

Testing measurement equivalence of the SF-36 questionnaire across patients on hemodialysis and healthy people

Zahra Bagheri¹ · Peyman Jafari¹ · Marjan Faghih¹ · Elahe Allahyari¹ · Tania Dehesh¹

Received: 27 May 2015 / Accepted: 18 August 2015 / Published online: 2 September 2015
© Springer Science+Business Media Dordrecht 2015

Abstract

Background Differential item functioning (DIF) occurs when members from different groups respond differently to particular items in a health-related quality of life (HRQoL) questionnaire after controlling for underlying HRQoL construct. This study aimed to assess DIF in the SF-36 questionnaire and its effect on comparing HRQoL scores across patients on HD and healthy people.

Methods One hundred fifty patients on maintenance hemodialysis (HD) and 642 healthy individuals filled out the Persian version of the SF-36 questionnaire. Multiple-group multiple-indicator multiple-causes (MG-MIMIC) model was used to assess DIF across patients on HD and healthy population.

Results Sixteen out of 36 (44.4 %) items were flagged with DIF. Six out of 16 items (37.5 %) were flagged with uniform DIF, nine items (56.2 %) with non-uniform DIF, and one item (6.2 %) with both uniform DIF and non-uniform DIF. DIF items were associated with all subscales with the exception of the limitation due to physical problems and bodily pain subscales. The significant lower HRQoL scores of patients on HD in comparison with healthy people in the physical functioning and vitality subscales did not change after removing the items with uniform DIF.

Conclusions Our findings revealed that patients on HD and healthy people perceived the meaning of the items in SF-36 questionnaire differently. Although the impact of DIF is minimal, the cross-group comparison across patients

on HD and healthy people should be performed with caution.

Keywords Quality of life · Patients on hemodialysis · Healthy population · Differential item functioning · SF-36

Introduction

End-stage renal disease (ESRD) is a major public health problem all over the world [1], with its prevalence increasing by 8 % from 2007 to 2012 [2]. Although maintenance hemodialysis (HD) is an effective life-preserving treatment for patients with ESRD [3], it poses a substantial burden on patients and their families [4] and also adversely affects the patients' health-related quality of life (HRQoL) [5–8]. While previous studies have mainly focused on improving the patients' survival, optimizing their HRQoL is now increasingly investigated [4, 8]. The importance of measuring HRQoL in patients on HD is due to higher rates of mortality and hospitalization among patients with lower HRQoL scores [9–13]. In addition, measuring HRQoL provides valuable information regarding the effect of disease and its treatment on the physical, psychological and social well-being of the patients with chronic conditions [14].

Several biological and psychological factors such as malnutrition, the presence of concomitant diseases, sleep complaints, anemia, anxiety, depression, dependency on caregivers, uncertainty about future and difficulties in undertaking and maintaining their job contribute to impaired HRQoL in patients on HD [4, 8, 9, 15–19]. Accordingly, previous studies have shown that do patients on HD evaluate their HRQoL scores not only lower than healthy people [17, 20–25], but also lower than patients with renal transplant, breast cancer, colon cancer, and

✉ Peyman Jafari
jafarip@sums.ac.ir

¹ Department of Biostatistics, Shiraz University of Medical Sciences, Shiraz, Iran

leukemia [26]. However, these cross-group comparisons are meaningful and valid as long as a prerequisite assumption which is termed as measurement equivalence is established [27]. Measurement equivalence of an instrument, which is examined by differential item functioning (DIF) analysis, means that respondents from different groups interpret items in a HRQoL questionnaire similarly, given the same level of underlying HRQoL [27, 28]. If measurement equivalence is not satisfactory, it is not clear whether the observed disparity in HRQoL scores of the studied groups is a real difference in the underlying construct of interest or, it is due to different interpretations of items in the questionnaire [27, 29].

The SF-36 is one of the most frequently used questionnaires to measure HRQoL of patients on HD, with its reliability and validity being confirmed in several studies [30–35]. Although in the previous research the measurement equivalence of the SF-36 has been evaluated across people with different cultural and socio-demographic backgrounds and also with different chronic conditions [36–42], this issue has never been examined across patients on HD and healthy people. Hence, it is not clear whether the observed disparity in their HRQoL scores which was reported in the previous studies is a real difference or not. This study aimed to assess whether patients on HD and healthy people perceive the meaning of the items in SF-36 differently or not, using multiple-group multiple-indicator multiple-causes (MG-MIMIC) model.

Methods

Patients and measurements

A total of 150 patients on HD from the Sadra dialysis center and two affiliated hospitals of Shiraz University of Medical Sciences in Shiraz (southern part of Iran) were recruited from July 2011 to December 2011. The patients who received HD on a thrice per week dialysis regime, aged over 18 years, undergoing HD for at least 6 months were included in our study [43]. A simulation study showed that when the sample size in the focal group (generally referred to as patients group) is around 100, in order to have MIMIC results with adequate type I error rate and power, the sample size in the reference group (generally referred to as healthy) should be greater than 500 [44]. Therefore, we randomly selected 642 healthy individuals aged over 18 from a general healthy population and those who had any chronic diseases or health conditions were excluded from the study. Written informed consent was obtained from the participants prior to enrollment in the study. The study was approved by the ethical committee of our institution, Shiraz University of Medical Sciences. The Persian version of

the SF-36, which was previously translated and validated in Iran [45, 46], was filled out by both healthy individuals and patients on HD, after completing consent form.

