CrossMark

# Negligible impact of differential item functioning between Black and White dialysis patients on the Kidney Disease Quality of Life 36-item short form survey (KDQOL^TM-36)

John D. Peipert[1,2,3] · Peter Bentler[4] · Kristi Klicko[5] · Ron D. Hays[6]

## Abstract

**Purpose** Black dialysis patients report better health-related quality of life (HRQOL) than White patients, which may be explained if Black and White patients respond systematically differently to HRQOL survey items.

**Methods** We examined differential item functioning (DIF) of the Kidney Disease Quality of Life 36-item (KDQOL^TM-36) Burden of Kidney Disease, Symptoms and Problems with Kidney Disease, and Effects of Kidney Disease scales between Black ($n = 18,404$) and White ($n = 21,439$) dialysis patients. We fit multiple group confirmatory factor analysis models with increasing invariance: a Configural model (invariant factor structure), a Metric model (invariant factor loadings), and a Scalar model (invariant intercepts). Criteria for invariance included non-significant $\chi^2$ tests, $> 0.002$ difference in the models' CFI, and $> 0.015$ difference in RMSEA and SRMR. Next, starting with a fully invariant model, we freed loadings and intercepts item-by-item to determine if DIF impacted estimated KDQOL^TM-36 scale means.

**Results** $\Delta$CFI was 0.006 between the metric and scalar models but was reduced to 0.001 when we freed intercepts for the burdens and symptoms and problems of kidney disease scales. In comparison to standardized means of 0 in the White group, those for the Black group on the Burdens, Symptoms and Problems, and Effects of Kidney Disease scales were 0.218, 0.061, and 0.161, respectively. When loadings and thresholds were released sequentially, differences in means between models ranged between 0.001 and 0.048.

**Conclusion** Despite some DIF, impacts on KDQOL^TM-36 responses appear to be minimal. We conclude that the KDQOL^TM-36 is appropriate to make substantive comparisons of HRQOL between Black and White dialysis patients.

**Keywords** Health-related quality of life · KDQOL-36 · Measurement invariance · Differential item functioning

✉ John D. Peipert
john.peipert@northwestern.edu

1 Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 625 N. Michigan Ave., Suite 2700, Chicago, IL 60611, USA

2 Division of Nephrology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

3 Terasaki Research Institute, University of California, Los Angeles, Los Angeles, CA, USA

4 Departments of Psychology and Statistics, University of California, Los Angeles, Los Angeles, CA, USA

5 Medical Education Institute, Inc., Madison, WI, USA

6 Division of General Internal Medicine and Health Services Research, University of California, Los Angeles, Los Angeles, CA, USA

## Abbreviations

| | |
|---|---|
| AV | Arteriovenous |
| CFA | Confirmatory factor analysis |
| CFI | Comparative fit index |
| CMS | Centers for Medicare and Medicaid Services |
| DIF | Differential item functioning |
| DOPPS | Dialysis outcomes and practice patterns study |
| ESRD | End-stage renal disease |
| HRQOL | Health-related quality of life |
| KDCS | Kidney Disease Component Summary |
| KDQOL-36 | Kidney Disease Quality of Life 36-item survey |
| KDQOL-SF | KDQOL-short form |
| MCS | Mental Component Summary |
| PCS | Physical Component Summary |
| PD | Peritoneal dialysis |

RMSEA      Root mean squared error of approximation
SRMR      Standardized root mean square residual
US      United States
WLSMV      Weighted least squares with mean and variance adjustment

## Introduction

There are over 670,000 individuals with end-stage renal disease (ESRD), or kidney failure, in the United States (U.S.) [1]. ESRD patients must have medical therapy to replace their kidney function to remain living. Over 70% of these patients are on dialysis, where a machine filters toxins from the blood [1]. There is a large disparity in the likelihood of developing ESRD between Black and White individuals. Blacks are 3.9 times more likely to develop ESRD than Whites [1]. Potentially driving this disparity are differences between Black and White patients in prevalence of underlying causes of ESRD, like diabetes and hypertension [1–3], and this problem is growing. The incidence of ESRD due to diabetes increased by 33% between 2000 and 2012 for Blacks, but only 2% for Whites during the same period [1].

Though lifesaving, dialysis has been associated with poor health-related quality of life (HRQOL). For example, in a study comparing SF-36 Physical Component Summary (PCS) scores across multiple chronic and infectious conditions including AIDS, epilepsy, diabetes, and others, ESRD patients on dialysis had lower (worse) scores than patients with any other condition except multiple sclerosis [4]. For this reason, significant efforts are made to improve HRQOL in dialysis patients. In their Conditions for Coverage (42 CFR § 494.90), the Centers for Medicare and Medicaid Services (CMS) mandated that each U.S. dialysis patient's physical and mental health be monitored, and this often occurs with the use of a standardized HRQOL measure [5].

