

# Improving CKD-Specific Patient-Reported Measures of Health-Related Quality of Life

John E. Ware Jr. <sup>1,2</sup>, Michelle M. Richardson,<sup>3,4</sup> Klemens B. Meyer,<sup>3,4</sup> and Barbara Gandek<sup>1,2</sup>

<sup>1</sup>John Ware Research Group, Outcomes Measurement Department, Watertown, Massachusetts; <sup>2</sup>Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts; <sup>3</sup>William B. Schwartz Division of Nephrology, Department of Medicine, Tufts Medical Center, Boston, Massachusetts; and <sup>4</sup>Dialysis Clinic, Inc., Outcomes Monitoring Program, Boston, Massachusetts

## ABSTRACT

**Background** Patient-reported outcome measures that are more practical and clinically useful are needed for patients with CKD. We compared a new CKD-specific quality-of-life impact scale (CKD-QOL) with currently used measures.

**Methods** Patients ( $n=485$ ) in different treatment groups (nondialysis stages 3–5, on dialysis, or post-transplant) completed the kidney-specific CKD-QOL and Kidney Disease Quality of Life-36 (KDQOL-36) forms and the generic SF-12 Health Survey at baseline and 3 months. New items summarizing quality of life (QOL) impact attributed to CKD across six QOL domains yielded single impact scores from a six-item static (fixed-length) form and from computerized adaptive tests (CATs) with three to six items. Validity tests compared the CKD-QOL, KDQOL-36 (Burden, Effects, and Symptoms/Problems subscales), and generic SF-12 measures across groups in four tests of clinical status and clinician assessment of change (CKD-specific tests), and number of comorbidities. ANOVA was used to test for group mean differences, variances in each measure explained by groups, and relative validity (RV) in comparison with the referent KDQOL-36 Burden subscale.

**Results** KDQOL-36 and CKD-QOL measures generally discriminated better than generic SF-12v2 measures. The pattern of variances across CKD-specific tests comparing validity favored CKD-QOL two-fold over KDQOL-36. Two RV test results confirmed CKD-QOL improvements over the referent KDQOL scale. Results for static and CAT CKD-QOL forms were similar. SF-12 Physical and KDQOL-36 Symptoms scores worsened with increasing comorbid condition counts.

**Conclusions** Overall, compared with the KDQOL-36, the new approach to summarizing CKD-specific QOL impact performed better across multiple tests of validity. CAT surveys were more efficient than static surveys.

*J Am Soc Nephrol* 30: 664–677, 2019. doi: <https://doi.org/10.1681/ASN.2018080814>

CKD, its comorbid conditions, and its treatment substantially burden patients' health-related quality of life (QOL). Kidney diseases, even those with diverse clinical presentations such as polycystic kidney disease and nephrotic syndrome, are associated with substantial QOL impairment.<sup>1</sup> QOL declines with increasing CKD severity: patients at earlier CKD stages experience smaller but still noteworthy QOL impairments compared with the general and hypertensive populations,<sup>2,3</sup> whereas those at CKD stage 5 who are about to

begin dialysis report physical health 3–4 SD below the general population average.

Received August 10, 2018. Accepted January 30, 2019.

Published online ahead of print. Publication date available at [www.jasn.org](http://www.jasn.org).

**Correspondence:** Dr. John E. Ware Jr., John Ware Research Group, 10 Wheeler Court, Watertown, MA 02472. Email: [john.ware@jwrginc.com](mailto:john.ware@jwrginc.com)

Copyright © 2019 by the American Society of Nephrology

Generic QOL measures have the advantage of enabling comparisons of disease burden across CKD and other conditions, whereas disease-specific measures have the advantages of greater validity, including responsiveness, for a specific condition.<sup>4</sup> The most common generic QOL tools used in CKD are the Short Form-36 Health Survey<sup>5–8</sup> and its 12-item subset, the Short Form-12 Health Survey.<sup>9</sup> The disease-specific tool most commonly used in kidney disease, the Kidney Disease Quality of Life 36-Item Short Form Survey (KDQOL-36), measures both CKD-specific and generic QOL domains.<sup>10–12</sup>

The KDQOL-36 augments the Short Form-12 generic core with 24 items used to score three kidney-specific scales: Burden of Kidney Disease (four items), Symptoms/Problems of Kidney Disease (12 items), and Effects of Kidney Disease (eight items).

The KDQOL-36 is widely used and has been reported to show satisfactory psychometric properties.<sup>12</sup> Despite its widespread use, the KDQOL-36 has disadvantages.<sup>13</sup> In an attempt to be short to reduce overall respondent burden, important CKD-specific domains are omitted. Short forms also may yield scores that are too imprecise for use in individual patient clinical care. In addition, static surveys such as the KDQOL-36 administer the same questions to everyone, including some questions that may be irrelevant to a specific individual. The range of reliable measurement is restricted, limiting the ability to detect score change associated with changes in disease severity or with treatment effects,<sup>14</sup> leading a recent Technical Expert Panel convened on behalf of the Centers for Medicare and Medicaid Services to question whether the KDQOL-36 is an effective patient-reported outcome measure for comparing facility performance.<sup>15</sup>