The SF-36 is a well-know generic questionnaire that consists of eight subscales, including physical functioning (ten items), role limitations due to physical problems (four items), bodily pain (two items), general health perceptions (five items), vitality (four items), social functioning (two items), role limitations due to emotional problems (three items), and perceived mental health (five items). Each scale was directly transformed into a 0–100 scale based on the assumption of equal weight for each item, with zero indicating the worst HRQoL and 100 indicating the best possible score. Then the T-scores being a linear transformation of the 0–100 possible range and having a mean of 50 and standard deviation of 10 for every subscale were used for subsequent analyses.

Statistical analysis

In this study, MG-MIMIC model was used to assess DIF. In technical point of view, identifying DIF through MG-MIMIC model is the same as a multiple-group confirmatory factor analysis (MG-CFA) analysis with covariate [47]. When assessing DIF, two types of DIF can occur, namely uniform and non-uniform DIF. Uniform DIF means a constancy of difference in item response probabilities of two groups across the scale. When there is a different direction of DIF in different parts of the construct scale, non-uniform DIF occurs [48, 49]. Uniform DIF was detected when a significant difference in thresholds of a categorical item between the two groups was observed, and non-uniform DIF was identified when there was a significant difference in the loadings of an item across the two groups [50]. The process of detecting DIF was an iterative procedure that involved comparing a series of nested models. In the first step, the most constraint model with equal factor loadings, thresholds and residual variance of all items, variance of latent trait, and scaling factor for both groups fit as a baseline model. In this model, the only parameter which was freely estimated for both groups was the latent variables' mean. The information from modification indices associated with threshold parameters and factor loadings was utilized to detect items with uniform and non-uniform DIF. In the next step, a model that relaxed a constraint on an item's threshold (which had the largest magnitude of modification indices associated with threshold parameters) was estimated and compared to the baseline model. Then, a model that relaxed constraint on factor loading of an item (which has the largest magnitude of modification indices associated with factor loading) was fitted and compared to the baseline model. Uniform or non-uniform DIF was detected based on comparison of the reduction in the value

of Chi-square statistics for these two refinements. If relaxing of threshold parameters resulted in larger improvement and became significant, uniform DIF was detected, and if relaxing of factor loading led to larger improvement, non-uniform DIF was identified. The resultant model, which was considered as new baseline model, was fitted, then the resulting modification indices examined, and the above-mentioned steps were repeated until no significant model modifications were recognized.

In our study, age and gender were considered as the confounding variables since they differed significantly among patients and healthy individuals. We applied the mean- and variance-adjusted weighted least square (WLSMV) estimation procedure being proposed for ordinal indicators in the Mplus 6.1 software to fit MG-MIMIC models.

To compare the SF-36 subscale scores between healthy population and patients on HD, Student's *t* test was used, with calculating Cohen's *d* effect size. The magnitude of effect size <0.2, 0.2–0.49, 0.5–0.79, and >0.8 is defined as negligible, small, moderate, and large, respectively [51]. To assess the impact of DIF, the items with uniform DIF were eliminated from the corresponding subscale and the subscale score was computed again.

Results

Descriptive statistics

The mean (\pm SD) age of the healthy population was 41.51 ± 13.10 , and 248 out of 642 (38.6 %) healthy population were male. Table 1 describes the demographic characteristics of the patients on HD in our study. As shown, the patients on HD are significantly older than the healthy ones (p value < 0.001) and the number of males was significantly higher among patients compared with the healthy (p value = 0.048).

Table 1 Demographic characteristics of patients on HD

	<i>N</i> (%)
<i>Gender</i>	
Male	88 (59)
Female	62 (41)
<i>Education level</i>	
Illiterate	29 (19)
High school	40 (26)
University studies	81 (54)
	Mean \pm SD
Age	50.5 \pm 15.1
HD time (min)	216.5 \pm 237

DIF assessment

Table 2 shows the items' parameters, including factor loadings and thresholds resulting from MG-MIMIC model for items identified with DIF across patients on HD and healthy general population. As mentioned earlier, uniform DIF is detected by different threshold parameters among the groups and non-uniform DIF is identified with different factor loadings. In general, 16 out of 36 (44.4 %) items were identified with DIF. Six out of 16 DIF items (37.5 %) were flagged with uniform DIF, nine items (56.2 %) with non-uniform DIF, and one item (6.2 %) with both uniform DIF and non-uniform DIF. In the physical functioning subscale, there are four items (#3, #5, #9, and #12) with non-uniform DIF and two items (#7 and #11) with uniform DIF. One non-uniform DIF item (#19) was associated with the role limitation due to emotional problems subscale, and two items, one uniform (#23) and one non-uniform DIF (#29), were associated with the vitality subscale. In addition, in the perceived mental health subscale, one item (#28) was detected as uniform DIF, one (#30) as non-uniform DIF, and one (#26) as both uniform DIF and non-uniform DIF. In the social functioning subscale, one item (#20) with non-uniform DIF and in the general health subscale, two items (#1 and #35) with uniform DIF and one item (#33) with non-uniform DIF were identified by MG-MIMIC. No DIF items were detected in the role limitation due to physical problems and bodily pain subscales.

