

Kidney Disease Quality of Life 36-Item Short Form Survey (KDQOL-36) Normative Values for the United States Dialysis Population and New Single Summary Score

John D. Peipert,^{1,2} Devika Nair,³ Kristi Klicko,⁴ Dorian R. Schatell,⁴ and Ron D. Hays⁵

¹Department of Medical Social Sciences and ²Northwestern University Transplant Outcomes Research Collaborative, Comprehensive Transplant Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois; ³Division of Nephrology and Hypertension, Vanderbilt University Medical Center, Nashville, Tennessee; ⁴Medical Education Institute, Inc., Madison, Wisconsin; and ⁵Division of General Internal Medicine and Health Services Research, University of California Los Angeles, Los Angeles, California

ABSTRACT

Background The Kidney Disease Quality of Life 36-item short form survey (KDQOL-36) is a widely used, patient-reported outcome measure for patients on dialysis. Efforts to aid interpretation are needed.

Methods We used a sample of 58,851 dialysis patients participating in the Medical Education Institute (MEI) KDQOL Complete program, and 443,947 patients from the US Renal Data System (USRDS) to develop the KDQOL-36 Summary Score (KSS) for the kidney-targeted KDQOL-36 scales (Burdens of Kidney Disease [BKD], Symptoms and Problems of Kidney Disease [SPKD], and Effects of Kidney Disease [EKD]). We also used the MEI and USRDS data to calculate normative values for the Short Form-12 Health Survey's Physical Component Summary (PCS) and Mental Component Summary (MCS), and the KDQOL-36's BKD, SPKD, and EKD scales for the United States dialysis population. We used confirmatory factor analysis (CFA) models for KDQOL-36 kidney-targeted items, evaluated model fit with the comparative fit index (CFI; >0.95 indicates good fit) and root-mean-squared error of approximation (RMSEA; <0.06 indicates good fit), and estimated norms by matching the joint distribution of patient characteristics in the MEI sample to those of the USRDS sample.

Results A bifactor CFA model fit the data well (RMSEA=0.046, CFI=0.990), supporting the KSS ($\alpha=0.91$). Mean dialysis normative scores were PCS=37.8 and MCS=50.9 (scored on a T-score metric); and KSS=73.0, BKD=52.8, SPKD=79.0, and EKD=74.1 (0–100 possible scores).

Conclusions The KSS is a reliable summary of the KDQOL-36. The United States KDQOL-36 normative facilitate interpretation and incorporation of patient-related outcome measures into kidney disease care.

J Am Soc Nephrol 30: 654–663, 2019. doi: <https://doi.org/10.1681/ASN.2018100994>

Stakeholder panels of kidney patients, nephrologists, psychometricians, and industry representatives have called for the continued incorporation of rigorously developed patient-reported outcomes (PROMs) into clinical trials and health policy studies.^{1–5} Health-related quality of life (HRQOL) is often measured by PROMs, and the Centers for Medicare and Medicaid Services (CMS) mandates that all patients with ESRD have HRQOL assessed annually.^{6–8} The Kidney Disease Quality of Life 36-item

Received August 10, 2018. Accepted January 28, 2019.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. John D. Peipert, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, 625 Michigan Avenue, Suite 2700, Chicago, IL 60611. Email: john.peipert@northwestern.edu

Copyright © 2019 by the American Society of Nephrology

short form survey (KDQOL-36) is one of the most widely-used measures of HRQOL for patients with ESRD.⁹ The KDQOL-36 includes both generic and ESRD-specific HRQOL scales, which facilitates comparison with other clinical subgroups as well as offering specificity and responsiveness to change for elements of HRQOL important to patients with ESRD.¹⁰

The psychometric properties of the KDQOL-36 have been evaluated in United States patients on dialysis, but there is a need for guidance on its interpretation to facilitate its use in clinical practice.^{11,12} Normative values of a PROM that are referenced to a relevant population are useful to help guide interpretation. As an example, the National Institutes of Health Patient-Reported Outcome Measurement Information System (PROMIS) generic HRQOL measures have been normed to the United States general population on a T-score metric, with a mean of 50 and SD of 10, representing the United States general population average and SD.¹³ In previous research, a Kidney Disease Component Summary (KDCCS) score that averaged 11 ESRD-targeted scales from the longer KDQOL short form instrument was used in the Dialysis Outcomes and Practice Patterns Study.^{7,14,15} Shortening measure length and minimizing measure burden on patients is also important, as these characteristics have the ability to ease implementation of PROMs into practice. A composite score from the KDQOL-36 short form (Burdens of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease) has not been constructed.

To facilitate implementation and interpretation of the KDQOL-36 in clinical practice, we examine whether there is psychometric support for a composite score of items from this measure's three kidney-targeted scales and to estimate normative values in the United States dialysis population for the Short Form (SF-12) version 1 Physical Component Summary (PCS), SF-12 version 1 Mental Component Summary (MCS), the KDQOL-36 kidney-targeted scales, and the new KDQOL-36 composite score.

METHODS

Data and Patient Sample Selection

The primary dataset used in this study was provided by the nonprofit Medical Education Institute (MEI), which administers the KDQOL Complete scoring program used primarily by small dialysis organization clinics throughout the United States.¹⁰ KDQOL Complete helps dialysis centers complete and score the KDQOL-36 instrument and uses responses to the KDQOL-36 scales to help develop personalized plans of care for patients on dialysis. Additional information on the KDQOL Complete program and MEI has been published elsewhere.¹¹ Because all data provided by the MEI were deidentified, an institutional review board exemption was granted by the University of California, Los Angeles Human Subjects Protection Committee (approval no. 17-000313), including waiver of informed consent.

