ABSTRACT

Background: Measurement equivalence is an essential prerequisite for making valid comparisons in mental health questionnaires across groups. In most methods used for assessing measurement equivalence, which is known as Differential Item Functioning (DIF), latent variables are assumed to be continuous. Objective: To compare a new method called Latent Class Regression (LCR) designed for discrete latent variable with the multiple indicators multiple cause (MIMIC) as a continuous latent variable technique to assess the measurement equivalence of the 12-item General Health Questionnaire (GHQ-12), which is a cross deferent subgroup of Iranian nurses. Methods: A cross-sectional survey was conducted in 2014 among 771 nurses working in the hospitals of Fars and Bushehr provinces of southern Iran. To identify the Minor Psychiatric Disorders (MPD), the nurses completed self-report GHQ-12 questionnaires and sociodemographic questions. Two uniform-DIF detection methods, LCR and MIMIC, were applied for comparability when the GHQ-12 score was assumed to be discrete and continuous, respectively. Results: The result of fitting LCR with 2 classes indicated that 27.4% of the nurses had MPD. Gender was identified as an influential factor of the level of MPD. LCR and MIMIC agree with detection of DIF and DIF-free items by gender, age, education and marital status in 83.3, 100.0, 91.7 and 83.3% cases, respectively. Conclusions: The results indicated that the GHQ-12 is to a great degree, an invariant measure for the assessment of MPD among nurses. High convergence between the two methods suggests using the LCR approach in cases of discrete latent variable, e.g. GHQ-12 and adequate sample size.

Key words: mental disorders, measurement equivalence, latent class regression, MIMIC model, GHQ-12.

1. INTRODUCTION

In recent years, the methods used in assessing validity and reliability of psychological instruments have changed from classical test theory (e.g. factor analysis) to modern approaches such as differential item functioning (DIF) analysis (1). Based on the questionnaires used in different groups such as age, gender, race or culture, there is a query whether the participants in the study have the same understanding of the questionnaire items. This aspect of the validity is called DIF (2).

Several statistical methods have been designed for DIF detection. Unidimensionality and local independence are two fundamental assumptions in most DIF detection methods including item response theory (IRT), Rasch model and ordinal logistic regression (OLR) approach (3, 4). Furthermore, two disadvantages of methods used in detecting DIF, except Latent Class Regression (LCR) and multiple indicators multiple causes (MIMIC) are, assumption that latent variable is continuous willy-nilly and inability to adjust categorical and continuous covariates concurrently (5).

MIMIC model in comparison with other DIF detection methods relaxes the assumptions of unidimensionality and conditional independence (6, 7). However, discrete latent variables are incorporated in MIMIC model complicate analysis (6).

Discrete latent variables give a more straightforward interpretation than continuous ones under certain conditions such as psychometric studies. LCR is a popular method for extracting the discrete latent variable from the discrete observed variable and facilitates estimation of the effects of covariates.
on predicting latent class membership (8).

This paper demonstrates an empirical example regarding Minor Psychiatric Disorders (MPD) among nurses for assessing DIF and proposes a new LCR based technique in comparison with MIMIC for assessing DIF in General Health Questionnaire (GHQ-12).

2. METHODS

Study design

A large population-based cross-sectional study was carried out among 771 nurses working in hospitals of the Fars and Bushehr provinces, Southern Iran, between October and December 2014. The data were fully described elsewhere (9, 10).

Statistical Analysis

The proposed method based on the LCR and MIMIC was applied to determine Uniform-DIF. Mplus version 7 software was used to analyze the data.