For the physical functioning subscale, the factor loading of items 3 (Vigorous activities, such as running, lifting...), 5 (Lifting or carrying groceries), and 12 (Bathing or dressing yourself) was greater for healthy individuals than for patients on HD, indicating that these three items have a greater ability to discriminate under and above a certain level of underlying physical functioning for healthy people as opposed to patients on HD. In contrast, the factor loading of item 9 (Walking more than a kilometer) was greater for patients than for healthy people, meaning that this item possesses a greater ability to discriminate under and above a certain level of underlying physical functioning for patients on HD than for healthy people. In addition in this subscale, the threshold parameters of items 7 (Climbing one flight of stairs) and 11 (Walking 100 m) were smaller for patients on HD than for healthy people, implying that given a higher level of underlying physical functioning, patients on HD probably reported fewer problems than healthy individuals. The factor loading of item 19 (Didn't do work or other activity as carefully as usual) in the role limitation due to emotional problems, items 26 (Felt calm and peaceful), and 30 (Been a happy person) in the perceived mental health and item 20 (Extent health problems interfered with normal social activities) in the social functioning subscales were greater for patients on

Table 2 Item parameters and standard errors for uniform and non-uniform DIF items in the SF-36 questionnaire across patients on HD and healthy general population

	Factor loading <i>b</i> (SE)	First threshold τ_1 (SE)	Second threshold τ_2 (SE)	Third threshold τ_3 (SE)	Fourth threshold τ_4 (SE)
Physical functioning					
q3: Vigorous activities, such as running, lifting heavy objects, strenuous sports					
Healthy	0.858 (0.075)	-2.055 (0.222)	-0.575 (0.21)	-	-
Patient	0.646 (0.149)	-2.055 (0.222)	-0.575 (0.21)	-	-
q5: Lifting or carrying groceries					
Healthy	1.724 (0.164)	-2.233 (0.353)	-0.716 (0.334)	-	-
Patient	0.945 (0.166)	-2.233 (0.353)	-0.716 (0.334)	-	-
q7: Climbing one flight of stairs					
Healthy	1.533 (0.12)	-2.488 (0.382)	-1.07 (0.374)	-	-
Patient	1.533 (0.12)	-5.018 (0.699)	-3.107 (0.65)	-	-
q9: Walking more than a kilometer					
Healthy	1.181 (0.094)	-2.529 (0.288)	-1.009 (0.275)	-	-
Patient	1.724 (0.346)	-2.529 (0.288)	-1.009 (0.275)	-	-
q11: Walking 100 m					
Healthy	2.531 (0.248)	-3.463 (0.665)	-2.132 (0.644)	-	-
Patient	2.531 (0.248)	-8.352 (1.267)	-5.933 (1.255)	-	-
q12: Bathing or dressing yourself					
Healthy	2.113 (0.235)	-2.798 (0.648)	-1.756 (0.463)	-	-
Patient	0.976 (0.172)	-2.798 (0.648)	-1.756 (0.463)	-	-
Role limitations due to emotional problems					
q19: Didn't do work or other activity as carefully as usual					
Healthy	1.161 (0.081)	-2.162 (0.264)	-0.821 (0.234)	0.483 (0.234)	1.545 (0.244)
Patient	2.105 (0.284)	-2.162 (0.264)	-0.821 (0.234)	0.483 (0.234)	1.545 (0.244)
Vitality					
q23: Feel full of life*					
Healthy	0.842 (0.057)	-2.637 (0.223)	-1.488 (0.211)	-0.254 (0.206)	1.199 (0.212)
Patient	0.842 (0.057)	-1.655 (0.656)	-0.434 (0.618)	0.859 (0.632)	2.585 (0.691)
29: Feel worn out*					
Healthy	1.299 (0.088)	-2.754 (0.283)	-1.182 (0.259)	0.148 (0.257)	1.174 (0.265)
Patient	0.932 (0.125)	-2.754 (0.283)	-1.182 (0.259)	0.148 (0.257)	1.174 (0.265)
Perceived mental health					
q26: Felt calm and peaceful*					
Healthy	0.855 (0.067)	-2.393 (0.23)	-1.234 (0.217)	-0.158 (0.214)	1.211 (0.22)
Patient	1.831 (0.239)	-2.356 (0.583)	-0.927 (0.592)	0.576 (0.578)	2.825 (0.692)
q28: Felt down*					
Healthy	1.572 (0.109)	-2.276 (0.315)	-1.031 (0.306)	0.355 (0.302)	1.757 (0.317)
Patient	1.572 (0.109)	-4.014 (0.756)	-2.144 (0.689)	-1.404 (0.683)	-0.532 (0.683)
q30: Been a happy person*					
Healthy	0.96 (0.068)	-3.083 (0.265)	-2.036 (0.237)	-0.593 (0.227)	1.065 (0.226)
Patient	1.682 (0.208)	-3.083 (0.265)	-2.036 (0.237)	-0.593 (0.227)	1.065 (0.226)
Social functioning					
q20: Extent health problems interfered with normal social activities					
Healthy	0.999 (0.074)	-2.334 (0.269)	-1.418 (0.239)	-0.572 (0.237)	0.679 (0.236)
Patient	1.75 (0.447)	-2.334 (0.269)	-1.418 (0.239)	-0.572 (0.237)	0.679 (0.236)
General health					
q1: Is your health: excellent, very good, good, fair, poor					
Healthy	1.024 (0.081)	-3.158 (0.26)	-1.851 (0.246)	-0.279 (0.231)	0.665 (0.23)