Significance Statement

Reliable, valid, and interpretable patient-reported outcome measures for kidney patients are needed for patient monitoring and use as outcomes in clinical trials. The Kidney Disease Quality of Life 36-item short form survey (KDQOL-36) is often used with patients on dialysis, but improvements are needed to facilitate interpretability of its scores. The authors calculate normative values for the KDQOL-36 scales referenced to the United States dialysis population, which allow comparison of group means and individual scores with national averages, such as by dialysis centers when fulfilling their required annual assessment of patients' quality of life. The authors also created the KDQOL-36 Summary Score (KSS), a composite of items from the KDQOL-36's kidney-targeted scales, which may be useful when kidney-targeted health-related quality of life needs to be summarized in a single score.

We started with a set of 77,072 surveys and analyzed data from 58,851 patients in the MEI dataset after excluding 4090 surveys that were completed a second time by patients (*i.e.*, only the first survey completed was used for patients who completed the survey multiple times). We excluded the following from our analyses: 1275 assessments with a missing assessment date ($n=1273$) or missing all survey data ($n=2$), 585 assessments from patients under the age of 18, 336 assessments from patients who had not yet started dialysis or who had received a previous transplant, 10,534 patients with missing information on race and ethnicity, and 1401 patients who self-identified as Native Hawaiian/Pacific Islander because as this ethnic subgroup was not included in the US Renal Data System (USRDS) dataset. As a secondary analysis dataset used for test-retest reliability estimation, we retained the second assessments of patients who completed the KDQOL-36 twice within 1–21 days ($n=103$). We used aggregated data from 443,947 patients on dialysis from the USRDS to represent the United States dialysis population.¹⁶ Because the data were aggregated, we did not have access to individual-level patients, and therefore did not perform individual case selection comparable with the case selection performed for the MEI dataset.

Measures

The KDQOL-36 has five scales, including two generic HRQOL scales from the SF-12 version 1 (12 items total) and three kidney-specific scales (24 items total). The SF-12 PCS and MCS are scored on a T-score metric (mean=50, SD=10, in the United States general population), with higher scores indicating better HRQOL.¹² The kidney-targeted scales include Burden of Kidney Disease (four items; *e.g.*, "My kidney disease interferes too much with my life"), Symptoms and Problems of Kidney Disease (12 items; *e.g.*, "Washed out or drained?"), and Effects of Kidney Disease (eight items; *e.g.*, "Your ability to work around the house").¹³ The Burden of Kidney Disease items are prompted with the context, "How true or false is each of the following statements?" and have five response options that range from "definitely true" to "definitely false." The Symptoms

and Problems with Kidney Disease items are given the context, “During the past 4 weeks, to what extent were you bothered by each of the following?” and have five response options ranging from “not at all bothered” to “extremely bothered.” The Effects of Kidney Disease scale’s items ask patients, “How much does kidney disease bother you in each of the following areas?” and have five response options ranging from “not at all bothered” to “extremely bothered.” Each of the KDQOL-36 kidney-targeted scales are scored by transforming all items linearly to a 0–100 possible range and averaging the items in the scale. On the KDQOL-36, higher scores indicate better HRQOL.¹⁴ For each KDQOL-36 scale, we also present results in the T-score metric, with a mean of 50 and SD of 10, in the United States dialysis population.

Relevant clinical and demographic characteristics were collected on the CMS Form 2728 in the dialysis facilities or by the dialysis facilities themselves for the MEI sample. We measured age, sex, race/ethnicity, dialysis type (hemodialysis, peritoneal dialysis), and presence of diabetes. The same variables were obtained for the USRDS national dialysis population.

Examination of a KDQOL-36 Composite

We examined whether a composite could be created from the items from the KDQOL-36 Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease scales using multiple approaches. First, exploratory factor analysis (EFA) was conducted using maximum likelihood estimation (Supplemental Appendix 1). To determine the number of factors to retain, we examined two criteria: (1) the Kaiser rule, which retains the number of factors equal to the number of eigenvalues >1 ; and (2) parallel analysis, which retains the number of factors wherein eigenvalues from the data analyzed are larger than the eigenvalues from an EFA on a random dataset with the same number of observations and variables. In the presence of evidence from these analyses, which includes a ratio of the first to second eigenvalue >4 ,¹⁷ an exploratory bifactor model was estimated using the psych package in R. This analysis extracts Schmid–Leiman transformed factor loadings to determine how well each item loads on a general factor in addition to specific factors.¹⁸ This analysis also generates ω_h , a measure of general factor saturation, or the proportion of total variance accounted for by the general factor.¹⁹ ω_h values of ≥ 0.70 indicate that a measure is essentially unidimensional.²⁰ Finally, we examined the explained common variance (ECV) from this analysis. The ECV is computed as the ratio of the general factor’s eigenvalue to the sum of all eigenvalues from the analysis and, like the ω_h , is a measure of general factor saturation. Values of ≥ 0.60 indicate the presence of a strong general factor.²⁰

Next, we conducted confirmatory factor analysis (CFA) model using the lavaan package in R.²¹ We compared several factorial structures, each with different CFA models. First, we fit a unidimensional model, where all items were set to load on

only one factor (Supplemental Appendix 2 [Supplemental Figure 2]). Next, we fit a three-correlated-factors CFA model, which represents the factor structure previously published for the KDQOL-36.¹¹ The three-correlated-factors model sets four items to load on one factor representing Burdens of Kidney Disease (Table 1, items i13–i16), 12 items to load on a second factor representing Symptoms and Problems of Kidney Disease (items i17–i28), and a third factor representing Effects of Kidney Disease (items i29–i36). Each of these factors is allowed to correlate with one another (Supplemental Appendix 2 [Supplemental Figure 3]). Finally, we fit a bifactor model. This model evaluates how well a general factor, in addition to specific factors, account for the latent structure of a measure.²² For this model, we included three specific factors matching those from the three-correlated-factors model, each with the same items loading on the Burdens of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease factors as described above, except the factors are left uncorrelated. In addition, each item loads directly onto a general factor (Supplemental Appendix 2 [Supplemental Figure 4]). Good fit of a bifactor model to the data provides support for a composite summary score.