Detection of uniform-DIF in MIMIC

One of the methods for assessing DIF is the use of MIMIC models, an application of Structural Equation Modeling (SEM) (11). The MIMIC model has many advantages as compared with traditional models and does not need large sample sizes and scale variables that are not restricted to particular values; also, latent variable can be predicted by at least one observed indicator variable (12). The MIMIC model provides information on the structural and measurement models and is also based on matching with a latent variable, which may be more accurate than an observed score (12). In MIMIC model, establishing a common metric does not seem to be necessary (13). However, it only detects Uniform-DIF (6). The mechanism of MIMIC for detection of DIF is to regress a latent variable and its indicators on a covariate. The model in Figure 1 permits DIF for item 2 while assuming that all other items are DIF free.

The goodness-of-fit of MIMIC model was demonstrated using the Root Mean Square Error of Approximation (0.072), Bentler comparative fit index (0.949) and Tucker-Lewis Index (0.937). All goodness-of-fit indexes were in the acceptable range.

Detection of uniform-DIF in LCR

Huang and Bandeen-Roche (14) extended LCR, which includes direct effects of covariates on observed indicator variables. Our idea for detection of DIF in the GHQ-12 questionnaire is based on LCR proposed by Huang and Bandeen-Roche (14). If the relationship between covariates, such as gender, and observed indicator variable, such as items of GHQ-12 in Figure 1 was statistically significant, therefore there is a Uniform-DIF for that item of the questionnaire. The difference between LCR and MIMIC methods that are latent variables in LCR can only be discrete.

For LCR approach, Vuong-Lo-Mendell-Rubin likelihood ratio (VLMR) tests, entropy, Akaike Information Criterion (AIC) and Bayesian Information Criteria (BIC) were used to evaluate how well latent classes are distinguished (15, 16). In this study, information criteria (AIC=18335 and BIC=18916) and classification quality (entropy=0.892 and VLMR test’s p-value is less than 0.001) show that the best fitting model of the LCR was achieved with two classes, which also had a suitable interpretation.

3. RESULTS

Table 1 summarizes the demographic characteristics of the participants included in DIF analyses and descriptive findings. There were significant differences between the nurses’ GHQ-12 scores regarding the categorical age and gender.

The result of fitting LCR with 2 classes showed that 27.4% of nurses were classified as MPD. Gender was identified as an influential factor for the level of MPD. Table 2 shows covariates coefficients of the effective level of MPD. Table 3 shows the results of fitting LCR and MIMIC methods for detection of Uniform-DIF in GHQ-12 questionnaire.

DIF by gender

The final model identified in two items (items 2 and 9) had significant direct effects among these items and gender in the LCR and MIMIC simultaneously. However, items 1 and 3 only in MIMIC had Uniform-DIF. The agreement between LCR and MIMIC to detect Uniform-DIF across gender was 83.3%.

DIF by age

Both LCR and MIMIC methods showed that item 9 had Uniform-DIF of age. The DIF analysis did not reveal any difference between the LCR and MIMIC methods by age.

DIF by education

Direct effect between items 6 and education in MIMIC model was significant. The agreement between LCR and MIMIC to detect Uniform-DIF across education was 91.7%.

DIF by marital status

Item 8 by MIMIC technique and item 2 with our proposed method had Uniform-DIF of marital status. The agreement between LCR and MIMIC to detect Uniform-DIF across marital status was 83.3%.

Table 1. Participants’ demographic characteristics and descriptive findings.

