [26], the Iranian version (QDS vs. RMDQ r = 0.76) [20], and the Korean version (FABQ vs. RMDQ r = 0.45) [27].

To our knowledge, this was the first study to assess the responsiveness of the Arabic versions of FABQ, RMDQ, and QDS and to evaluate their correlation with VAS. We believe that our findings could help the clinicians and physicians easily assess the condition and outcomes of the...
local Arabic-speaking population.

There were a few limitations of this study. The number of patients recruited in this study was low, which might raise problems during the generalization of our results. Our follow-up duration was small, which might lead to biased results because the patients’ condition might further improve over time, which could lead to lower scores during follow-up. Furthermore, the type and duration of interventions that the patients received were not taken into account. This might lead to a discrepancy with respect to overall responsiveness and MCID. The effects of other factors, such as the psychological and sociodemographic factors, should also have been considered. This study’s statistical data must be considered an overall guideline since our approach restricted further division of the study sample. Furthermore, our results differed from the general trends that are observed in previous findings. For instance, none of the tools showed association with VAS.

CONCLUSION

LBP is a globally prevalent disorder that affects around 80% of adults. The clinicians are gradually shifting their focus towards the patient’s self-assessment, making the self-report questionnaires and surveys extremely important. Furthermore, it is also important to accurately adapt the already available self-report instruments to different world subpopulations. In the present study, the responsiveness of Arabic versions of FABQ, QDS, and RMDQ was evaluated and compared with VAS. Overall, we observed good reliability, acceptable responsiveness, and low sensitivity for FABQ and QDS for Arabic patients with LBP. However, RMDQ showed high sensitivity and specificity compared to the other two scales, along with comparable MDD90% and MCID. However, on analysis of AUC, SRM, and ES, the effect size of QDS was observed to be the largest among the three scales. However, none of those mentioned above values were significant. The ICC values of all the scales were significantly high, which indicated good reliability among all the Arabic versions.

Furthermore, all three scales showed significant positive correlations with each other—however, scores of none of the scales correlated with VAS scores. Future studies need to be performed with larger samples and longer follow-up periods, which might provide better data on the three scales’ sensitivities.

Acknowledgments

The authors would like to thank Dr. Mohamed ELdesoky and Dr. Ayman Honin for their data collection help.

REFERENCES