Because the KDQOL-36 items have ordered categorical responses, CFA models were estimated using the diagonally weighted least squares estimator. Model fit was evaluated using the model’s chi-squared statistic, the comparative fit index (CFI), and the root mean squared error of approximation (RMSEA). Nonsignificant chi-squared statistics, CFI values >0.95 , and RMSEA values <0.06 indicate good model fit.²³ However, it is worth noting that the chi-squared statistic is sensitive to sample size and may be statistically significant ($P<0.05$) even in the presence of good model fit when the sample size is large.

We examined construct validity of a kidney-specific composite by examining Pearson correlations between the composite and the SF-12 PCS and MCS. The following conventions were used to interpret the magnitude of correlation coefficients, which correspond to standardized effect sizes: small, $0.10 \leq r < 0.243$; medium, $0.243 \leq r < 0.371$; and large, $r \geq 0.371$.²⁴ Finally, we examined reliabilities of the KDQOL-36 scales using two approaches. First, coefficient α was used to estimate the internal consistency reliability of each KDQOL-36 kidney-specific scale, both overall and for race/ethnicity \times age subgroups. The magnitude of α coefficients was assessed using the following standard cut-off criteria: acceptable, $0.70 \leq \alpha < 0.80$; good, $0.80 \leq \alpha < 0.90$; and excellent, $\alpha \geq 0.90$.²⁵ Because of the nature of the scoring of the SF-12 version 1 PCS and MCS, coefficient α does not provide an appropriate estimate of reliability. Hence, we used the subsample of patients with second assessments between 1–21 days after the first assessment, all scales’ test-retest reliability was estimated for the overall sample ($n=103$) using intraclass correlation coefficients. The magnitude of intraclass correlation coefficients for test-retest reliability was assessed with the following cut-off criteria: marginal, <0.40 ; good, $0.40–0.75$; and excellent, >0.75 .²⁶

Table 1. Bifactor CFA model factor loadings for the KDQOL-36

Item	General Factor	Specific Factor: Burden of Kidney Disease	Specific Factor: Symptoms/Problems of Kidney Disease	Specific Factor: Effects of Kidney Disease
My kidney disease interferes too much with my life (i13)	0.63	0.62	—	—
Too much time is spent dealing with kidney disease (i14)	0.60	0.68	—	—
I feel frustrated dealing with my kidney disease (i15)	0.67	0.48	—	—
I feel like a burden on my family (i16)	0.61	0.32	—	—
Soreness in your muscles? (i17)	0.53	—	0.33	—
Chest pain? (i18)	0.46	—	0.39	—
Cramps? (i19)	0.40	—	0.32	—
Itchy skin? (i20)	0.37	—	0.67	—
Dry skin? (i21)	0.41	—	0.65	—
Shortness of breath? (i22)	0.47	—	0.38	—
Faintness or dizziness? (i23)	0.50	—	0.35	—
Lack of appetite? (i24)	0.46	—	0.31	—
Washed out or drained? (i25)	0.65	—	0.30	—
Numbness in hands or feet? (i26)	0.47	—	0.34	—
Nausea or upset stomach? (i27)	0.51	—	0.36	—
Problems with your access/catheter site? (i28)	0.41	—	0.17	—
Fluid restriction? (i29)	0.57	—	—	0.50
Dietary restriction? (i30)	0.61	—	—	0.67
Your ability to work around the house? (i31)	0.75	—	—	−0.01
Your ability to travel? (i32)	0.70	—	—	0.07
Being dependent on doctors and other medical staff? (i33)	0.75	—	—	0.04
Stress or worries caused by kidney disease? (i34)	0.84	—	—	−0.01
Your sex life? (i35)	0.58	—	—	0.02
Your personal appearance? (i36)	0.73	—	—	0.02

—, item was not set to load on this factor.

Calculating Normative Scores

Normative scores were estimated using weights to match the joint distribution of age, sex, race/ethnicity, dialysis type, and etiology of ESRD in the study sample to the United States general dialysis population sourced from the USRDS.¹⁵ We divided both the MEI ($n=58,851$) and USRDS ($n=443,947$) samples into cells derived from nested demographic groups, starting with sex at the highest level (male versus female). We divided the sample within both sex groups into racial/ethnic subgroups (Hispanic, non-Hispanic white, black, Asian, American Indian), then by type of dialysis (hemodialysis versus peritoneal dialysis), and then by etiology of ESRD (diabetes versus no diabetes). Finally, we divided each of these subdivisions by age range (18–29, 30–44, 45–59, 60–74, and ≥ 75 years). This created a matrix of 200 cells.

We calculated cell weights by dividing each cell percentage in the USRDS sample by the corresponding cell percent from the MEI sample to create the MEI weighted sample. For example, among the USRDS sample, the cell percentage of patients who were female, non-Hispanic White, on hemodialysis, with diabetes, in the 45–59 age range was 1.443%, and the corresponding cell percentage in the MEI sample was 1.383%, resulting in a weight of 1.04 for this cell. This weight was used in the calculation of mean scores for each of the KDQOL-36 scales using the survey means procedure in SAS v9.4.²⁷ Normative values were calculated for the overall sample, as well as by race/ethnic and age groups.

RESULTS

Study Samples

The characteristics of the MEI sample, MEI weighted sample, and USRDS sample are shown in Table 2. Differences were apparent between demographic and clinical characteristics between the MEI sample and the USRDS sample. For example, there was a 6.1% difference in the proportion of Hispanic patients between the MEI sample and the USRDS sample. After weighting the MEI sample (MEI weighted sample), the proportions of patients in each different subgroup categories differed by no greater than 0.1%.

KDQOL-36 Summary Score

The EFA results supported a three factor solution with strong potential for a bifactor structure. The ω_h was 0.78, which exceeds the cut-off to evidence essential unidimensionality and indicates that 78% of the items' variance is explained by the general factor. The ECV was 0.68, also exceeding the cut-off to indicate essential unidimensionality. Results of these analyses are given in the Supplemental Materials.