<table>
<thead>
<tr>
<th>Gender</th>
<th>N (%)</th>
<th>Mean (SD) of GHQ-12</th>
<th>Statistic (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>269 (34.9)</td>
<td>3.30±3.29</td>
<td>2.492 (0.015*)</td>
</tr>
<tr>
<td>Female</td>
<td>502 (65.1)</td>
<td>2.70±2.99</td>
<td></td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
<td>0.977 (0.329)</td>
</tr>
<tr>
<td>Single</td>
<td>222 (28.8%)</td>
<td>3.27±3.27</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>540 (71.2%)</td>
<td>3.02±3.17</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>0.847 (0.429)</td>
</tr>
<tr>
<td>Associate Degree</td>
<td>48 (6.2%)</td>
<td>2.76±3.35</td>
<td></td>
</tr>
<tr>
<td>Bachelor degrees</td>
<td>702 (91.1%)</td>
<td>3.14±3.21</td>
<td></td>
</tr>
<tr>
<td>Master of Arts</td>
<td>21 (2.7%)</td>
<td>2.76±3.35</td>
<td></td>
</tr>
<tr>
<td>Age Categorical**</td>
<td></td>
<td></td>
<td>4.442 (0.012*)</td>
</tr>
<tr>
<td>&lt; 30</td>
<td>381 (50.1%)</td>
<td>3.43±3.38</td>
<td></td>
</tr>
<tr>
<td>30-40</td>
<td>310 (40.7%)</td>
<td>2.79±2.98</td>
<td></td>
</tr>
<tr>
<td>&gt; 40</td>
<td>70 (9.2%)</td>
<td>2.57±3.12</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Parameters estimated for factors affecting MPD by LCR. Reference category * Significant at 5% Reference category: nurse with MPD

Table 3. Covariates coefficients of the effective level of MPD.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Odds ratio</th>
<th>95%Confidence interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Ref = Women)</td>
<td>1.20</td>
<td>(1.01,1.42)</td>
<td>0.049*</td>
</tr>
<tr>
<td>Marital status (Ref = single)</td>
<td>1.05</td>
<td>(0.88,1.24)</td>
<td>0.680</td>
</tr>
<tr>
<td>Level of education (Ref = Associate Degree)</td>
<td>1.09</td>
<td>(0.65,1.83)</td>
<td>0.790</td>
</tr>
<tr>
<td>Age</td>
<td>1.04</td>
<td>(0.99,1.09)</td>
<td>0.150</td>
</tr>
</tbody>
</table>

*Reference category * Significant at 5% Reference category: nurse with MPD
Table 3. Detection of Uniform-DIF in LCR (p-value for 2 classes) and MIMIC. Plus indicates item with DIF and minus indicates item with free-DIF.

<table>
<thead>
<tr>
<th>ITEM</th>
<th>Concentration</th>
<th>Lost sleep</th>
<th>Play useful part</th>
<th>Making decision</th>
<th>Under strain</th>
<th>Overcome difficulties</th>
<th>Enjoy daily activities</th>
<th>Face up to problems</th>
<th>Feeling depressed</th>
<th>Losing confidence</th>
<th>Worthless person</th>
<th>Reasonably happy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LCR</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MIMIC</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Education</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Marital Status</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 1. MIMIC and LCR model for detecting Uniform-DIF for only item 2.

4. DISCUSSION

This study attempted to determine whether the GHQ-12 was invariant in Iranian nurses using LCR and MIMIC modeling. Although, the GHQ-12 is widely used in various settings to assess MPD, there is a dearth of research on examining it with DIF analysis (17, 18).

The results of the fitting LCR demonstrated that gender was a significant factor in determining the level of MPD. More details on the factors affecting MPD in southern Iranian nurses using LCR and robust regression were reported elsewhere (9, 10).

The results of MIMIC and LCR fitting indicate that the GHQ-12 is to a great degree, an invariant measure for the assessment of MPD among nurses. The measurement equivalence analyses with LCR revealed the presence of DIF for two items: Item 2 “lost sleep”, and item 9 “feeling depressed”. LCR and MIMIC are two model-based procedures, which have fewer assumptions than others, for detecting DIF. The agreement between the LCR and MIMIC for detection of DIF and DIF-free items was desirable. Overall, six items of GHQ-12 (items 1, 2, 3, 6, 8 and 9) in both methods were identified as DIF,2 of which (items 2 and 9) were common. Smith and his colleagues(17) using Rash model showed that items 1, 2, 3 and 5 of GHQ-12 had a DIF by gender, treatment aims and age group in the sample of cancer patients. Another study using Rash model found 3 items (4, 7 and 10) with DIF by gender in a sample of cancer patients (18). Different approaches for detecting DIF led to different results. So, an appropriate selection technique for assessing DIF according to the data structure is very important.