The most common instrument for implementing this requirement is the Kidney Disease Quality of Life 36-item survey (KDQOL-36™) [6]. The KDQOL-36 is a subset of items and scales from a larger, parent instrument, the KDQOL-SF [7]. The KDQOL-SF features 12 scales covering multiple dimensions of HRQOL, including generic HRQOL (SF-36), symptoms associated with kidney disease, effects of kidney disease, ability to work, cognitive function, quality of social interactions, sexual function, sleep quality, social support, encouragement from dialysis staff, and satisfaction with care. The KDQOL-36 short form survey reduces the SF-36 to the SF-12 and retains the Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease scales [6].

Over the past few decades, evidence has emerged suggesting that Black dialysis patients may have better HRQOL than Whites [8, 9]. For example, in the Dialysis Outcomes and Practice Patterns Study (DOPPS), a nationally representative surveillance study of dialysis patients, Black patients had significantly better HRQOL than Whites on SF-36 PCS, Mental Component Summary (MCS), and the Kidney Disease Component Summary (KDCS), a composite of the KDQOL-SF scales [10]. If accurate, these findings could have considerable clinical importance, leading to potentially different treatment decisions for Black and White patients. The absolute low levels of HRQOL on dialysis is often the justification for seeking alternative treatments associated with superior outcomes, like kidney transplantation, or to adjust the care plan of the patient while remaining on dialysis.

Before making clinical decisions based on evidence of differences in HRQOL between Black and White patients, the measurement invariance of HRQOL measures between Blacks and Whites needs to be established. Measurement invariance refers to whether the items within a measure tap the same construct equally well, and ultimately have the same meaning for different subgroups of patients. Violation of measurement invariance occurs when some items are not performing equally across groups, i.e., exhibit differential item functioning (DIF). In order to examine whether reported HRQOL differences between Black and White dialysis patients are genuine, this paper evaluates whether there is measurement invariance between Black and White respondents for the KDQOL-36 scales.

## Methods

### Patient sample

The data used in this study were collected between 6/1/2015 and 5/31/2016 as part of the Medical Education Institute's KDQOL-Complete program, a program that supports dialysis providers in assessing their patients with the KDQOL-36 to meet the aforementioned CMS requirement. During this period, 77,072 KDQOL-36 assessments were entered. After omitting duplicate records of patients, excluding patient records with missing assessment dates and incomplete KDQOL-36 assessments, excluding records for patients who were not adults, and were not on dialysis, 70,786 remained. The details of this patient selection procedure are described elsewhere [6]. Of these, we retained 39,843 patients for analysis: Black = 18,404 and White = 21,439. The UCLA Human Subjects Protection Committee (UCLA IRB #17–000313) granted an IRB exemption.

### Measures

The KDQOL-36 items are categorical and are combined to create multi-item scales. A generic scale, the SF-12, is

embedded within the KDQOL-36 and its Physical and Mental Health Component Summary measures (PCS and MCS) are scored on a T-score (mean = 50, SD = 10, in U.S. general population). It also includes 3 kidney-targeted scales: Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease. Each of these is scored by linearly transforming all items to a 0-100 possible range and averaging the items. KDQOL-36 items are all scored so that higher scores indicate better health.

In addition to the KDQOL-36, we obtained demographic and clinical data from CMS Form-2728 or from the dialysis centers. These variables included race (Black vs. White), age, whether etiology of ESRD was diabetes, dialysis type [in-center hemodialysis, peritoneal dialysis (PD), conventional home hemodialysis], dialysis access site [arteriovenous (AV) fistula, AV graft, venous catheter, PD catheter], and employment status.

## Tests of measurement invariance

Tests of measurement invariance for the KDQOL-36 Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease scales were based on correlated factors confirmatory factor analysis (CFA) models (KDQOL-36 Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease modeled as 3 correlated latent factors). In the multiple CFA models described below, several model parameters are constrained or freed, including factor loadings (correlation between items and latent factors), item category thresholds (standardized score determining which category each individual is in based on item response), item intercepts (when estimated, expected value of each item when the latent factor is equal to 0), and residual variances (variance of observed item responses not explained by the latent factor).

Tests of sequential, nested, multi-group CFA models for categorical items were conducted using the Mplus software Version 8 [11]. Based on recent recommendations for testing measurement invariance with categorical data [12], we examined results from multiple model estimators, including robust ML estimation (MLR) and weighted least squares with mean and variance adjusted estimation (WLSMV). For MLR estimation, we used standard approaches to factorial invariance testing. In the results, these models are referred to as MLR models. First, we estimated Configural models wherein all model parameters were freely estimated across groups. This step shows whether the factor structure is the same for each measure across each group. Second, if Configural invariance was found, then Metric invariance was examined by constraining the factor loadings to be equal across groups. Third, if Metric invariance was found, then Scalar invariance was examined by constraining intercepts (equal) as well as factor loadings across the groups. This step

in the procedure determines if patients from the different groups provide similar ratings on each item or not.