Table 2 continued

	Factor loading b (SE)	First threshold τ_1 (SE)	Second threshold τ_2 (SE)	Third threshold τ_3 (SE)	Fourth threshold τ_4 (SE)
Patient	1.024 (0.081)	-2.492 (0.545)	-1.087 (0.524)	0.507 (0.49)	0.985 (0.548)
q33: I seem to get sick a little easier than other people					
Healthy	0.783 (0.068)	-1.403 (0.196)	-0.709 (0.19)	-0.124 (0.188)	0.482 (0.19)
Patient	0.381 (0.111)	-1.403 (0.196)	-0.709 (0.19)	-0.124 (0.188)	0.482 (0.19)
q35: I expect my health to get worse					
Healthy	0.918 (0.075)	-1.893 (0.228)	-0.83 (0.217)	-0.106 (0.212)	0.432 (0.213)
Patient	0.918 (0.075)	-2.156 (0.662)	-1.871 (0.659)	-1.398 (0.656)	-1.07 (0.657)

b factor loading, τ_i threshold parameters, SE standard error

* The stated items have six categories, but in our sample no one chose the last category, and so these items had four-threshold parameter as same as five-category items

HD as compared to healthy population. However, a reverse pattern was observed for item 29 (Feel worn out) in the vitality subscale in which the factor loading was greater for healthy individuals.

Furthermore, the threshold parameters of items 23 (Feel full of life) in the vitality, item 26 (Felt calm and peaceful) in the perceived mental health, and item 1 (In general, is your health...?) in the perceived general health were smaller for healthy people than for patients on HD, meaning that given a higher level of underlying physical functioning, healthy individuals probably reported fewer problems than patients on HD. Conversely, this pattern was reverse for item 28 (Felt down) in the perceived mental health and item 35 (I expect my health to get worse) in the general health subscales.

Impact of DIF

Table 3 presents the HRQoL scores (mean \pm SD) of general healthy population and patients on HD in each subscale of SF-36 before and after removing the items with uniform DIF. Before removing DIF items, the HRQoL score of patients was significantly lower than that of the healthy individuals in all subscales except for the role limitation due to emotional problems which did not differ significantly in different groups. The values of effect size showed that the discrepancy was negligible for role limitation due to emotional problems and perceived mental health (0.02 and 0.15, respectively), small for role limitation due to physical problems, bodily pain, general health perception, vitality, and social functioning (effect size ranged from 0.23 to 0.39), and moderate for physical functioning with the effect size being 0.51.

After removing the items with uniform DIF, the HRQoL scores of the patients on HD were also significantly lower for the physical functioning and vitality subscales with the moderate (0.55) and negligible (0.18) effect sizes,

respectively. We did not remove the items with uniform DIF in the social functioning subscale because there are only two items in these subscales. In addition, in the general health and perceived mental health subscales the effect of one item with uniform DIF can be canceled out by another uniform DIF item in the opposite direction with respect to the coefficients of threshold parameters reported in Table 2, so these subscales' scores were not recomputed, as well.

Discussion

The results of this study provide a perspective for researchers to find out whether SF-36 is an invariant measure for cross-group comparison across patients on HD and healthy people, the issue which has never been investigated in the previous studies. Our results revealed that 16 out of 36 items were flagged with DIF, implying that patients on HD and healthy individual perceived the meaning of these items differently.

For uniform DIF items, the threshold parameters of the items were greater or smaller for patients on HD than healthy people, suggesting that given a certain level of underlying HRQoL, patients on HD reported greater or fewer problems than healthy people. For non-uniform DIF items, factor loadings of the items were greater or smaller for healthy people compared with patients on HD, implying that these items had more or less discrimination ability under and above a certain level of underlying HRQoL for healthy people than for patients on HD.