The single-factor CFA model fit the data poorly (chi square, 143,137.08; degrees of freedom [df], 252; $P < 0.001$; RMSEA=0.105; CFI=0.945). By comparison, both the three-correlated-factors model (chi square, 51,252.81; df=249; $P < 0.001$; RMSEA=0.063; CFI=0.980) and the bifactor model

Table 2. Dialysis patient characteristics in the MEI sample, MEI weighted sample, and prevalent adult United States national dialysis sample from the USRDS

Patient Characteristics	MEI Raw Sample, n=58,851	MEI Weighted Sample, n=58,846	USRDS Sample, n=443,947
Age, yr, % (n)			
18–29	2.4 (1389)	2.2 (1276)	2.1 (9610)
30–44	10.5 (6196)	10.5 (6174)	10.4 (46,586)
45–59	29.1 (17,120)	28.4 (16,711)	28.4 (126,154)
60–74	38.5 (22,650)	38.4 (22,584)	38.4 (170,355)
≥75	19.5 (11,498)	20.6 (12,101)	20.6 (91,242)
Sex, % (n)			
Women	43.3 (25,458)	43.4 (25,567)	43.4 (192,787)
Men	56.7 (33,393)	56.6 (33,393)	56.6 (251,160)
Race/ethnicity, % (n)			
Hispanic	23.1 (13,594)	17.0 (9995)	17.0 (75,447)
Non-Hispanic white	37.3 (21,972)	40.8 (24,008)	40.8 (181,091)
Non-Hispanic black	32.0 (18,855)	35.8 (21,048)	35.8 (158,788)
Non-Hispanic Asian	5.8 (3414)	5.3 (3414)	5.3 (23,498)
Non-Hispanic American Indian	1.7 (1016)	1.2 (1016)	1.2 (5123)
Dialysis type, % (n)			
Hemodialysis	88.2 (51,904)	90.3 (53,141)	90.3 (400,920)
Peritoneal dialysis	11.8 (6947)	9.7 (5705)	9.7 (43,027)
Diabetes, % yes (n)	53.5 (31,522)	46.3 (27,269)	46.4 (205,807)

(chi square, 24,960.99; df=228; $P<0.001$; RMSEA=0.046; CFI=0.990) fit the data very well, but the bifactor model fit better (Table 3). Higher loadings on the general factor in comparison to specific factors indicates that the general factor explains more variance for the item, providing support for the general factor. Of the 24 KDQOL-36 kidney-specific items, only one item loaded on the general factor of the bifactor model at a magnitude of <0.40 (range: 0.37–0.84): i20, “itchy skin.” Only four of the 24 items had a higher loading on any of the specific factors than on the general factor. One of these was from the Burden of Kidney Disease scale, i14 (“too much time is spent dealing with kidney disease”). Two were from the Symptoms and Problems of Kidney Disease scale, i20 (“itchy skin”) and i21 (“dry skin”), and the last item with a higher loading on a specific factor was on the Effects of Kidney Disease scale, i30 (“dietary restriction”).

The correlations among three factors were noteworthy in the three-correlated-factors model: Burden of Kidney Disease with Symptoms and Problems of Kidney Disease, $r=0.57$; Burden of Kidney Disease with Effects of Kidney Disease, $r=0.74$; and Symptoms and Problems of Kidney Disease with Effects of Kidney Disease, $r=0.74$. These correlations suggest the presence of a general factor that underlies these three scales.

Table 3. Fit of CFA models for kidney-specific scales of the KDQOL-36

Model	RMSEA	CFI	Chi Square; df; P Value
Unidimensional	0.105	0.945	143,137.08; 252; $P<0.001$
Three-correlated-factors model	0.063	0.980	51,252.81; 249; $P<0.001$
Bifactor model	0.046	0.990	24,960.99; 228; $P<0.001$

RMSEA good fit <0.06 ; CFI good fit >0.95 .

Informed by these results, we developed the KDQOL-36 Summary Score (KSS) scale. We averaged together the 24 items from the Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease scales. In addition to its original scoring on a 0–100 possible range, the KSS was also scored on a T-score metric with a mean of 50 and an SD of 10 in the dialysis population.

Each KDQOL-36 kidney-specific scale had a “good” coefficient α (≥ 0.80), and the KSS had an “excellent” coefficient α (≥ 0.90), at 0.91 (Table 4). Of note, none of the scales’ reliabilities could be increased by removing items. Coefficient α scores within race/ethnicity \times age subgroups were all in the range of the overall sample’s estimates (Supplemental Appendix 3 [Supplemental Table 4]). For the test-retest sample ($n=103$), all KDQOL-36 kidney-targeted scales and the SF-12 PCS had excellent test-retest reliabilities (>0.75), and the MCS had good test-retest reliability (0.40–0.75) (Table 4). The KSS correlated at a large magnitude with the SF-12 PCS ($r=0.51$; $P<0.001$) and MCS ($r=0.59$; $P<0.001$).

Normative Scores Referenced to the United States Dialysis Population

Normative scores for the SF-12 PCS, SF-12 MCS, KSS, Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease referenced to the United States dialysis population using the MEI weighted sample are presented in Tables 4 and 5 as mean scale scores with 95% confidence intervals. Table 4 presents the overall sample normative scores in both the original scoring and T-score metrics. As SEMs were small (0.04–0.12), estimates of normative scores were precise.