Perhaps the most important limitation is that the KDQOL and its iterations were developed decades ago and have not kept pace with conceptual and methodological advances in patient-reported outcome measurement. These improvements include additional and improved QOL descriptions not as well represented in CKD-specific KDQOL-36 scales, but known with both general and kidney disease attributions to be affected by CKD.<sup>5,16</sup> Psychometric progress using item response theory (IRT) has improved scale construction and scoring<sup>17–20</sup> and enabled computerized adaptive tests (CATs) that reduce patient burden and/or increase score precision by matching items better to score levels.<sup>21</sup> Finally, evidence that items across QOL domains, all with attributions to one disease, are sufficiently unidimensional to construct a single summary disease impact score has been independently reported across multiple chronic conditions (MCC).<sup>22–24</sup>

This study evaluated an improved and briefer approach to measuring QOL impact attributed specifically to CKD with a single summary score. As illustrated in Figure 1, the new approach was on the basis of published distinctions among categories of outcomes throughout a conceptual continuum that ranges from the most CKD-specific clinical parameters to generic QOL.<sup>4,25</sup>

On the left of the figure, in box 1, are CKD-specific clinical parameters (*e.g.*, eGFR), followed in box 2 by patient-reported

### Significance Statement

Monitoring patient-reported outcomes to capture CKD's effects on health-related quality of life (QOL) is important for population health and individual care. Current measures such as the Kidney Disease Quality of Life-36 (KDQOL-36) do not incorporate some proven measurement advances, and measures incorporating such advances are rarely compared with current methods. The authors evaluated the validity of a new approach to CKD-specific QOL measurement that comprehensively represents CKD-specific QOL, yields a single summary QOL impact score, and generally requires only 1 minute. Across CKD stages 3–5, dialysis, and transplant patients, results favored the new approach over the KDQOL-36 in comparisons of validity, including responsiveness (sensitivity to clinical change), across multiple clinical tests. Computerized adaptive test versions of the new approach were more efficient than static versions.

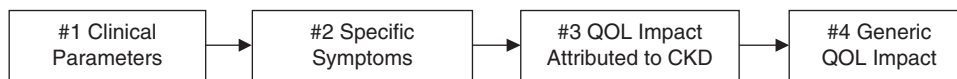
frequency of CKD-specific symptoms, such as cramps, then in box 3 by the QOL impact (*e.g.*, on relationships with family and friends) attributed specifically to CKD, and on the far right, in box 4, by generic measures of functional health and wellbeing. Measures in boxes 3 and 4 are both considered QOL impact to the extent that operational definitions go beyond clinical status and symptom frequency to reflect what people can do (functioning) in everyday life, how they feel (ill- and wellbeing), and how they evaluate those states qualitatively. A major difference between them is their attribution, respectively, to CKD versus health in general. Boxes 2 and 3 are more useful in detecting changes that are specific to CKD whereas box 4 is necessary to compare QOL burden and outcomes across various diseases and their treatments.<sup>4</sup>

We report results from tests of a new approach to the measurement of CKD-specific QOL impact. This approach reduces respondent burden by measuring CKD-specific QOL with either a six-item static survey or a three- to six-item CAT. We evaluate the new tool's psychometric properties in a clinical setting and test its validity, including responsiveness, in head-to-head comparisons with three CKD-specific KDQOL-36 scales<sup>10,12</sup> and with two generic Short Form-12 version 2 Health Survey (SF-12v2) summary measures.

## METHODS

### Study Design and Participants

This was a cross-sectional multisite survey of a convenience sample of patients with CKD, with a follow-up survey administered at one site, to compare the validity of CKD-specific and generic measures. The study enrolled a sample of patients with CKD (nondialysis stages 3–5, on dialysis, and post-transplant not requiring dialysis; *n*=485) during routine outpatient appointments at nephrology practices affiliated with Tufts Medical Center and Brigham and Woman's Hospital and at ten Dialysis Clinic, Inc. sites in Boston (*n*=4), Jacksonville, Florida



**Figure 1.** New approach based on distinctions throughout a continuum ranging from clinical to CKD-specific and generic QOL outcomes.

(*n*=2), Farmington, Connecticut (*n*=1) and central New York State (*n*=3). Patients age ≥18 years who spoke and read English were eligible, and patients judged by treating

nephrologists to have cognitive impairments were excluded. Patients self-administered all surveys using tablet computers, without difficulty. Three months after baseline, Boston-area