Most of DIF items (six out of 15) were associated with the physical functioning subscale. This different perception of the items by patients on HD as oppose to healthy population may partly be explained by fatigue, reduced energy, limitation in the use of arm or leg, and also restriction on their daily activities which are frequent among patients

Table 3 Mean comparison of the HRQoL scores in the SF-36 questionnaire's subscale between patients on HD and healthy general population before and after DIF calibration

	Healthy Mean \pm SD*	Patients Mean \pm SD*	<i>t</i> value [†] (<i>p</i> [‡]), <i>d</i> ^c
<i>Scale</i>			
Physical functioning	47.20 \pm 9.96	34.86 \pm 10.88	13.41 (<0.001) 0.51
Role limitations due to physical problems	42.68 \pm 9.17	37.85 \pm 9.43	5.71 (<0.001) 0.25
Bodily pain	38.86 \pm 12.24	31.58 \pm 9.48	6.82 (<0.001) 0.31
General health perceptions	46.34 \pm 8.44	39.88 \pm 6.47	8.78 (<0.001) 0.39
Vitality	48.72 \pm 6.79	45.18 \pm 6.41	5.80 (<0.001) 0.26
Social functioning	33.43 \pm 5.64	30.02 \pm 8.26	6.04 (<0.001) 0.23
Role limitations due to emotional problems	39.80 \pm 11.34	39.20 \pm 13.62	0.56 (0.57) 0.02
Perceived mental health	38.80 \pm 7.44	36.48 \pm 7.47	3.43 (<0.001) 0.15
<i>Total score corrected for DIF</i>			
Physical functioning ^a	46.52 \pm 10.14	32.62 \pm 10.74	14.49 (<0.001) 0.55
Vitality ^b	50.06 \pm 9.00	46.90 \pm 7.69	3.97 (0.0001) 0.18

* SD

† Independent *t* test‡ *p* value^a Total score corrected for items 7 and 11 with uniform DIF^b Total score corrected for item 23 with uniform DIF^c Effect size

on HD [8, 52, 53]. Furthermore, in the vitality subscale patients on HD and healthy individuals responded differently to two out of four items. This discrepancy could be attributed to fatigue and exhaustion which prevail in patients on HD. Interestingly, a previous research showed that vitality grows in proportion to the physical health more than to the mental health [54]. In both the perceived mental health and general health subscales, three out five items were identified with DIF. A possible explanation for this could be that patients on HD usually suffer from psychological problems; consequently, a high prevalence of depression and anxiety among patients on HD was reported in previous research [16, 17, 55]. Hence, these problems can adversely affect their perception of HRQoL concept.

In contrast, surprisingly, patients on HD and healthy population perceived the meaning of the items in the role limitations due to physical problems and bodily pain subscales similarly. From a cognitive and psychological view point, this may be due to the fact that patients on HD may change their personal standards or their expectations when assessing their daily function. Moreover, the adverse effect of their chronic disease may become blunted because of psychological adaptation [8, 50, 52]. With regard to bodily pain, it is worth mentioning that patients on HD can manage their pain and have higher pain threshold [52]. Since this is the first study appraising DIF across patients on HD and healthy people, there were no comparable studies in this context. However, our findings are in accordance with those of previous studies, indicating that health status of people can affect the way they respond to the items of

HRQoL questionnaires [36–42]. For instance, in a previous research assessing DIF of the SF-36 across healthy adults with and without functional limitations (e.g., deafness and blindness), almost all of the items in the physical functioning subscale exhibited DIF [37]. Dallmeijer et al. [38] also reported that certain items of SF-36 showed DIF across patients with stroke, multiple sclerosis and amyotrophic lateral sclerosis [38]. Moreover, in another study which examined DIF in the SF-36 across patients with different chronic conditions (namely, hypertension, rheumatic conditions, diabetes, respiratory diseases, and depression), a few items with DIF were distinguished, even though the effect of DIF on cross-group comparison was minimal [42].

It should be noted that detecting DIF is one issue and its effect on the subscale scores is another. In the present study, we used a removing and retaining strategy to assess whether mean group differences in subscales scores alter statistically, with and without considering DIF items. Since removing a large number of items with DIF could affect the content validity of the measure, we only removed those with uniform DIF [56–58]. Our findings revealed that removing or retaining items with DIF from the physical functioning and vitality subscales did not change the results principally. For the other subscales, DIF cancellation occurred at the subscale level, so there was no indication to perform removing and retaining strategy. However, a recent study reported that full DIF cancellation rarely arises in practice [59]. Violation of the measurement equivalence assumption in the present study, which indicates that the

SF-36 functioned differently across the healthy people and the patients on HD, could be a threat to generalizability of the measure to heterogeneous groups [60]. Hence, considerable caution should be warranted when using the SF-36 to compare HRQoL scores across the two groups. However, it should be mentioned that the SF-36 is a well-known questionnaire with sound psychometric properties in different samples and languages [30–35, 45, 46]; consequently, detecting DIF does not necessarily undermine the validity of the SF-36 when we intend to measure HRQoL in a specific group [60].

An important strength of our study is that the impact of age and gender as confounding variables on DIF analysis was controlled by MG-MIMIC method, which is the distinct advantage of MG-MIMIC over other methods, such as item response theory (IRT) or MG-CFA analysis [61]. The importance of this issue is owing to the fact that without controlling the effect of confounding variables, DIF detection procedure and also subsequent group comparison regarding HRQoL score may distort. Although ordinal logistic regression (OLR) and MIMIC models also possess this distinct advantage, each has its own limitations. In MIMIC model, only uniform DIF can be detected [61], and in OLR, the observed score, not the latent one, is used in DIF detection procedure [62].