Table 5 presents normative scores within age and race/ethnicity groups on the T-score metric. For the SF-12 PCS, scores decreased monotonically with age group, regardless of

Table 4. KDQOL-36 scale weighted means

Scale	Mean	SEM	95% Confidence Interval	Cronbach	Test-Retest Reliability ICC
Original Scoring					
SF-12 PCS ^a	37.8	0.04	37.7 to 37.9	—	0.81
SF-12 MCS ^a	50.9	0.04	50.8 to 51.0	—	0.63
KSS	73.0	0.07	72.9 to 73.1	0.91	0.88
KDQOL-36 Burden of Kidney Disease	52.8	0.12	52.6 to 53.1	0.85	0.80
KDQOL-36 Symptoms/Problems of Kidney Disease	79.0	0.07	78.9 to 79.2	0.83	0.89
KDQOL-36 Effects of Kidney Disease	74.1	0.09	74.0 to 74.3	0.85	0.82
T-scores					
KSS	50.2	0.04	50.1 to 50.2	—	—
KDQOL-36 Burden of Kidney Disease	50.2	0.04	50.1 to 50.3	—	—
KDQOL-36 Symptoms/Problems of Kidney Disease	50.1	0.04	50.0 to 50.2	—	—
KDQOL-36 Effects of Kidney Disease	50.1	0.04	50.0 to 50.2	—	—

ICC, intraclass correlation coefficient; —, not estimated.

^aScored on T-score metric with mean of 50 and SD of 10 in United States general population.

race/ethnicity, indicating worsening HRQOL with age. Overall, and for non-Hispanic white and black patients, the SF-12 MCS scores were larger (more positive) with increasing age. The PCS and MCS scores for non-Hispanic black patients were higher than other race/ethnic groups, as has been observed in other research.¹² No specific patterns in the PCS or MCS across age groups emerged for other race/ethnic groups.

Black patients had the highest (best HRQOL) KSS scores, followed by non-Hispanic white patients, Hispanic patients, and finally, Asian patients. No clear pattern in KSS scores emerged across age groups. Black patients tended to have the highest and Asian patients the lowest Burden of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease scale scores. Strong patterns did not emerge across age groups for these scales.

DISCUSSION

The use and relevance of PROMs in dialysis quality monitoring and clinical research has increased in the past decade,^{1,2,15,28,29} and the KDQOL-36 is among the most common PROMs used with patients on dialysis. Despite its widespread use, efforts to aid interpretation of the KDQOL-36 scores are needed. This study advances the current knowledge of how to interpret scales of the KDQOL-36 by providing reference values for the national United States dialysis population, overall and for subgroups. In addition, a new KDQOL-36 composite, the KSS summary scale, demonstrated excellent reliability and allows ease of implementation of this PROM into clinical practice and research studies.

Normative values referenced to the national United States dialysis population, as represented in the USRDS, were calculated in both the original scoring metric of each KDQOL-36 scale and using a T-score metric that set each scale to a mean of 50 with an SD of 10. These normative values make possible a comparison of groups and individual patients to a national “average” patient. As the value of 50 is referenced to the United

States national dialysis population average on the T-score metric, an individual’s or patient group’s scores can be referenced easily in terms of SD. For example, a group of patients on dialysis with a mean T-score of 40 would be 1 SD worse than the national average. T-scores are also used by mature HRQOL measurement systems, including the Medical Outcomes Study SF-36^{30,31} and PROMIS.^{13,32}

Use of normative values have several important applications for dialysis clinics and patients. The National Quality Forum and expert reviews have listed appropriate normative values as critical for interpretation of PROMs used as quality performance measures.^{33,34} A dialysis clinic’s medical director may want to compare the average score for his or her clinic to the national average. This type of group-level comparison may help facilitate quality improvement initiatives. For example, a dialysis center may set a quality goal to exceed the clinic-level national average for HRQOL, and these normative values could assist in that project. On the individual level, similar comparisons can be made. Values below the national average may signal a particularly sick or disabled patient who could benefit from a modified dialysis dose or from a referral (*e.g.*, in the case of low scores). These normative scores could also be utilized to enhance the dialysis patient–provider conversation. Because the CMS requires patient-specific care plans on the basis of HRQOL scores,³⁵ the KDQOL-36 normative scores could assist in explaining to patients how they compare with the national average. Currently, without additional interpretation from programs like KDQOL Complete, patients on dialysis and providers may not have the appropriate context to interpret HRQOL scale scores, including the KDQOL-36 scales, and comparison to national norms can give the scores context for what is “normal.”

Because scores have been calculated by age and racial/ethnic group, these comparisons can be better targeted to a patient’s demographic characteristics. Case-mix–adjusted norms, like race/ethnicity-adjusted scores, can help to give additional context to normative scores by providing reference values for the adjusted group. For example, clinicians may choose to compare their Asian, 49-year-old patient’s score on the KDQOL-36

Table 5. KDQOL-36 scale normative scores in T-score metric for race/ethnicity and age groups