The approach to model identification and Metric setting varies among each of these models. First, the Configural models sets 1 loading per factor to 1 and sets factor means to 0. The Metric model is similar, except all loadings are constrained to be equal across groups. For the Scalar model, where all factor loadings and intercepts are constrained across groups, factor means are estimated in 1 group (Black). If measurement invariance was indicated for any models, the relevant item parameters were sequentially freed until no further invariance was found.

For the MLR models, model fit was examined with several fit indices. Comparative fit index (CFI) values of 0.95 or above [13], root mean square error of approximation (RMSEA) value of 0.06 or less [13, 14], and standardized root mean square residual (SRMR) of 0.08 or less [13]. Due to the large sample size used in these analyses, we expect $\chi^2$ tests of model fit to be significant even in the presence of good model fit. In each of these series of nested models, the corrected $\chi^2$ difference test, change in CFI ($\Delta$CFI), change in RMSEA ($\Delta$RMSEA), and change in SRMR ($\Delta$SRMR) were examined. Each change in goodness of fit index test has its own cut-off criterion suggesting non-invariance. Meade et al's criteria were used for $\Delta$CFI ($\geq -0.002$), which is calculated based on the number of factors and items in the model [15]. We used Chen's criteria for $\Delta$RMSEA ($\geq 0.015$ for Metric and Scalar invariance) and $\Delta$SRMR ($\geq 0.015$ for Metric invariance and $\geq 0.010$ for Scalar invariance) [16]. Chen's criteria for $\Delta$RMSEA and $\Delta$SRMR should be interpreted when there is also a $\Delta$CFI $\geq -0.01$. In examining differences in standardized factor loadings across groups, the standards set by Yoon and Millsap were adopted: $\geq 0.1 =$ small; $\geq 0.2 =$ medium; $\geq 0.3 =$ large [17].

In addition to these approaches, we employed Wu and Estabrook's method for categorical items with ordered responses [18]. This method holds that the standard measurement invariance approach of adding increasingly strict invariance constraints to the Configural model is not appropriate because the conditions established to identify latent responses' scales for the Configural Model may not be appropriate for subsequent models (i.e., Metric, Scalar), leading to incorrect conclusions about invariance of models. Instead, Wu and Estabrook recommend selecting a series of nested models wherein the identification conditions and invariance constraints are explicitly identified so that when model parameters are increasingly constrained in invariance testing, the scale of the underlying factors is comparable.

We implemented the Wu and Estabrook method using the WLSMV estimator in Mplus with theta parameterization. In the results, the models are referred to as WLSMV models. We selected a baseline model (WLSMV1) with non-invariant thresholds, loadings, and intercepts, but with invariant

residual variances. In addition, some parameters were not compared, including factor covariances (estimated), factor means (invariant), and factor variances (invariant). The metric was set by setting the first factor loading in the Black group to 1 and setting factor variances to 1 in the White group, and the model is identified by setting factors means to 0. The second model (WLSMV2) constrains loadings to be invariant. This model sets the first threshold of each item to 1, maintains the same constraints on factor variances as WLSMV1, and the factor means are also set to 0. The third model (WLSMV3) constrains thresholds to be invariant. This model maintains the factor variance constraints, and sets factor means to 0. Two models, WLSMV4a and WLSMV4b, hold invariant thresholds, loadings, intercepts, and residual variances. WLSMV4b alters the identification conditions so that factor variances are set to 1 and factor means are estimated in the Black group.

For WLSMV1-WLSMV4b, each model was compared with two different $\chi^2$ statistics. The first of these, $\chi^2_T$, is a standard model $\chi^2$ calculated as the fit function minimum*(2n), where $n$ is the total sample size [19]. $\chi^2_T$ is known to not be $\chi^2$ distributed, and therefore does not lend itself to calculation of $p$ values. The second, $\chi^2_{T3}$, theoretically should allow for improved $p$ value calculation but is known to misbehave by not increasing monotonically with nested, more restricted, models [20]. The WLSMV1-4b models were compared with a corrected $\chi^2$ difference test based on $\chi^2_{T3}$ in Mplus (DIFFTEST command) [21], but $\chi^2_T$ is also reported to verify increasing lack of fit with increasingly restricted models. For each of these models, RMSEA

and CFI based on $\chi^2_{T3}$ were calculated to indicate model fit. However, comparing these RMSEAs and CFIs for nested models is inappropriate.

Then, we compared model WLSMV4b to a series of models wherein each item's loadings and thresholds were released item-by-item. In these models, only the loadings and thresholds for the respective model are released; loadings and thresholds were not released cumulatively across the sequence of models. Each of these models was compared statistically to WLSMV4b with the corrected $\chi^2$ difference test. In addition, estimated factor means for the Black group from each model were compared to WLSMV4b.

## Results

### Participants

There were some differences in background characteristics between Black and White patients (Table 1). White patients tended to be older, more likely to use peritoneal dialysis, and more likely to be retired. A higher proportion of Black patients were unemployed.