In addition, another advantage of MG-MIMIC model is that the purification of anchor items is considered in DIF detection process. Purification is an iterative process of removing items currently identified as DIF with the aim of obtaining a purified set of items, unaffected by DIF. This issue is important since overlooking purification results in over- or under-estimation in the number of DIF items [47, 50]. In MG-MIMIC approach, parameters detecting and controlling for DIF are estimated in an iterative process, so the underlying HRQoL scores are purified effectively in this study [47].

Our study had some limitations that merit attention when interpreting our findings. First, the effects of other demographic variables such as income and education on the DIF analysis were not taken into account in this study owing to lack of information in this regard. Previous research showed that higher levels of education and income can lead to improvement in HRQoL score of patients on HD [52, 63]. Furthermore, it is generally accepted that anxiety and depression are the most common psychological disorders that patients on HD suffer from [17, 18, 53, 63–65]; this may influence their perception of HRQoL questionnaires' items. Therefore, further in-depth studies should be undertaken to evaluate the effects of these factors on the DIF analysis of SF-36 questionnaire in these two groups.

In conclusion, our findings revealed that SF-36 is not an invariant measure across patients on HD and healthy people; hence, caution should be warranted for comparing

HRQoL scores between healthy people and the patients on HD. As researchers have recently aimed for measures that function similarly across heterogeneous populations, we recommended revising the items that exhibit DIF, so the questionnaire can be used across different groups. Yet, identifying the underlying causes of DIF using qualitative methodology, such as cognitive interviewing techniques, can provide valuable information regarding how the DIF items are interpreted differently by patients on HD and healthy people. Finally, detecting DIF may vary from measure to measure and from one sample to another. Hence, future studies should replicate our study of SF-36 across people with other health conditions and also across different cultures or extend it to other generic HRQoL instruments such as GHQ and WHOQOL-BREF.

Acknowledgments The authors are grateful to Dr Nasrin Shokrpour, Mr Hosain Argasi, and Mr Amin Borhani for English editing of this paper.

Compliance with ethical standards This work was supported by the Grant Number 89-5299 from Shiraz University of Medical Sciences Research Council. This article was extracted from a scientific project conducted by a PhD student, Marjan Faghieh. Written informed consent was obtained from the participants prior to enrollment in the study. The study was approved by the ethical committee of our institution, Shiraz University of Medical Sciences, and it was in accordance with the 1964 Helsinki declaration.

Conflict of interest The authors declare that they have no conflict of interests.

References

1. Khakurel S, Agrawal RK, Hada R (2009) Pattern of end stage renal disease in a tertiary care center. *J Nepal Med Assoc* 48:126–130
2. Hinkle JL, Cheever KH (2014) The 13th edition of Brunner and Suddarth's textbook of medical-surgical nursing. Lippincott, Williams and Wilkins, Philadelphia
3. Hinoshita F, Ando R, Sakai R, Kuriyama S (2014) Hemodialysis-associated problems to solve: current and future. *Sci World J* 2014:382170
4. Wasserfallen JB, Halabi G, Saudan P (2004) Quality of life on chronic dialysis: comparison between haemodialysis and peritoneal dialysis. *Nephrol Dial Transpl* 19:1594–1599
5. Kusleikaite NBI, Kuzminskis V, Vaiciuniene R (2010) The association between health-related quality of life and mortality among hemodialysis patients. *Medicina (Kaunas)* 46:531–537
6. Noshad H, Sadreddini S, Nezami N, Salekzamani Y, Ardalan MR (2009) Comparison of outcome and quality of life: haemodialysis versus peritoneal dialysis patients. *Singap Med J* 50:185–192
7. Pakpour AH, Saffari M, Yekaninejad MS, Panahi D, Harrison AP, Molsted S (2010) Health-related quality of life in a sample of Iranian patients on hemodialysis. *Iran J Kidney Dis* 4:50–59
8. Mittal SK, Ahern L, Flaster E, Maesaka JK, Fishbane S (2001) Self-assessed physical and mental function of haemodialysis patients. *Nephrol Dial Transpl* 16:1387–1394
9. Hedayati SS, Bosworth HB, Briley LP, Sloane RJ, Pieper CF, Kimmel PL, Szczech LA (2008) Death or hospitalization of