Scale	All Races Mean (95% CI)	Hispanic Mean (95% CI)	Asian Mean (95% CI)	Black Mean (95% CI)	White Mean (95% CI)
SF-12 PCS					
All ages	37.8 (37.7 to 37.9)	38.8 (38.6 to 38.9)	38.3 (37.9 to 38.6)	39.2 (39.1 to 39.4)	36.0 (35.9 to 36.1)
18–29 yr	43.0 (42.4 to 43.5)	45.0 (44.2 to 45.8)	45.2 (43.2 to 47.2)	42.5 (41.6 to 43.5)	41.7 (40.6 to 42.7)
30–44 yr	40.7 (40.4 to 40.9)	42.0 (41.6 to 42.5)	41.8 (40.7 to 42.9)	41.2 (40.8 to 41.6)	38.8 (38.2 to 39.4)
45–59 yr	38.3 (38.1 to 38.4)	38.6 (38.3 to 38.9)	39.0 (38.3 to 39.7)	39.7 (39.4 to 40.0)	36.2 (35.9 to 36.5)
60–74 yr	37.2 (37.1 to 37.3)	38.3 (38.0 to 38.6)	38.1 (37.5 to 38.6)	38.5 (38.3 to 38.8)	35.6 (35.4 to 35.8)
≥75 yr	36.1 (35.9 to 36.3)	36.6 (36.1 to 37.1)	35.9 (35.2 to 36.6)	37.4 (37.0 to 37.8)	35.4 (35.1 to 35.6)
SF-12 MCS					
All ages	50.9 (50.8 to 51.0)	49.2 (49.0 to 49.4)	49.7 (49.4 to 50.1)	51.4 (51.2 to 51.5)	51.3 (51.2 to 51.4)
18–29 yr	49.8 (49.3 to 50.4)	49.5 (48.7 to 50.4)	49.4 (46.7 to 52.1)	50.6 (49.7 to 51.6)	49.2 (48.1 to 50.4)
30–44 yr	49.9 (49.6 to 50.1)	49.5 (49.0 to 50.0)	50.4 (49.3 to 51.5)	50.2 (49.8 to 50.6)	49.5 (49.0 to 50.1)
45–59 yr	50.1 (49.9 to 50.3)	48.6 (48.3 to 48.9)	49.5 (48.8 to 50.1)	51.0 (50.8 to 51.2)	50.0 (49.7 to 50.3)
60–74 yr	51.2 (51.0 to 51.3)	49.2 (48.9 to 49.5)	50.0 (49.5 to 50.6)	51.7 (51.5 to 52.0)	51.6 (51.4 to 51.8)
≥75 yr	52.1 (51.9 to 52.3)	50.0 (49.5 to 50.5)	49.4 (48.6 to 50.2)	52.6 (52.3 to 53.0)	52.6 (52.4 to 52.9)
KSS					
All ages	50.2 (50.1 to 50.2)	49.2 (49.0 to 49.3)	47.8 (47.4 to 48.2)	51.2 (51.0 to 51.3)	50.0 (49.8 to 50.9)
18–29 yr	50.2 (49.7 to 50.8)	50.5 (49.7 to 51.3)	51.0 (48.7 to 53.4)	50.3 (49.3 to 51.3)	49.9 (48.8 to 50.9)
30–44 yr	49.2 (49.0 to 49.5)	49.2 (48.8 to 49.7)	48.9 (47.8 to 50.0)	49.6 (49.2 to 50.0)	48.8 (48.3 to 49.4)
45–59 yr	49.2 (49.1 to 49.4)	48.2 (47.9 to 48.5)	47.4 (46.7 to 48.0)	50.5 (50.3 to 50.8)	48.4 (48.1 to 48.7)
60–74 yr	50.4 (50.3 to 50.5)	49.4 (49.1 to 49.7)	48.0 (47.5 to 48.6)	51.7 (51.5 to 51.9)	50.0 (49.8 to 50.2)
≥75 yr	51.5 (51.3 to 51.6)	50.2 (49.7 to 50.7)	47.2 (46.4 to 48.0)	53.2 (52.8 to 53.5)	51.5 (51.3 to 51.7)
KDQOL-36 Burden of Kidney Disease					
All ages	50.2 (50.1 to 50.3)	48.9 (48.8 to 49.1)	47.3 (47.0 to 47.6)	51.6 (51.5 to 51.8)	49.8 (49.7 to 49.9)
18–29 yr	49.5 (48.9 to 50.0)	49.6 (48.8 to 50.4)	49.1 (46.9 to 51.3)	49.6 (48.7 to 50.6)	49.3 (48.2 to 50.3)
30–44 yr	49.4 (49.2 to 49.7)	49.0 (48.6 to 49.5)	48.2 (47.1 to 49.3)	50.0 (49.6 to 50.4)	49.0 (48.5 to 49.5)
45–59 yr	49.7 (49.6 to 49.9)	48.1 (47.8 to 48.4)	47.1 (46.5 to 47.8)	51.3 (51.1 to 51.6)	48.9 (48.7 to 49.2)
60–74 yr	50.5 (50.3 to 50.6)	49.1 (48.8 to 49.4)	47.3 (46.8 to 47.9)	52.1 (51.8 to 52.3)	50.1 (49.9 to 50.3)
≥75 yr	50.7 (50.5 to 50.9)	50.1 (49.6 to 50.6)	46.9 (46.2 to 47.6)	53.0 (52.6 to 53.3)	50.3 (50.1 to 50.6)
KDQOL-36 Symptoms and Problems of Kidney Disease					
All ages	50.1 (50.0 to 50.2)	49.7 (49.5 to 49.9)	48.8 (48.4 to 49.2)	50.5 (50.4 to 50.6)	50.1 (49.9 to 50.2)

Table 5. Continued

Scale	All Races Mean (95% CI)	Hispanic Mean (95% CI)	Asian Mean (95% CI)	Black Mean (95% CI)	White Mean (95% CI)
18–29 yr	51.0 (50.4 to 51.5)	51.6 (50.7 to 52.5)	52.8 (50.2 to 55.4)	50.5 (49.4 to 51.6)	50.8 (49.7 to 51.8)
30–44 yr	49.7 (49.5 to 50.0)	50.6 (50.1 to 51.1)	50.1 (48.9 to 51.2)	49.6 (49.2 to 50.0)	49.4 (48.9 to 49.9)
45–59 yr	49.4 (49.2 to 49.5)	49.2 (48.9 to 49.5)	48.6 (47.9 to 49.4)	50.0 (49.8 to 50.3)	48.8 (48.5 to 49.1)
60–74 yr	50.2 (50.0 to 50.3)	49.7 (49.4 to 50.0)	49.0 (48.4 to 49.6)	50.8 (50.6 to 51.1)	50.0 (49.8 to 50.2)
≥75 yr	51.0 (50.8 to 51.2)	49.8 (49.3 to 50.3)	47.9 (47.1 to 48.7)	51.6 (51.2 to 52.0)	51.4 (51.2 to 51.6)
KDQOL-36 Effects of Kidney Disease					
All ages	50.1 (50.0 to 50.2)	49.0 (48.9 to 49.2)	47.9 (47.6 to 48.3)	51.2 (51.0 to 51.3)	50.0 (49.8 to 50.1)
18–29 yr	49.8 (49.2 to 50.4)	49.7 (48.8 to 50.5)	49.9 (47.6 to 52.1)	50.3 (49.3 to 51.3)	49.3 (48.3 to 50.4)
30–44 yr	48.9 (48.6 to 49.1)	48.3 (47.8 to 48.8)	48.5 (47.4 to 49.6)	49.3 (48.9 to 49.7)	48.6 (48.1 to 49.2)
45–59 yr	49.0 (48.8 to 49.1)	47.8 (47.5 to 48.1)	47.2 (46.5 to 47.9)	50.3 (50.0 to 50.5)	48.2 (47.9 to 48.5)
60–74 yr	50.4 (50.3 to 50.5)	49.6 (49.3 to 49.9)	48.3 (47.7 to 48.3)	50.8 (50.6 to 51.1)	49.9 (49.7 to 50.1)
≥75 yr	51.9 (51.7 to 52.1)	50.6 (50.1 to 51.0)	47.8 (47.0 to 48.6)	53.7 (53.4 to 54.1)	51.8 (51.6 to 52.0)