LCR is a useful tool to deal with discrete multivariate response data. LCR is a mixed model that divides a population into mutually exclusive and exhaustive subgroups. LCR can model the data in which indicator and latent variables are discrete and make covariates to be related to latent and indicator variables (14). Researchers are interested in using discrete latent variable and modeling it in medicine, psychology and social science (19, 20). Despite widespread use of the LCR in practice, there is no commonly accepted statistical index for deciding on the number of classes in a study population (21). On the other hand, this method requires a large sample size (22). Previous studies have shown that latent class analyses have better performance than latent trait model when data are in the Likert-type scales (23). To the best of the authors’ knowledge, in psychology studies, LCR has not been used to detect DIF.

MIMIC method is widely used in psychometric context to test for DIF between groups on the measurement of a latent variable. MIMIC, as compared to traditional methods (i.e. IRT- Likelihood Ratio DIF), has some favorable properties such as having generally small sample size requirements while simultaneously maintaining reliably low Type 1 error rates and sufficient DIF detection power (12). However, this model has disadvantages such as sensitivity to only one type of DIF (uniform DIF), inability to estimate and interpret c-parameter as a guessing parameter in the three-parameter model, inflated type 1 error rates found for shorter test and when ignoring data structures, effect-size estimates have not been derived and this model assumes that variances are equal across groups (24-26). Teresi (27) discussed the advantages and disadvantages of MIMIC method and compared them with other DIF detection techniques. Previous studies have shown that the MIMIC model outperformed Rash when there was no mean ability difference and the number of DIF items reached 30% (28). Also, MIMIC approaches to effect size estimation performed somewhat better than IRT at shorter test lengths (29). Correlation between the items and dimensions can be considered by MIMIC, while it does not have fundamental assumptions unidimensionality and conditional independence.

In general, we cannot say that the MIMIC model is better than the other models, but some statistical properties have made it one of the most popular methods used to detect DIF.
There are some limitations in the current study. Although two types of DIF, uniform and non-uniform, can be identified, in this present study, only Uniform-DIF was considered. It should be noted that for items with non-uniform DIF, specifically symmetric non-uniform DIF, the direction of DIF differs along the subscale, leading to bias sizes which will cancel out at the scale level and cannot influence the results. In this study, the effect of magnitude of DIF across covariates was not reviewed. The best practice for comparing the proposed method with competing methods in the evaluation of DIF is the iterative quantitative processes. In future studies with simulation, the statistical properties of this model will be described and compared with other models to detect DIF.

5. CONCLUSIONS

The results of this study have both practical and theoretical implications. From a practical perspective, the present study results demonstrated that, in general, the GHQ-12 can provide accurate estimates of GHQ-12 among Iranian nurses. From a theoretical viewpoint, advanced multivariate statistical methods enable us to adjudicate between continuous and categorical models of MPD using quantitative methods; yet, no prior studies have attempted to make such a comparison. According to the high agreement between MIMIC and LCR methods, the proposed method based on the LCR is highly recommended when latent variable is discrete and the sample size is adequate.

- Competing interests: The authors declare that they have no competing interests.
- Acknowledgments:

This work was a part of the Ph.D. thesis of Jamshid Jamali and was supported by grant number 94-10488 from Shiraz University of Medical Sciences, Shiraz, Iran. We would like to thank the nurses who participated in the study.

- Abbreviations:

MPD: Minor Psychiatric Disorders; GHQ: General Health Questionnaire; DIF: Differential Item Functioning; MIMIC: multiple indicators multiple causes; LCR: Latent Class Regression; IRT: Item Response Theory

REFERENCES

3. Zumbo BD. Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. Language assessment quarterly. 2007; 4(2): 223-33.