### Measurement invariance with MLR models

Table 2 shows the loadings for White versus Black patients for KDQOL-36 Burden, Symptoms and Problems, and Effects of Kidney Disease scales from the configural

**Table 1** Patient characteristics by race

| | White ($n=21,439$) | Black ($n=18,404$) | $p$ value |
|---|---|---|---|
| Age (mean, range) | 66 (18–100) | 59 (18–97) | < 0.001 |
| Dialysis type (%, $n$) | | | < 0.001 |
|   In-center hemodialysis | 81% (17,307) | 86% (15,857) | |
|   Peritoneal dialysis | 14% (3090) | 9% (1692) | |
|   Conventional home hemodialysis | 3% (546) | 3% (529) | |
|   Other | 2% (496) | 2% (326) | |
| Dialysis access site (%, $n$) | | | < 0.001 |
|   Arteriovenous fistula | 43% (9154) | 43% (7836) | |
|   Arteriovenous graft | 5% (1103) | 9% (1734) | |
|   Venous catheter | 24% (5130) | 23% (4204) | |
|   PD catheter | 14% (3066) | 10% (1803) | |
|   Missing | 14% (2986) | 15% (2827) | |
| Diabetes status (% yes, $n$) | 51% (10,862) | 50% (9261) | 0.49 |
| Employment status (%, $n$) | | | < 0.001 |
|   Unemployed | 5% (1176) | 15% (2711) | |
|   Employed full or part time | 11% (2362) | 11% (2027) | |
|   Retired | 73% (15,573) | 61% (11,320) | |
|   Other | 4% (865) | 4% (728) | |
|   Missing | 7% (1176) | 9% (1618) | |

**Table 2** Configural model standardized factor loadings for White versus Black patients on KDQOL-36 Burden, Symptoms and Problems, and Effects of Kidney Disease scales—robust maximum likelihood estimation

| | KDQOL-36 Burden of Kidney Disease | | KDQOL-36 Symptoms and Problems of Kidney Disease | | KDQOL-36 Effects of Kidney Disease | | $R^2$ | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | White | Black | White | Black | White | Black | White | Black |
| My kidney disease interferes too much with my life (i13) | 0.83 | 0.82 | – | – | – | – | 0.69 | 0.68 |
| Too much time is spent dealing with kidney disease (i14) | 0.83 | 0.83 | – | – | – | – | 0.69 | 0.69 |
| I feel frustrated dealing with my kidney disease (i15) | 0.78 | 0.78 | – | – | – | – | 0.61 | 0.60 |
| I feel like a burden on my family (i16) | 0.60 | 0.59 | – | – | – | – | 0.36 | 0.35 |
| Soreness in your muscles? (i17) | – | – | 0.60 | 0.60 | – | – | 0.36 | 0.36 |
| Chest pain? (i18) | – | – | 0.45 | 0.48 | – | – | 0.21 | 0.23 |
| Cramps? (i19) | – | – | 0.47 | 0.51 | – | – | 0.22 | 0.26 |
| Itchy skin? (i20) | – | – | 0.52 | 0.55 | – | – | 0.28 | 0.31 |
| Dry skin? (i21) | – | – | 0.57 | 0.58 | – | – | 0.32 | 0.34 |
| Shortness of breath? (i22) | – | – | 0.53 | 0.57 | – | – | 0.28 | 0.33 |
| Faintness or dizziness? (i23) | – | – | 0.53 | 0.57 | – | – | 0.28 | 0.32 |
| Lack of appetite? (i24) | – | – | 0.50 | 0.52 | – | – | 0.25 | 0.27 |
| Washed out or drained? (i25) | – | – | 0.70 | 0.70 | – | – | 0.50 | 0.49 |
| Numbness in hands or feet? (i26) | – | – | 0.54 | 0.55 | – | – | 0.29 | 0.31 |
| Nausea or upset stomach? (i27) | – | – | 0.58 | 0.58 | – | – | 0.22 | 0.33 |
| Problems with your access/catheter site? (i28) | – | – | 0.32 | 0.37 | – | – | 0.11 | 0.14 |
| Fluid restriction? (i29) | – | – | – | – | 0.55 | 0.59 | 0.21 | 0.34 |
| Dietary restriction? (i30) | – | – | – | – | 0.59 | 0.62 | 0.35 | 0.38 |
| Your ability to work around the house? (i31) | – | – | – | – | 0.69 | 0.67 | 0.47 | 0.45 |
| Your ability to travel? (i32) | – | – | – | – | 0.66 | 0.69 | 0.44 | 0.48 |
| Being dependent on doctors and other medical staff? (i33) | – | – | – | – | 0.69 | 0.70 | 0.48 | 0.49 |
| Stress or worries caused by kidney disease? (i34) | – | – | – | – | 0.76 | 0.77 | 0.58 | 0.59 |
| Your sex life? (i35) | – | – | – | – | 0.51 | 0.52 | 0.26 | 0.27 |
| Your personal appearance? (i36) | – | – | – | – | 0.64 | 0.63 | 0.41 | 0.39 |