**Table 1.** Summary of information for CKD-specific and generic measures (*n*=485)

Measure/Scoring	<i>k</i>	% Missing	Mean (SD)	Reliability <sup>a</sup>		Interpretation of Lowest Score	Interpretation of Highest Score
				Internal Consistency	Test-Retest		
<b>KDQOL-36</b>							
Burden (+)	4	1.0	61.0 (29.5)	0.88	0.79	Kidney disease interferes with life, frustrating, feels like burden on family	Kidney disease does not interfere with life, is not frustrating, does not feel like burden on family
Effects (+)	8	1.4	72.2 (22.4)	0.86	0.86	Extremely bothered by eight effects of kidney disease on daily life	Not at all bothered by eight effects of kidney disease on daily life (see Figure 2)
Symptoms/Problems (+)	12	1.2	75.8 (17.0)	0.86	0.85	Extremely bothered by 12 symptoms	Not at all bothered by 12 symptoms (see Figure 2)
<b>CKD-QOL</b>							
CKD-QOL-6 (-)	6	0.2	47.6 (9.3)	0.94	0.89	No impact on role, social, emotional, cognitive functioning, fatigue, or quality of life because of kidney disease	Severe and frequent impact on role, social, emotional, cognitive functioning, fatigue, quality of life because of kidney disease
CKD-QOL-CAT (-)	6	0.0	47.1 (10.2)	0.93 <sup>b</sup>	0.89	(Same as above)	(Same as above)
3- or 6-item CAT (-)	3-6	0.0	47.4 (10.0)	0.90 <sup>b</sup>	0.88	(Same as above)	(Same as above)
<b>SF-12v2</b>							
PCS (+)	12	3.7	40.4 (11.1)	0.89	0.87	Limited physical, work, or daily activities; severe pain; low energy; health poor	No limits in physical, work, or daily activities; no pain; high energy; health excellent
MCS (+)	12	3.7	50.1 (10.4)	0.84	0.84	Frequent emotional distress and low energy; limited in everyday activities because of emotional problems	Frequent positive affect, not at all limited in everyday activities because of emotional problems

Scoring: (+) higher score is better, (-) lower score is better. *k*, number of items.

<sup>a</sup>Reliability: internal-consistency reliability for scales estimated with Cronbach  $\alpha$ . Test-retest reliability estimated with intraclass correlation; sample size is *n*=65 for KDQOL-36 and CKD-QOL measures, and *n*=61 for SF-12v2 PCS/MCS. Reliability for PCS/MCS estimated on the basis of reliability of eight SF-12v2 scales, scale covariances, and scale weights used to estimate each component.<sup>48</sup>

<sup>b</sup>Mean reliability for CKD-QOL-CAT and three or six-item CAT on the basis of SEM for each patient score. Percentage of patients (for three- or six-item CKD-QOL-CAT) with reliability ≥0.90 is 78%/66%, reliability ≥0.80 is 82%/73%, and reliability ≥0.70 is 100%/100%.<sup>49</sup>

Domains and Abbreviated Item Content for SF-12v2, CKD-QOL-6 and KDQOL-36 Items†	Generic	CKD-Specific			
	SF-12v2	CKD-QOL-6	KDQOL		
			Burden	Effects	Symptoms
<b>Physical functioning</b>					
Limited in moderate activities, climbing stairs	•				
<b>Role functioning</b>					
Limited in kind of work/activities; accomplished less; did work/activities less carefully	•				
Kidney disease restricted usual daily activities(a)		•			
Kidney disease bothers ability to work around house				•	
<b>Pain (bodily)</b>					
Pain interfered with normal work including housework	•				
Bothered by soreness in muscles					•
Bothered by chest pain					•
<b>General health evaluation</b>					
Health is excellent/poor	•				
<b>Energy/fatigue</b>					
Amount of time had a lot of energy	•				
Kidney disease left too tired for work, daily activities(c)		•			
Bothered by being washed out or drained					•
<b>Social functioning/burden</b>					
Health interfered with social activities	•				
Kidney disease kept from enjoying social activities(b)		•			
Feel like a burden on family			•		
<b>Mental health (psychological distress/well-being)</b>					
Felt downhearted, depressed; felt calm, peaceful	•				
Fed up or frustrated because of kidney disease(c)		•			
Feel frustrated dealing with kidney disease			•		
Bothered by stress/worries caused by kidney disease				•	
<b>Life interference/quality of life</b>					
Kidney disease limited usual activities, enjoyment of everyday life (d)		•			
Kidney disease interferes too much with life			•		
Too much time spent dealing with kidney disease			•		
Kidney disease bothers ability to travel				•	
Other effects of kidney disease on daily life*				•	
<b>Cognitive functioning</b>					
Kidney disease limited ability to concentrate on work or daily activities(c)		•			
<b>Other symptoms**</b>					•

† Columns distinguish SF-12v2, CKD-QOL-6 and KDQOL items.

\* Other KDQOL Effects: Bothered by fluid restriction, dietary restriction, dependent on doctors/staff, sex life, personal appearance.

\*\* Other KDQOL Symptoms: Bothered by cramps, itchy skin, dry skin, shortness of breath, faintness/dizziness, lack of appetite, numbness in hands/feet, nausea/upset stomach, problems with access or catheter site.

KDQOL Burden responses: Definitely true, mostly true, don't know, mostly false, definitely false.

KDQOL Effects and Symptoms responses: Not at all bothered, somewhat bothered, moderately bothered, very much bothered, extremely bothered.

CKD-QOL-6 responses (a) Not at all, very little, somewhat, quite a lot, could not do activities; (b) None of the time, a little of the time, some of the time, most of the time, all of the time; (c) Never, almost never, sometimes, very often, always; (d) Not at all, a little, moderately, quite a lot, extremely.