Burdens of Kidney Disease scale to the adjusted normative T-score of 47.1 instead of the general dialysis population mean of 50. Therefore, in some cases, case-mix–adjusted normative scores may be most appropriate. However, use of case-mix–adjusted scores may also mask disparities in HRQOL between subgroups by normalizing a lower or higher level of HRQOL in comparison to a reference subgroup. When considered at the clinic level, some have argued that case-mix adjustment can set a lower performance bar for clinics serving many at-risk patients or incentivize serving patients at lower risk to improve scores.^{36,37} For this reason, users should remain cautious when comparing clinic-level scores to national norms when these scores are stratified by race/ethnicity.

In addition to the KDQOL-36 scales' normative values, this study found evidence for a new composite of the KDQOL-36's Burdens of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease items: the KSS. There are some instances where a single, combined score may be preferred to three separate kidney-specific scales, especially in research applications. For instance, in clinical trials, a composite outcome is often preferable to multiple separate outcomes because of higher reliability, statistical power, and reducing multiple comparisons.^{38,39} The continued use of the Burdens, Symptoms and Problems, or Effects of Kidney Disease scales is still recommended when the measurement of more specific outcomes is desired.

Other investigators have scored the KDCS from the larger KDQOL Short Form instrument by combining all of its

11 scales.^{7,14,15,40} Although the KDCS has been shown to be predictive of clinical outcomes, its psychometric properties are unknown. Because the KDCS combines information from a heterogeneous set of scales (which include measures of cognitive function, satisfaction with care, *etc.*), further research should be conducted to determine if it offers similar reliability to that of the KSS. Although a single-score summary representing the diverse constructs represented in the KDQOL short form subscales could be useful, it is questionable whether a coherent, unidimensional construct could be represented by such heterogeneous content.

This study has limitations that should be considered when interpreting its results. First, research was not the original objective of the data used in this study, which were collected as part of the KDQOL Complete program to monitor the physical and mental health of patients on dialysis. Because of the sample's size, we were able to draw a subsample representative of the national dialysis population. However, a study designed for this purpose that drew a probability sample would have been able to cover more specific demographic and clinical categories in its representation of the national dialysis population. Second, because of the deidentified nature of the MEI data, we could not make patient-level links to the USRDS dataset to enable statistical comparisons between the cohorts. Third, although the dataset we used was rich in terms of sample size and patient diversity, few clinical or other variables were available to examine the KSS's validity. We plan to study the KSS's construct validity in future research. Finally, there

are several steps that still need to be taken to optimize the interpretation of the KDQOL-36 that could not be covered by this study. Perhaps most important among these is the estimation of minimally important differences, or the smallest difference in scores deemed important to a patient in terms of a change in his or her health, or that would lead to a change in treatment decisions.⁴¹ Estimation of minimally important differences for the KDQOL-36 scales, including the new KSS, will help guide clinical decision-making as well as estimate effect sizes for group differences to support planning of future research studies.

In conclusion, this study has advanced the use of the KDQOL-36 in two ways. We created a new kidney-specific HRQOL composite from the KDQOL-36 Burdens of Kidney Disease, Symptoms and Problems of Kidney Disease, and Effects of Kidney Disease items. The KSS composite captures kidney-specific HRQOL in a single score. In addition, we calculated normative values referenced to the national United States dialysis population that can be used to compare with group and individual scores on the KDQOL-36 scales, giving context against the national average. These developments may serve to improve the interpretation and implementation of the KDQOL-36 instrument for patients and providers.

ACKNOWLEDGMENTS

J.D.P. and R.D.H. conceived of the study, designed the study, analyzed study data, and wrote the manuscript. D.N., K.K., and D.R.S. conceived of the study and wrote the manuscript.

There was no direct financial support for the research reported in this manuscript.

DISCLOSURES

None.

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at <http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2018100994/-/DCSupplemental>.

Supplemental Appendix 1. EFA results.

Supplemental Appendix 2. CFA model results.

Supplemental Appendix 3. Reliability of KDQOL-36 scales by age and race/ethnicity groups.