MLR model. Many loadings were very similar across the racial groups and all fell below the range of small differences (< 0.1) [17]. In Table 3, formal tests of measurement invariance from the Configural MLR model are shown. The Configural model (MLR1) showed good fit on RMSEA and SRMR but fell below standard for CFI: $\chi^2 = 37{,}293.22$, $df = 498$, $p < 0.001$; CFI = 0.886; RMSEA = 0.054; and SRMR = 0.041. In comparison to the Configural model, the Metric model (MLR2) constrained 21 factor loadings to be equal between the Black and White groups to test the factor loading invariance hypothesis. This resulted in a model with similar fit to the Configural model: $\chi^2 = 37{,}620.77$, $df = 519$, $p < 0.001$; CFI = 0.885; RMSEA = 0.053; and SRMR = 0.042. Finally, in comparison to the Metric model, the Scalar model (MLR3) estimated 3 additional factor means and 24 fewer intercepts: $\chi^2 = 39{,}681.72$, $df = 540$, $p < 0.001$; CFI = 0.879; RMSEA = 0.053; and SRMR = 0.044. This difference exceeded the ΔCFI cutoff for invariance (− 0.002). Therefore, to test partial Scalar invariance, the intercepts were freed for the KDQOL

Burdens scale (MLR4), and for the KDQOL Burdens and Symptoms and Problems scales simultaneously (MLR5). The second of these models demonstrated partial Scalar invariance.

## Measurement invariance with WLSMV models

Looking at the Wu and Estabrook method results, model WLSMV1 estimated 45 factor loadings, 24 item intercepts, 168 thresholds, 3 factor variances, and 6 factor covariances with $\chi^2_{T3} = 70{,}723.93$, $df = 498$, $p < 0.001$ (Table 4). Model WLSMV2 estimated 21 less factor loadings than WLSMV1, but was otherwise the same. Model WLSMV2's $\chi^2_{T3}$ was smaller than WLSMV1's: $\chi^2_{T3} = 49{,}577.68$, $df = 519$, $p < 0.001$. In comparison to WLSMV2, WLSMV3 estimated 72 fewer thresholds. WLSMV3's $\chi^2_{T3}$ was larger than WLSMV2's: $\chi^2_{T3} = 51{,}652.15$, $df = 591$, $p < 0.001$. Model WLSMV4a estimated 24 fewer intercepts than WLSMV3, and its $\chi^2_{T3}$ was smaller: $\chi^2_{T3} = 49{,}519.66$, $df = 615$, $p < 0.001$. Model WLSMV4b estimated the same number of parameters

**Table 3** Measurement invariance tests of White versus Black patients for KDQOL-36 Burden, Symptoms and Problems, and Effects of Kidney Disease scales—robust maximum likelihood estimation

| Model (#) | Ref model | $\chi^2$ | df | p | CFI | RMSEA | SRMR | $\Delta\chi^{2,a}$ | $\Delta df$ | p | $\Delta$CFI | $\Delta$RMSEA | $\Delta$SRMR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Configural (MLR1) | – | 37,293.22 | 498 | <0.001 | 0.886 | 0.054 | 0.041 | – | – | – | – | – | – |
| Metric (MLR2) | MLR1 | 37,620.77 | 519 | <0.001 | 0.885 | 0.053 | 0.042 | 280.61 | 21 | <0.001 | −0.001 | −0.001 | 0.001 |
| Scalar (MLR3) | MLR2 | 39,681.72 | 540 | <0.001 | 0.879 | 0.053 | 0.044 | 2233.69 | 21 | <0.001 | −0.006 | 0 | 0.002 |
| Partial Scalar invariance—free intercepts for KDQOL-36 Burden of Kidney Disease scale (i13-i16) (MLR4) | MLR2 | 39,204.69 | 537 | <0.001 | 0.880 | 0.053 | 0.044 | 1675.41 | 18 | <0.001 | −0.005 | 0 | 0.002 |
| Partial Scalar invariance—free intercepts for KDQOL-36 Burden, Problems and Symptoms of Kidney Disease scales (i13-i28) (MLR5) | MLR2 | 38,013.79 | 526 | <0.001 | 0.884 | 0.053 | 0.042 | 357.54 | 7 | <0.001 | −0.001 | 0 | 0 |

*Ref model* reference model, *Δ* corrected $\chi^2$ difference test for model in row above Metric versus Configural, Scalar versus Metric, *Configural* invariant factor structure, *Metric* Configural + invariant factor loadings, *Scalar* Metric + invariant intercepts, *CFI* comparative fit index, *RMSEA* root mean squared error of approximation, *SRMR* standardized root mean square residual