**Figure 2.** Comparisons of QOL domain content reveal overlap and gaps in generic and CKD-specific measures.

patients completed a follow-up survey to assess responsiveness. Clinician-investigators abstracted baseline data, including dialysis and transplant status, eGFR (patients not on dialysis), comorbid conditions, and medical history, and assessed CKD clinical status at 3 months to determine better, same, or worse clinical status from baseline. This study was approved by the New England Institutional Review Board, Tufts Health Sciences Institutional Review Board, the Partners Human Research Committee, and underwent administrative review by DCI's Administrative Review Office.

## Measures

The measures evaluated included three CKD-specific scales from the widely used KDQOL-36, a new CKD-specific QOL impact scale (see below), and the generic SF-12v2 Health Survey.

### KDQOL-36

The KDQOL-36,<sup>12</sup> derived from the longer KDQOL-SF instrument,<sup>10</sup> includes 24 items in three CKD-specific scales (Burden of Kidney Disease, Effects of Kidney Disease, and

Symptoms/Problems) (see Table 1); all were scored positively (higher score indicating better QOL) on a 0–100 scale using developer-recommended scoring (available at [https://www.rand.org/health/surveys\\_tools/kdqol.html](https://www.rand.org/health/surveys_tools/kdqol.html)).

#### CKD-QOL Forms

In contrast to the KDQOL-36, which spreads QOL-related item content across CKD-specific Burden, Effects, and Symptoms/Problems scales, the new approach aggregates all QOL items, with attribution specifically to kidney disease (Figure 1, box 3), into a single measure of QOL impact. The Supplemental Appendix summarizes how, in a precursor study conducted before this validation study, a new 34-item CKD-specific quality-of-life (CKD-QOL) impact item bank was constructed using data from an independent developmental sample of 1236 individuals with CKD representing CKD stages 3–5 not treated by dialysis, treated by dialysis, and treated by kidney transplant. CKD-QOL items measured the impact of kidney disease across multiple functional health and wellbeing domains (role and social functioning, fatigue, psychologic distress, cognitive functioning) and overall quality of life (see Figure 2). CKD-QOL items were derived in a multistage process, including review of existing questionnaires such as the CHOICE Study Health Experience Questionnaire,<sup>26</sup> four focus groups ( $n=40$ ) of patients not on dialysis with CKD stages 3–5 and patients on dialysis, recommendations from a clinical advisory board,<sup>27</sup> and cognitive testing with patients with CKD. Although most items were derived from item content previously tested in other conditions,<sup>17,28–30</sup> new items were added to represent descriptions of QOL impact reported by patients and/or recommended by clinicians.<sup>27</sup> In contrast to generic health surveys, all items were asked with kidney-specific attribution (for example, “In the past 4 weeks, how much did your kidney disease limit your usual activities or enjoyment of everyday life?” with five responses ranging from “not at all” to “extremely”). Extensive psychometric evaluation of data from the developmental sample including factor analysis confirmed that items were sufficiently unidimensional to compute and meaningfully interpret a single summary score. Item parameters were estimated using IRT models, as in previous research.<sup>17,31</sup>

IRT standardizes and better quantifies the units of measurement underlying item responses. In contrast to traditional “static” surveys that ask the same questions of everyone regardless of answers, CAT individualizes each assessment so that the most informative questions are matched to their level of health. The result is more efficient tools for comparing scores across individuals and time points regardless of whether items are the same. At baseline, validation study patients completed the full 34-item CKD-QOL item bank. A static subset of six items (CKD-QOL-6) selected to represent distinct QOL domains (Figure 2, Table 1) was scored. These items are appended. In addition, real-data CAT simulation methods were applied to 34-item bank responses (CKD-QOL-CAT) to identify the six most informative items for each patient.<sup>21</sup> To further reduce respondent burden as well as maximize

important information, CAT estimates also were limited to only three items for patients with little or no CKD impact (three- or six-item CAT). Norm-based scoring was used to enable direct comparisons of scores across all CKD-QOL forms, which were scored negatively (higher score indicating worse QOL impact) and linearly transformed to have a mean of 50 and an SD of 10 in the developmental sample.

#### SF-12v2 Health Survey

Generic Physical Component Summary (PCS) and Mental Component Summary (MCS) scores were calculated using developer-recommended scoring methods for the SF-12v2 Health Survey.<sup>33</sup> PCS and MCS scores were scored positively (higher score indicating better QOL) and were transformed to have a mean of 50 and SD of 10 in the United States general population.

Figure 2 compares QOL domains and item content across generic (SF-12v2) and CKD-specific (KDQOL-36 and CKD-QOL-6) scales, which differ in ways that could affect results. Although both instruments represent four or five of the seven SF-12v2 domains, in the KDQOL-36, QOL domains are distributed across Burden, Effects, and Symptoms/Problems scales. Rather than distributing this information, in CKD-QOL they are all scored in a single summary. There are also differences in how the domains are operationalized. For example, the CKD-QOL-6 Role item asks more broadly about “usual daily activities,” whereas the KDQOL-36 item asks about “work around the house.” The CKD-QOL-6 Social item asks about “social activities,” whereas the KDQOL-36 item asks about feeling “like a burden” on family. In the KDQOL-36, the same domain is measured in two separate scales (e.g., psychologic distress items in KDQOL-36 Burden and KDQOL-36 Effects). Finally, in comparison with the KDQOL-36 Effects and Symptoms ratings of how “bothered” the patient was by an effect or symptom, CKD-QOL items explicitly evaluate the impact of CKD on functioning and wellbeing. Both instruments include items measuring CKD-specific life interference/QOL, whereas CKD-QOL-6 also adds cognitive functioning.