- patients on chronic hemodialysis is associated with a physician-based diagnosis of depression. *Kidney Int* 74:930–936
10. Kimmel PL, Patel SS (2006) Quality of life in patients with chronic kidney disease: focus on end-stage renal disease treated with hemodialysis. *Semin Nephrol* 26:68–79
 11. Lopes AA, Bragg-Gresham JL, Goodkin DA, Fukuhara S, Mapes DL, Young EW et al (2007) Factors associated with health-related quality of life among hemodialysis patients in the DOPPS. *Qual Life Res* 16:545–557
 12. Mapes DL, Lopes AA, Satayathum S, McCullough KP, Goodkin DA, Locatelli F et al (2003) Health-related quality of life as a predictor of mortality and hospitalization: the dialysis outcomes and practice patterns study (DOPPS). *Kidney Int* 64:339–349
 13. Marquis P, Trudeau E (2001) Quality of life and patient satisfaction: two important aspects in asthma therapy. *Curr Opin Pulm Med* 7:S18–S20
 14. Paniagua R, Amato D, Vonesh E, Guo A, Mujais S, Mexican Nephrology Collaborative Study Group (2005) Health-related quality of life predicts outcomes but is not affected by peritoneal clearance: the ADEMEX trial. *Kidney Int* 67:1093–1104
 15. Mujais SK, Story K, Brouillette J, Takano T, Soroka S, Franek C et al (2009) Health-related quality of life in CKD Patients: correlates and evolution over time. *Clin J Am Soc Nephrol* 4:1293–1301
 16. Tsay SL, Healstead M (2002) Self-care self-efficacy, depression, and quality of life among patients receiving hemodialysis in Taiwan. *Int J Nurs Stud* 39:245–251
 17. Finkelstein FO, Finkelstein SH (2000) Depression in chronic dialysis patients: assessment and treatment. *Nephrol Dial Transpl* 15:1911–1913
 18. Rayyani M, Malekyan L, Azzizadeh Forouzi M, Razban F (2014) Self-care self-efficacy and quality of life among patients receiving hemodialysis in South-East of Iran. *Asian J Nurs Educ Res* 4:165–171
 19. Zimmermann PR, Poli de Figueiredo CE, Fonseca NA (2001) Depression, anxiety and adjustment in renal replacement therapy: a quality of life assessment. *Clin Nephrol* 56:387–390
 20. Blake C, Codd MB, Cassidy A, O'Meara YM (2000) Physical function, employment and quality of life in end-stage renal disease. *J Nephrol* 13:142–149
 21. Johansen KL, Shubert T, Doyle J, Soher B, Sakkas GK, Kent-Braun JA (2003) Muscle atrophy in patients receiving hemodialysis: effects on muscle strength, muscle quality, and physical function. *Kidney Int* 63:291–297
 22. Kimmel PL, Peterson RA, Weihs KL, Simmens SJ, Alleyne S, Cruz I, Veis JH (1998) Psychosocial factors, behavioral compliance and survival in urban hemodialysis patients. *Kidney Int* 54:245–254
 23. Painter P, Carlson L, Carey S, Paul SM, Myll J (2000) Physical functioning and health-related quality-of-life changes with exercise training in hemodialysis patients. *Am J Kidney Dis* 35:482–492
 24. Painter P, Carlson L, Carey S, Paul SM, Myll J (2000) Low-functioning hemodialysis patients improve with exercise training. *Am J Kidney Dis* 36:600–608
 25. Tawney KW, Tawney PJ, Kovach J (2003) Disablement and rehabilitation in end-stage renal disease. *Semin Dial* 16:447–452
 26. Chen M, Ku N (1998) Factors associated with quality of life among patients on hemodialysis. *Nurs Res* 6:393–404
 27. Teresi JA, Fleishman JA (2007) Differential item functioning and health assessment. *Qual Life Res* 16:33–42
 28. Vandenberg RJ, Lance CE (2000) A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 3:4–70
 29. Kim ES, Yoon M (2011) Testing measurement invariance: a comparison of multiple-group categorical CFA and IRT. *Struct Equ Model* 18:212–223
 30. Curtin RB, Lowrie EG, DeOreo PB (1999) Self-reported functional status: an important predictor of health outcomes among end-stage renal disease patients. *Adv Ren Replace Ther* 6:133–140
 31. Diaz-Buxo JA, Lowrie EG, Lew NL, Zhang H, Lazarus JM (2000) Quality-of-life evaluation using Short Form 36: comparison in hemodialysis and peritoneal dialysis patients. *Am J Kidney Dis* 35:293–300
 32. Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT (1993) The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 306:1437–1440
 33. Kalantar-Zadeh K, Kopple JD, Block G, Humphreys MH (2000) Association among SF-36 quality of life measures and nutrition, hospitalisation, and mortality in hemodialysis. *J Am Soc Nephrol* 12:2797–2806
 34. McHorney CA, Ware JE Jr, Raczek AE (1993) The MOS 36-item Short Form Health Survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 31:247–263
 35. Ware JE, Sherbourne CD (1992) The MOS 36-item Short Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30:473–483
 36. Pollard B, Johnston M, Dixon D (2013) Exploring differential item functioning in the SF-36 by demographic, clinical, psychological and social factors in an osteoarthritis population. *BMC Musculoskelet Disord* 14:346
 37. Horner-Johnson W, Krahn GL, Suzuki R, Peterson JJ, Roid G, Hall T, RRTC Expert Panel on Health Measurement (2010) Differential performance of SF-36 items in healthy adults with and without functional limitations. *Arch Phys Med Rehabil* 91:570–575
 38. Dallmeijer AJ, de Groot V, Roorda LD, Schepers VP, Lindeman E, van den Berg LH et al (2007) 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. *J Rehabil Med* 39:163–169
 39. Taylor WJ, McPherson KM (2007) 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 Rheum* 57:723–729
 40. Bjorner JB, Kreiner S, Ware JE, Damsgaard MT, Bech P (1998) Differential item functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 51:1189–1202
 41. Perkins AJ, Stump TE, Monahan PO, McHorney CA (2006) Assessment of differential item functioning for demographic comparisons in the MOS SF-36 health survey. *Qual Life Res* 15:331–348
 42. Yu YF, Yu AP, Ahn J (2007) Investigating differential item functioning by chronic diseases in the SF-36 health survey: a latent trait analysis using MIMIC models. *Med Care* 45:851–859
 43. Dehesh T, Zare N, Jafari P, Sagheb MM (2014) Psychometric assessment of the Persian version of the Ferrans and Powers 3.0 index in hemodialysis patients. *Int Urol Nephrol* 46:1183–1189
 44. Woods CM (2009) Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivar Behav Res* 44:1–27
 45. Jafari H, Lahsaeizadeh S, Jafari P, Karimi M (2008) Quality of life in thalassemia major: reliability and validity of the Persian version of the SF-36 questionnaire. *J Postgrad Med* 54:273–275
 46. Montazeri A, Goshtasebi A, Vahdaninia M, Gandek B (2005) The Short Form Health Survey (SF-36): translation and validation study of the Iranian version. *Qual Life Res* 14:875–882