**Table 4** Measurement invariance tests of White versus Black patients for KDQOL-36 Burden, Symptoms and Problems, and Effects of Kidney Disease scales—Wu and Estabrook method using weighted least squares with mean and variance adjustment estimation

| Model | Model description | $\chi^2_{T1}$ | $\chi^{2\,a}_{T3}$ | df | p | $\Delta\chi^{2b,c}$ | $\Delta df$ | p | RMSEA | CFI |
|---|---|---|---|---|---|---|---|---|---|---|
| WLSMV1 | Constrains residual variances to be equal; freely estimates factor loadings, intercepts, and thresholds | 51,336.09 | 70,723.93 | 498 | <0.001 | – | – | – | 0.074 | 0.925 |
| WLSMV2 | Constrains residual variances and factor loadings to be equal; freely estimates intercepts, and thresholds | 51,734.11 | 49,577.68 | 519 | <0.001 | 113.71 | 21 | <0.001 | 0.061 | 0.948 |
| WLSMV3 | Constrains residual variances, factor loadings, and thresholds to be equal; freely estimates intercepts | 52,303.01 | 51,652.15 | 591 | <0.001 | 934.71 | 72 | <0.001 | 0.058 | 0.945 |
| WLSMV4a | Constrains residual variances, factor loadings, thresholds, and intercepts to be equal. Factor means constrained to 0 in both groups. Factor variances freed in Black group | 62,082.19 | 49,519.66 | 615 | <0.001 | 2,555.98 | 24 | <0.001 | 0.056 | 0.948 |
| WLSMV4b | Constrains residual variances, factor loadings, thresholds, and intercepts to be equal. Freed factor means in Black group. Factor variances constrained in both groups | 56,712.63 | 44,970.01 | 615 | <0.001 | 1,172.71 | 24 | <0.001 | 0.053 | 0.953 |

[a] $\chi^2_{T3}$ Calculated from WLSMV models may not be ordered as is typical for nested models

[b] $\Delta\chi^2$ Corrected $\chi^2$ difference test for model in row above except for model WLSMV4b (compared to WLSMV3)

[c] $\Delta\chi^2$ Implemented with the DIFFTEST option in Mplus

as WLSMV4a, but freed 3 factor means in the Black group and constrained 3 factor variances. This model had a smaller $\chi^2_{T3}$ than WLSMV3: $\chi^2_{T3} = 44{,}970.01$, $df = 615$, $p < 0.001$. As expected, $\chi^2_{T1}$ values were ordered and ascending from WLSMV1-WLSMV4b, indicating worse fit with increasingly constrained models. All corrected $\chi^2$ difference tests were statistically significant, RMSEA ranged between 0.051 and 0.074, and CFI ranged between 0.925 and 0.953, indicating borderline or good fit (Table 4).

Table 5 shows sequential item-by-item released models (loadings and thresholds) in comparison to model WLSMV4b. With the exception of the i27 released model, all corrected $\chi^2$ difference test were significant at $p < 0.001$. However, differences in factor means were small.

## Discussion

A recent report provided support for the dimensional structure of the KDQOL-36 among U.S. dialysis patients [6]. Using the same dataset, this study found minimal impact of DIF between Black and White patients on KDQOL-36 score estimation. Though some DIF was observed, the KDQOL-36 scales have largely similar measurement properties for U.S. Black and White dialysis patients. These results indicate that substantive comparisons between Black and White patients using the KDQOL are appropriate.

In the MLR analyses, Configural and Metric invariance was observed, indicating that each KDQOL-36 item has similar importance in its contribution to measuring HRQOL among Black and White patients. This level of invariance is most important for allowing valid substantive comparisons of HRQOL between Black and White patients. However, only partial Scalar invariance was evidenced, indicating the Black and White patients may start from significantly

**Table 5** Difference in factor means between fully invariant and released loadings and thresholds item-by-item models using weighted least squares with mean and variance adjustment estimation