#### Statistical Analyses

Data completeness, descriptive statistics, and psychometric properties of all measures were evaluated, including replication of developmental tests of assumptions underlying scale construction and scoring. Reliability for all scales was estimated using test-retest methods and also using internal consistency methods for static scales (see Table 1). Reliability of CAT scores was calculated on the basis of the SEM of each individual’s score. Reliability of 0.70 or higher, as recommended for group-level comparisons, and a minimum reliability of 0.90, as recommended for individual patients, were deemed acceptable. Product-moment correlations among the six CKD-specific and two generic measures were also estimated to evaluate construct validity. Substantial ( $r>0.40$ ) associations were hypothesized among CKD-specific measures and between generic

**Table 2.** Patient characteristics by CKD treatment group

Characteristic	Stage 3–5, n=121	Dialysis, n=253	Transplant, n=111	P Value
Sex				<0.01
Women	47.9	48.2	34.2	
Men	52.1	49.4	65.8	
Missing	0.0	2.4	0.0	
Age, yr				<0.01
18–44	11.6	26.5	19.8	
45–64	44.6	40.3	62.2	
65–74	22.3	17.8	13.5	
≥75	21.5	12.6	4.5	
Missing	0.0	2.8	0.0	
Mean (SD)	61.7 (13.8)	55.8 (16.8)	53.6 (11.4)	
Race/ethnicity				<0.01
White, non-Hispanic	86.0	41.9	82.0	
Black, non-Hispanic	6.6	45.4	11.7	
Hispanic	2.4	3.6	2.7	
Other, non-Hispanic	5.0	6.7	3.6	
Missing	0.0	2.4	0.0	
Education				<0.01
Less than high-school graduate	5.0	16.6	5.4	
High-school graduate or general education diploma	24.0	30.8	22.5	
Some college or technical school	23.1	28.8	34.2	
College graduate	45.4	19.4	37.8	
Missing	2.5	4.4	0.0	

All data are percentages except for Mean (SD) and *P* values.

and CKD-specific measures—less so for the latter. Known groups methods were used to evaluate discriminant validity across groups differing in CKD clinical treatment status, eGFR level, self-evaluated severity, and number of comorbid conditions, as defined below. One-way ANOVA compared the statistical efficiency of measures in discriminating between groups; each ANOVA *F* statistic indicated how strongly each measure separated group means and minimized within group error. To facilitate interpretations and comparisons, Cohen<sup>47</sup>  $\eta^2$ , the variance in each measure explained by group membership ( $\eta^2$ =between/total sums of squares), was estimated. For each group comparison, as in previous analyses, relative validity (RV) was also defined as the ratio of the *F* statistic for each comparator scale divided by the *F* statistic for the referent scale, which was the best-performing KDQOL-36 scale (Burden, RV=1.0). Estimates of 95% confidence intervals for RV statistics were derived using empirical bootstrap<sup>38</sup> to account for statistical significance for all comparisons except clinician assessments of change, which were precluded by small samples. For *F* tests for differences in group means and confidence interval-based comparisons of differences in RV ratios, results with a chance probability <0.05 or <0.01 are noted in the tables.

Known groups for four tests of validity for CKD-specific measures were defined as follows: (1) CKD treatment status, comprising dialysis, nondialysis stages 3–5, and post-transplant, with the dialysis group hypothesized to report greater average QOL impact than patients not treated with dialysis, and transplant recipients hypothesized to report the least average impact; (2) three eGFR groups were defined (<15, 15–29,

and 30–59 ml/min per 1.73 m<sup>2</sup>) for patients not on dialysis and were hypothesized to be ordered with patients with lower eGFR experiencing worse QOL; (3) self-rated severity groups defined from the question, “How would you rate the severity of your kidney disease symptoms in the past 4 weeks?” were hypothesized to be ordered with greater severity being associated with worse QOL; (4) to extend validity with a preliminary test of responsiveness,<sup>39</sup> groups selected to vary in self-evaluated outcomes were formed on the basis of clinician assessment of the change in CKD status (better/same/worse) between baseline and 3-month follow-up. Assessments of change were on the basis of CKD symptoms, disease severity and progression, and hospitalizations. For the above four comparisons, all CKD-specific measures were hypothesized to discriminate better than the generic SF-12v2 PCS and MCS. Within CKD-specific measures, the new CKD-QOL approach was hypothesized to discriminate better than the three KDQOL-36 scales.

Finally, comorbid condition groups, defined by a count (0, 1, 2, 3+) of conditions from a list of 20 abstracted by clinician-investigators from the medical record, were compared. In contrast to the hypotheses above for CKD-specific measures, the generic SF-12v2 PCS score was hypothesized to worsen with increasing comorbidity to a greater extent than scores on CKD-specific measures.