47. Jones RN (2006) Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination: detecting differential item functioning using MIMIC modeling. *Med Care* 44:124–133
48. Traebert J, Foster Page LA, Thomson WM, Locker D (2010) Differential item functioning related to ethnicity in an oral health-related quality of life measure. *Int J Paediatr Dent* 20:435–441
49. Traebert J, de Lacerda JT, Thomson WM, Page LF, Locker D (2010) Differential item functioning in a Brazilian–Portuguese version of the Child Perceptions Questionnaire (CPQ). *Community Dent Oral Epidemiol* 38:129–135
50. Huang IC, Leite WL, Shearer P, Seid M, Revicki DA, Shenkman EA (2011) Differential item functioning in quality of life measure between children with and without special health-care needs. *Value Health* 14:872–883
51. Cohen J (1988) *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, New Jersey
52. Abdelbasit A, Katherine N, PamelaW Kamal HH (2014) Comparing health-related quality of life between haemodialysis patients and a community sample in the United Arab Emirates. *Ren Soc Australas J* 10:34–43
53. Sathvik BS, Parthasarathi G, Narahari MG, Gurudev KC (2008) An assessment of the quality of life in hemodialysis patients using the WHOQOL-BREF questionnaire. *Indian J Nephrol* 18:141–149
54. Cotârlă L (2009) Possibilities of increasing the quality of life of renal patients, chronically dialyzed. Dissertation. Lucian Blaga University, Sibiu, Romania
55. Parker KP, Kutner NG, Bliwise DL, Bailey JL, Rye DB (2003) Nocturnal sleep, daytime sleepiness, and quality of life in stable patients on hemodialysis. *Health Qual Life Outcomes* 1:68
56. Crane PK, Hart DL, Gibbons LE, Cook KF (2006) A 37-item shoulder functional status item pool had negligible differential item functioning. *J Clin Epidemiol* 59:478–484
57. Jafari P, Stevanovic D, Bagheri Z (2015) Cross-cultural measurement equivalence of the KINDL questionnaire for quality of life assessment in children and adolescents. *Child Psychiatry Hum Dev*. doi:10.1007/s10578-015-0568-5
58. Jafari P, Allahyari E, Salarzadeh M, Bagheri Z (2015) Item-level informant discrepancies across obese-overweight children and their parents on the PedsQL™ 4.0 instrument: an iterative hybrid ordinal logistic regression. *Qual Life Res*. doi:10.1007/s11136-015-1046-z
59. Wyse AE (2013) DIF cancellation in the Rasch Model. *J Appl Meas* 14:118–128
60. Dimitrov DM (2008) *Quantitative research in education. Intermediate and advanced methods*. Whittier Publications, New York
61. Teresi JA (2006) Different approaches to differential item functioning in health applications: advantages, disadvantages and some neglected topics. *Med Care* 44:152–170
62. Swaminathan H, Rogers HJ (1990) Detecting differential item functioning using logistic regression procedures. *J Educ Meas* 27:361–370
63. Fukuhara S, Lopes AA, Bragg-Gresham JL, Kurokawa K, Mapes DL, Akizawa T (2003) Health-related quality of life among dialysis patients on three continents: the dialysis outcomes and practice patterns study. *Kidney Int* 64:1903–1910
64. Acaray A, Pinar R (2005) Quality of life in Turkish haemodialysis patients. *Int Urol Nephrol* 37:595–602
65. Borowiak E, Braksator E, Nowicki M, Kostka T (2009) Quality of life of chronic hemodialysis and peritoneal dialysis patients. *Med Sci Tech* 50:RA37–RA42