| Scale | Item | $\chi^{2,a}$ difference | | | Factor mean difference | | |
|---|---|---|---|---|---|---|---|
| | | $\Delta\chi^2$ | $\Delta df$ | $p$ value | Invariant model[b] | Released model | $\Delta$ |
| KDQOL-36 Burden of Kidney Disease | i13 | 203.76 | 5 | < 0.001 | 0.218 | 0.246 | 0.028 |
| | i14 | 98.02 | 5 | < 0.001 | – | 0.229 | 0.011 |
| | i15 | 74.58 | 5 | < 0.001 | – | 0.232 | 0.014 |
| | i16 | 415.45 | 5 | < 0.001 | – | 0.17 | 0.048 |
| KDQOL-36 Symptoms and Problems of Kidney Disease | i17 | 40.70 | 5 | < 0.001 | 0.061 | 0.053 | 0.008 |
| | i18 | 72.51 | 5 | < 0.001 | – | 0.071 | 0.01 |
| | i19 | 96.31 | 5 | < 0.001 | – | 0.064 | 0.003 |
| | i20 | 86.24 | 5 | < 0.001 | – | 0.07 | 0.009 |
| | i21 | 159.26 | 5 | < 0.001 | – | 0.076 | 0.015 |
| | i22 | 74.23 | 5 | < 0.001 | – | 0.059 | 0.002 |
| | i23 | 132.27 | 5 | < 0.001 | – | 0.055 | 0.006 |
| | i24 | 112.81 | 5 | < 0.001 | – | 0.074 | 0.013 |
| | i25 | 650.03 | 5 | < 0.001 | – | 0.017 | 0.044 |
| | i26 | 40.12 | 5 | < 0.001 | – | 0.055 | 0.006 |
| | i27 | 5.12 | 5 | 0.40 | – | 0.062 | 0.001 |
| | i28 | 121.90 | 5 | < 0.001 | – | 0.071 | 0.01 |
| KDQOL-36 Effects of Kidney Disease | i29 | 119.81 | 5 | < 0.001 | 0.161 | 0.169 | 0.008 |
| | i30 | 137.05 | 5 | < 0.001 | – | 0.165 | 0.004 |
| | i31 | 170.47 | 5 | < 0.001 | – | 0.139 | 0.022 |
| | i32 | 125.01 | 5 | < 0.001 | – | 0.151 | 0.01 |
| | i33 | 75.20 | 5 | < 0.001 | – | 0.157 | 0.004 |
| | i34 | 108.19 | 5 | < 0.001 | – | 0.171 | 0.01 |
| | i35 | 70.93 | 5 | < 0.001 | – | 0.173 | 0.012 |
| | i36 | 51.60 | 5 | < 0.001 | – | 0.164 | 0.003 |

[a] $\Delta\chi^2$ Corrected $\chi^2_{T3}$ difference test against invariant model

[b] Invariant model is Model WLSMV4b from Table 4

different levels of HRQOL when responding to the KDQOL-36 items. This concern is not to be overlooked, though it is somewhat allayed when the WLSMV models are considered.

The WLSMV models showed mixed results regarding model fit with increasing invariance constraints. The $\chi^2_{T1}$ values increased across these models (indicating worse fit), while the values of the fit indexes decreased (indicating better fit), though these fit indexes may not be valid for comparison of the fit of nested models. However, when any one item's loadings and thresholds were released, changes in factor means were trivial. A similar set of results was found when the loadings and thresholds of each kidney-targeted KDQOL scale were released simultaneously. These results indicate that whatever DIF was present, it seems not to have an appreciable impact on average levels of HRQOL when estimated by the KDQOL-36. Nonetheless, additional studies examining measurement invariance between Black and White dialysis patients on the KDQOL-36 scales would be useful to provide further evidence.

A previous study examined measurement invariance of the KDQOL-SF scales between Veterans and non-Veterans [22]. That study found that there may be DIF on some of the kidney-targeted scales in the KDQOL survey. Like in the current study, these scales evidenced Metric invariance (invariant factor loadings), but not Scalar invariance, indicating difference in intercepts between groups. The authors concluded that additional study is needed before DIF between Veterans and non-Veterans could be confirmed.

Black patients had higher estimated factor means in comparison to Whites, indicating better HRQOL. These results are consistent with the results of the DOPPS study, which found that Black dialysis patients had higher scores on the KDQOL Burden and Effects of Kidney Disease scales [9]. Additionally, another large study, the Hemodialysis (HEMO) Study, which included 1,156 Black and 657 non-Black dialysis patients, found that Black patients had significantly higher scores on the KDQOL Symptoms and Problems of Kidney Disease, Effects of Kidney Disease, and Burden of Kidney Disease scales in comparison to non-Black patients [23]. One smaller study of Black and White dialysis patients found that Whites perceived dialysis to be more burdensome than Black patients and were less satisfied with their lives on dialysis [24]. This evidence indicates that Black dialysis patients have better HRQOL than White dialysis patients.

Like all studies, this study has limitations to consider when interpreting its results. First, at 0.866, the CFI value for model MLR1 (Configural model) is somewhat lower than the threshold indicating good fit ($> 0.95$). In tests of measurement invariance, severe misfit of the Configural model can lead to non-meaningful results from subsequent nested models. We argue that though the value is somewhat low, it does not indicate severe misfit, especially considering RMSEA $= 0.054$ and SRMR $= 0.041$. Therefore, despite this

lower than ideal CFI value, estimation of additional models in this paper is valid. Next, though sample size was very large, the patients participating in this study are a convenience sample and may be subject to bias. Nationally representative studies of dialysis patients' HRQOL could be used to confirm our findings from the current study. Finally, as we have pointed-out elsewhere, the data used in this study were not collected for the study's objectives [6].