REFERENCES

1. Peipert JD, Hays RD: Methodological considerations in using patient reported measures in dialysis clinics. *J Patient Rep Outcomes* 1: 11, 2017
2. Peipert JD, Hays RD: Using patient-reported measures in dialysis clinics. *Clin J Am Soc Nephrol* 12: 1889–1891, 2017
3. Food and Drug Administration: Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims. Rockville, MD, US Department of Health and Human Services, 2009, pp 2
4. Calvert M, Kyte D, Mercieca-Bebber R, Slade A, Chan AW, King MT, et al.: the SPIRIT-PRO Group: Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: The SPIRIT-PRO extension. *JAMA* 319: 483–494, 2018
5. Urquhart-Secord R, Craig JC, Hemmelgarn B, Tam-Tham H, Manns B, Howell M, et al.: Patient and caregiver priorities for outcomes in hemodialysis: An international nominal group technique study. *Am J Kidney Dis* 68: 444–454, 2016
6. Lowrie EG, Curtin RB, LePain N, Schatell D: Medical Outcomes Study Short Form-36: A consistent and powerful predictor of morbidity and mortality in dialysis patients. *Am J Kidney Dis* 41: 1286–1292, 2003
7. Mapes DL, Lopes AA, Satayathum S, McCullough KP, Goodkin DA, Locatelli F, et al.: 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, 2003
8. Molnar-Varga M, Molnar MZ, Szeifert L, Kovacs AZ, Kelemen A, Becze A, et al.: Health-related quality of life and clinical outcomes in kidney transplant recipients. *Am J Kidney Dis* 58: 444–452, 2011
9. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Gheorghe A, Keeley T, et al.: Measurement properties of patient-reported outcome measures (PROMs) used in adult patients with chronic kidney disease: A systematic review. *PLoS One* 12: e0179733, 2017
10. Hays RD, Kallich JD, Mapes DL, Coons SJ, Carter WB: Development of the Kidney Disease Quality Of Life (KDQOL) instrument. *Qual Life Res* 3: 329–338, 1994
11. Peipert JD, Bentler PM, Klicko K, Hays RD: Psychometric properties of the Kidney Disease Quality Of Life 36-item short-form survey (KDQOL-36) in the United States. *Am J Kidney Dis* 71: 461–468, 2018
12. Peipert JD, Bentler P, Klicko K, Hays RD: 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™-36). *Qual Life Res* 27: 2699–2707, 2018
13. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al.: Representativeness of the patient-reported outcomes measurement information system internet panel. *J Clin Epidemiol* 63: 1169–1178, 2010
14. Mapes DL, Bragg-Gresham JL, Bommer J, Fukuhara S, McKeivitt P, Wikström B, et al.: Health-related quality of life in the Dialysis Outcomes and Practice Patterns Study (DOPPS). *Am J Kidney Dis* 44[Suppl 2]: 54–60, 2004
15. Fukuhara S, Lopes AA, Bragg-Gresham JL, Kurokawa K, Mapes DL, Akizawa T, et al.: Worldwide Dialysis Outcomes and Practice Patterns Study: Health-related quality of life among dialysis patients on three continents: The dialysis outcomes and practice patterns study. *Kidney Int* 64: 1903–1910, 2003
16. United States Renal Data System: 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, 2017
17. Slocum-Gori SL, Zumbo BD: Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Soc Indic Res* 102: 443–461, 2011
18. Schmid J, Leiman JM: The development of hierarchical factor solutions. *Psychometrika* 22: 53–61, 1957
19. Revelle W: *Using R and the Psych Package to Find ω*, Evanston, IL, Northwestern University, 2018
20. Reise SP, Scheines R, Widaman KF, Haviland MG: Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educ Psychol Meas* 73: 5–26, 2013
21. Rosseel Y: lavaan: An R package for structural equation modeling. *J Stat Softw* 48: 1–36, 2012
22. Reise SP, Kim DS, Mansolf M, Widaman KF: Is the bifactor model a better model or is it just better at modeling implausible responses?

- Application of iteratively reweighted least squares to the rosenberg self-esteem scale. *Multivariate Behav Res* 51: 818–838, 2016
23. Hu LT, Bentler PM: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model* 6: 1–55, 1999
 24. Cohen J: *Statistical Power Analysis for the Behavioral Sciences*, New York, Academic Press, 1988
 25. Nunnally JC: *Psychometric Theory*, New York, McGraw-Hill, 1978
 26. Fleiss JL, Levin B, Paik MC: *Statistical Methods for Rates and Proportions*, Hoboken, NJ, John Wiley & Sons, Inc., 2004, pp 761–768
 27. SAS Institute: *I: Base SAS[®] 9.4 Procedures Guide: Statistical Procedures*, Cary, NC, SAS Institute, Inc., 2013
 28. Cavanaugh KL: Patient experience assessment is a requisite for quality evaluation: A discussion of the In-Center Hemodialysis Consumer Assessment of Health Care Providers and Systems (ICH CAHPS) survey. *Semin Dial* 29: 135–143, 2016
 29. Galla JH; The Renal Physicians Association and the American Society of Nephrology: Clinical practice guideline on shared decision-making in the appropriate initiation of and withdrawal from dialysis. *J Am Soc Nephrol* 11: 1340–1342, 2000
 30. Ware JE Jr., Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30: 473–483, 1992
 31. Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, et al.: US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. *Med Care* 45: 1162–1170, 2007
 32. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, Cella D: Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol* 63: 1195–1204, 2010
 33. National Quality Forum: Patient Reported Outcomes (PROs) in Performance Measurement. Washington, DC, National Quality Forum, 2013
 34. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, et al.: *Patient-Reported Outcomes in Performance Measurement*, Research Triangle Park, NC, Research Triangle Institute, 2015
 35. Centers for Medicare & Medicaid Services; Department of Health and Human Services: Medicare and Medicaid Programs; Conditions for Coverage for End-Stage Renal Disease Facilities; Final Rule. Baltimore, MD, Federal Register, 2008
 36. Romano PS: Should health plan quality measures be adjusted for case mix? *Med Care* 38: 977–980, 2000
 37. Zaslavsky AM, Hochheimer JN, Schneider EC, Cleary PD, Seidman JJ, McGlynn EA, et al.: Impact of sociodemographic case mix on the HEDIS measures of health plan quality. *Med Care* 38: 981–992, 2000
 38. Cordoba G, Schwartz L, Woloshin S, Bae H, Gøtzsche PC: Definition, reporting, and interpretation of composite outcomes in clinical trials: Systematic review. *BMJ* 341: c3920, 2010
 39. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C: Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA* 289: 2554–2559, 2003
 40. Lopes AA, Bragg-Gresham JL, Satayathum S, McCullough K, Pifer T, Goodkin DA, et al.: Worldwide Dialysis Outcomes and Practice Patterns Study Committee: 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). *Am J Kidney Dis* 41: 605–615, 2003
 41. Revicki D, Hays RD, Cella D, Sloan J: Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 61: 102–109, 2008

See related editorial, “Patient-Reported Outcomes: Toward Better Measurement of Patient-Centered Care in CKD,” on pages 523–525.