## RESULTS

Table 2 compares participant characteristics by treatment status: dialysis (*n*=253), CKD stages 3–5 not treated by dialysis



( $n=121$ ), and post-transplant ( $n=111$ ). As expected, the three groups differed significantly in sociodemographic characteristics; however, diversity was observed within all groups. Transplant patients tended to be younger and male; patients not on dialysis with CKD stages 3–5 were older on average, and patients on dialysis were disproportionately black.

The quality of data collected was high and associations among items and between items and scale and item bank totals corresponded well with assumptions underlying scale scoring methods, as in the developmental study (see Supplemental Appendix). In support of their reliability, test-retest estimates ranged from 0.79 to 0.89 and internal consistency estimates ranged from 0.86 to 0.94 for CKD-specific measures, and the reliability of individual CAT scores averaged 0.93 for CKD-QOL-CAT and 0.90 for the three- or six-item CAT (Table 1). The observation that all CKD-specific measures were highly intercorrelated ( $r=0.54$ – $0.94$ ) and were less highly correlated with the generic PCS ( $r=0.43$ – $0.63$ ) and MCS ( $r=0.38$ – $0.49$ ) supported their construct validity. As expected for alternate forms measuring the same construct, static and CAT CKD-QOL scores correlated highly ( $r=0.94$ ) and consistently yielded nearly equal means across the groups compared (Tables 3–6).

In the 21st percentile, the combined CKD sample generic PCS mean of 40.4 was well below the general population average,<sup>33</sup> whereas the mean MCS score was not (Table 1). On average, patients on dialysis had much (about 1 SD) worse QOL than those with CKD stages 3–5 not treated by dialysis and transplant patients, who generally had similar mean scores (Table 3). This significant pattern was observed for six CKD-specific measures and for the generic PCS scale, resulting in the second largest observed set of  $F$  and  $\eta^2$  statistics, but was not significant for the generic MCS scale. The trend in  $\eta^2$  and RV results consistently favored CKD-specific over generic measures. In tests comparing discriminant validity within the CKD-specific measures, the trends in results favored a CKD-QOL form over the KDQOL forms for both  $\eta^2$  (0.25–0.29 for CKD-QOL versus 0.11–0.22 for KDQOL) and RV (1.16–1.42 for CKD-QOL and 0.44–1.0 for KDQOL). Taking into account confidence intervals, only the CKD-QOL-CAT form (RV=1.42;  $P<0.05$ ) discriminated significantly better.

For the eGFR validity comparison, the three groups were ordered as hypothesized and were differentiated significantly by all CKD-specific measures, except KDQOL-36 Symptoms/Problems. However, variances explained by groups and  $F$  statistics were among the weakest observed across tests (Table 4). Neither of the generic PCS or MCS scales discriminated among eGFR groups ( $P>0.05$ ). In comparisons between CKD-specific measures, the ranges of variances slightly favored the three CKD-QOL forms ( $\eta^2=0.08$ – $0.10$ ) over the KDQOL forms ( $\eta^2=0.02$ – $0.08$ ). Taking confidence intervals into account, RVs did not differ significantly.

In the self-reported severity test, the four groups were ordered as hypothesized and separated substantially and

significantly (Table 5). For each of the three categories of measures, this test yielded the largest sets of variances explained (KDQOL,  $\eta^2=0.28$ – $0.33$ ; CKD-QOL,  $\eta^2=0.45$ – $0.48$ ; and PCS/MCS,  $\eta^2=0.08$ – $0.23$ ). Consistent with the nonoverlapping ranges of variances explained for referent versus comparator CKD-QOL scales, the largest CKD-QOL RV ratios, about 60%–80% better, were observed (RV=1.65–1.83; all  $P<0.05$ ).

Over half the patients (53 out of 95, 56%) with complete data for longitudinal follow-up were judged by clinicians to have the same CKD status after 3 months (Table 7). Although a change for the better was two times more likely than for the worse, samples were sufficient for preliminary estimation of  $F$  statistics. Average score changes on all CKD-specific measures were ordered as hypothesized across clinical change groups. On average, groups that were judged clinically to have gotten better improved more in QOL scores and those that were judged clinically to have gotten worse declined more in QOL scores. Group differences were significant ( $P<0.05$ ) for the two CKD-QOL-CAT ( $P<0.001$ ) and generic PCS ( $P<0.01$ ) and MCS ( $P<0.05$ ) measures. However, given the very small samples for groups that changed, RV estimates that would otherwise be considered substantial<sup>38</sup> could not be differentiated from what would be expected by chance.

As hypothesized, the generic PCS measure declined significantly ( $P<0.05$ ) in the presence of, and with increases in, the number of comorbid conditions (Table 6). Five of the six CKD-specific measures did not respond to the presence or number of comorbid conditions. The exception was KDQOL-36 Symptoms/Problems scale, which worsened ( $P<0.05$ ) in the presence of comorbid conditions.