## Conclusion

This paper examined whether the KDQOL-36, a widely used HRQOL measure among dialysis patients, had equivalent measurement properties between Black and White patients. Since we found negligible impact of measurement non-invariance on KDQOL-36 factor mean estimation between these groups, substantive comparisons of HRQOL between Black and White dialysis patients can be made using the KDQOL-36. The ability to make such comparisons is critical for epidemiological surveillance of dialysis patients' health, and to use in interventions to improve HRQOL among dialysis patients.

## Compliance with ethical standards

**Conflict of interest** All authors declare no conflict of interest.

**Ethical approval** This article does not contain any studies with human subjects performed by any of the authors.

## References

1. United States Renal Data System. (2017). 2017 USRDS annual data report: Epidemiology of kidney disease in the United States. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.
2. Hertz, R. P., Unger, A. N., Cornell, J. A., & Saunders, E. (2005). Racial disparities in hypertension prevalence, awareness, and management. *Archives of Internal Medicine, 165*(18), 2098–2104.
3. Centers for Disease Control and Prevention. (2017). National Diabetes Statistics Report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services.
4. Hays, R. D., Cunningham, W. E., Sherbourne, C. D., Wilson, I. B., Wu, A. W., Cleary, P. D., McCaffrey, D. F., Fleishman, J. A., Crystal, S., Collins, R., Eggan, F., Shapiro, M. F., & Bozzette, S. A. (2000). Health-related quality of life in patients with human immunodeficiency virus infection in the United States: Results from the HIV cost and services utilization study. *The American Journal of Medicine, 108*(9), 714–722.

5. Sledge, R. (2010). KDQOL-36 and the interdisciplinary team. *Nephrology News & Issues, 24*(7), 36–38.

6. Peipert, J. D., Bentler, P. M., Klicko, K., & Hays, R. D. (2017). Psychometric properties of the kidney disease quality of life 36-item short-form survey (KDQOL-36) in the United States. *American Journal of Kidney Diseases*. https://doi.org/10.1053/j.ajkd.2017.07.020.

7. Hays, R. D., Kallich, J. D., Mapes, D. L., Coons, S. J., & Carter, W. B. (1994). Development of the kidney disease quality of life (KDQOL) instrument. *Quality of Life Research, 3*(5), 329–338.

8. Tell, G. S., Mittelmark, M. B., Hylander, B., Shumaker, S. A., Russell, G., & Burkart, J. M. (1995). Social support and health-related quality of life in black and white dialysis patients. *ANNA Journal, 22*(3), 301–308; discussion 309–310.

9. Lopes, A. A., Bragg-Gresham, J. L., Satayathum, S., McCullough, K., Pifer, T., Goodkin, D. A., Mapes, D. L., Young, E. W., Wolfe, R. A., Held, P. J., & Port, F. K. (2003). Health-related quality of life and associated outcomes among hemodialysis patients of different ethnicities in the United States: The Dialysis Outcomes and Practice Patterns Study (DOPPS). *American Journal of Kidney Diseases, 41*(3), 605–615.

10. Mapes, D. L., Bragg-Gresham, J. L., Bommer, J., Fukuhara, S., McKevitt, P., Wikstrom, B., & Lopes, A. A. (2004). Health-related quality of life in the Dialysis Outcomes and Practice Patterns Study (DOPPS). *American Journal of Kidney Diseases, 44*(5 Suppl 2), 54–60.

11. Muthén, L. K., & Muthén, B. O. (2017). Mplus user's guide. Los Angeles, CA.

12. Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structure Equation Modeling-a Multidisciplinary Journal, 21*(2), 167–180.

13. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling-a Multidisciplinary Journal, 6*(1), 1–55.

14. Bentler, P. M. (2006). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

15. Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592.

16. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504.

17. Yoon, M., & Millsap, R. E. (2007). Detecting violations of factorial invariance using data-based specification searches: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 435–463.

18. Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045.

19. Asparouhov, T., & Muthen, B. (2010). Simple second order chi-square correction. Los Angeles, CA: Muthén and Muthén.

20. Fischer, F., Gibbons, C., Coste, J., Valderas, J. M., Rose, M., & Leplege, A. (2018). Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Quality of Life Research*. https://doi.org/10.1007/s11136-018-1785-8.

21. Asparouhov, T., & Muthen, B. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics*. Los Angeles, CA: Mplus.

22. Saban, K. L., Bryant, F. B., Reda, D. J., Stroupe, K. T., & Hynes, D. M. (2010). Measurement invariance of the kidney disease and quality of life instrument (KDQOL-SF) across veterans and non-veterans. *Health and Quality of Life Outcomes, 8*(1), 120.

23. Unruh, M., Miskulin, D., Yan, G., Hays, R. D., Benz, R., Kusek, J. W., Meyer, K. B., & HEMO Study Group. (2004). Racial differences in health-related quality of life among hemodialysis patients. *Kidney International, 65*(4), 1482–1491.

24. Kutner, N. G., & Devins, G. M. (1998). A comparison of the quality of life reported by elderly whites and elderly blacks on dialysis. *Geriatric Nephrology and Urology, 8*(2), 77–83.