## DISCUSSION

These results support the hypothesized advantages of a new approach to CKD-specific QOL impact measurement. The new CKD-QOL measures improved validity in head-to-head comparisons with generic SF-12v2 and CKD-specific KDQOL-36 measures. In addition to self-evaluated severity, multiple clinical comparisons included differences across treatment status groups, severity (eGFR) within the nondialysis groups, and responsiveness to clinician-evaluated change in severity. In summary, results from comparisons of measures across groups in four CKD-specific validation tests (Tables 3–6), showed that KDQOL-36 and CKD-QOL measures have greater discriminant validity than generic SF-12v2 in nearly all tests. As shown graphically in Figure 3, comparisons between CKD-specific measures reveal clear patterns of results across KDQOL and CKD-QOL measures in each test. For example, the top (blue) bar in the figure shows the range of variances (0.11–0.22 from Table 3) in three KDQOL scales explained by differences in treatment status groups, followed by the second (red) bar estimates for the three CKD-QOL scales in that test. Across the four comparative validity tests in Figure 3, results favored CKD-QOL scales over the referent (Burden) and other KDQOL-36 scales with two-fold

**Table 3.** Comparison of mean scores and validity statistics by treatment group

Measure	Mean (SD) Score			F Statistic	$\eta^2$ <sup>a</sup>	RV (95% CI)
	Stage 3–5, n=118	Dialysis, n=232	Transplant, n=111			
KDQOL-36						
Burden (+)	76.4 (24.7)	47.6 (27.1)	74.6 (25.2)	66.8 <sup>b</sup>	0.22	1.0
Effects (+)	82.9 (18.8)	62.1 (22.2)	83.1 (14.9)	64.3 <sup>b</sup>	0.22	0.96 (0.68 to 1.37)
Symptoms (+)	80.0 (15.8)	70.5 (17.2)	83.2 (13.2)	29.1 <sup>b</sup>	0.11	0.44 (0.25 to 0.70)
CKD-QOL						
CKD-QOL-6 (–)	42.9 (7.9)	51.8 (8.6)	42.3 (6.7)	77.3 <sup>b</sup>	0.25	1.16 (0.87 to 1.60)
CKD-QOL-CAT (–)	41.2 (9.4)	52.1 (8.0)	41.1 (8.5)	95.0 <sup>b</sup>	0.29	1.42 (1.07 to 1.97) <sup>c</sup>
Three- or six-item CAT (–)	41.9 (8.8)	52.2 (8.4)	41.5 (8.0)	89.3 <sup>b</sup>	0.29	1.34 (0.99 to 1.86)
SF-12v2						
PCS (+)	43.6 (10.5)	36.6 (10.1)	45.5 (10.6)	35.3 <sup>b</sup>	0.13	0.53 (0.32 to 0.91) <sup>c</sup>
MCS (+)	50.0 (10.1)	49.6 (10.7)	51.1 (10.0)	0.8	0.00	0.01 (0.00 to 0.05) <sup>c</sup>

Means are unadjusted. 95% CI, 95% confidence interval.

<sup>a</sup> $\eta^2$  estimate of variance explained by groups (see text); significance is same as F ratio.

<sup>b</sup> $P < 0.001$  (see text).

<sup>c</sup> $P < 0.05$  (see text).

median differences (0.22 versus 0.10) in variances ( $\eta^2$ ) explained. Figure 3 also shows nonoverlapping ranges of variances for referent and comparator scales and relatively wider and narrower ranges of variances across tests that have implications for measurement validity and interpretation. The relatively wider ranges observed across the three KDQOL scales reflect the consistently superior discriminant validity of the referent KDQOL Burden scale over the Symptoms scale in all four tests. Relatively narrower ranges of variances explained for the three CKD-QOL scales are consistent with alternate form measures of the same construct, presumably CKD-specific QOL. The implication is essentially the same or similar discriminant validity. An exception apparent in the physician-assessed change test reflects an unresponsive static CKD-QOL in comparison with the better performing two CAT scales (see Table 6). This pattern favoring CAT over static CKD-QOL forms was observed in all tests. The

pattern of variance estimates shown graphically also corroborates the only two significant RV estimates that favored CKD-QOL over the KDQOL referent measure by about 40% and 80%, respectively (Tables 3 and 5). Reasons for these and other differences in sets of measures studied, including the comorbidity test (Table 7), their implications for improving QOL measurement in clinical research and practice and other issues are discussed below.

To the best of our knowledge, this is the first study to report head-to-head comparisons of validity, including responsiveness, between KDQOL-36 and alternative CKD-specific measures. It also identified differences in discriminant validity between the three KDQOL scales. The largest improvement in validity with CKD-QOL over KDQOL-36 measures was observed for comparisons between patients differing in their own assessments of the severity of their CKD symptoms,

**Table 4.** Comparison of mean scores and validity statistics by eGFR<sup>a</sup>

Measure	Mean (SD) Score			F Statistic	$\eta^2$ <sup>b</sup>	RV (95% CI)
	GFR 30–59, n=153	GFR 15–29, n=50	GFR <15, n=17			
KDQOL-36						
Burden (+)	79.9 (24.0)	69.6 (24.8)	55.5 (23.1)	9.9 <sup>c</sup>	0.08	1.0
Effects (+)	84.8 (15.7)	81.5 (18.9)	70.3 (17.4)	6.1 <sup>d</sup>	0.05	0.61 (0.20 to 1.62)
Symptoms (+)	82.6 (14.1)	78.6 (16.3)	77.4 (16.8)	2.0	0.02	0.20 (0.00 to 0.67)
CKD-QOL						
CKD-QOL-6 (–)	41.4 (6.8)	44.8 (7.4)	47.7 (8.1)	9.1 <sup>c</sup>	0.08	0.92 (0.30 to 2.03)
CKD-QOL-CAT(–)	39.4 (8.63)	44.4 (8.6)	47.5 (8.4)	11.5 <sup>c</sup>	0.10	1.15 (0.51 to 2.56)
Three- or six-item CAT (–)	40.1 (7.9)	44.5 (8.3)	48.5 (8.3)	12.4 <sup>c</sup>	0.10	1.25 (0.54 to 2.85)
SF-12v2						
PCS (+)	45.4 (10.5)	42.2 (9.9)	43.1 (10.2)	2.0	0.02	0.20 (0.01 to 0.77)
MCS (+)	50.7 (10.2)	50.7 (10.3)	48.3 (9.9)	0.4	0.00	0.04 (0.00 to 0.19)

Means are unadjusted. 95% CI, 95% confidence interval.

<sup>a</sup>Abstracted from medical record.

<sup>b</sup> $\eta^2$  estimate of variance explained by groups (see text); significance is same as F ratio.

<sup>c</sup> $P < 0.001$  (see text).

<sup>d</sup> $P < 0.05$  (see text).



**Table 5.** Comparison of mean scores and validity statistics by self-evaluated severity of kidney disease symptoms<sup>a</sup>

Measure	Mean (SD) Score				F Statistic	$\eta^2$ <sup>b</sup>	RV (95% CI)
	Not Notable, n=148	Mild, n=120	Moderate, n=131	Severe, n=50			
KDQOL-36							
Burden (+)	81.6 (22.9)	63.8 (23.1)	48.0 (25.7)	31.6 (25.2)	74.5 <sup>c</sup>	0.33	1.0
Effects (+)	86.9 (13.2)	75.9 (17.4)	61.6 (21.6)	52.8 (23.7)	66.2 <sup>c</sup>	0.31	0.89 (0.65 to 1.21)
Symptoms (+)	86.1 (10.8)	79.4 (12.5)	68.0 (16.0)	62.5 (18.3)	58.6 <sup>c</sup>	0.28	0.79 (0.53 to 1.16)
CKD-QOL							
CKD-QOL-6 (-)	40.0 (5.3)	45.6 (6.4)	52.9 (7.7)	56.5 (8.4)	123.0 <sup>c</sup>	0.45	1.65 (1.24 to 2.21) <sup>d</sup>
CKD-QOL-CAT (-)	37.8 (7.2)	45.9 (7.3)	53.0 (7.7)	56.3 (6.3)	136.2 <sup>c</sup>	0.48	1.83 (1.36 to 2.45) <sup>d</sup>
Three- or six-item CAT (-)	38.6 (6.2)	45.9 (7.5)	53.2 (8.0)	56.5 (6.5)	131.7 <sup>c</sup>	0.47	1.77 1.29 to 2.39) <sup>d</sup>
SF-12v2							
PCS (+)	46.6 (9.7)	42.6 (9.8)	35.1 (9.9)	32.5 (9.7)	45.1 <sup>c</sup>	0.23	0.61 (0.37 to 0.95) <sup>d</sup>
MCS (+)	53.8 (9.1)	50.7 (10.1)	47.6 (9.8)	46.1 (11.8)	12.6 <sup>c</sup>	0.08	0.17 (0.07 to 0.31) <sup>d</sup>

Means are unadjusted. 95% CI, 95% confidence interval.

<sup>a</sup>Self-reported response to "How would you rate the severity of your kidney disease symptoms in the past 4 weeks?": not noticeable, mild, moderate, severe/very severe.

<sup>b</sup> $\eta^2$  estimate of variance explained by groups (see text); significance is same as F ratio.

<sup>c</sup>P<0.001 (see text).

<sup>d</sup>P<0.05 (see text).

followed closely by results from treatment status comparisons. These results suggest that patients view their CKD, symptoms, and treatment status as more severe when they have more extreme QOL impact. This association warrants further study. Trends in preliminary results from both cross-sectional and longitudinal analyses show the potential of further reductions in respondent burden with three- to six-item CAT forms that are shorter than four- to 12-item KDQOL-36 and static CKD-QOL measures. CAT forms were the only forms that discriminated across groups significantly in all four CKD-specific tests, including responsiveness.

The heart of this new approach is a bank of CKD-specific QOL impact items with improvements that are closer to the content of the Short Form-36 and other widely used generic health surveys. The main difference between these new and

generic items with common content is their attribution to kidney disease, as opposed to health in general. A crucial assumption underlying this approach to improving validity is that individuals can make valid attributions to CKD in the presence of MCCs. Because about two thirds of study participants had two or more chart-confirmed MCCs, this was a challenge for all CKD-specific measures studied. A weakness of the discriminant test (Table 6) of whether patients with MCCs can validly rate the impact of their CKD is that the tests were limited to evidence that the presence of MCCs did not appear to worsen CKD-specific QOL ratings. Alternative hypotheses are discussed below. Stronger discriminant tests, which are very rare in the QOL literature, require analyses of concurrent disease-specific clinical, severity, and QOL ratings for CKD and each comorbid condition.

**Table 6.** Comparison of mean change scores and validity statistics by clinician assessment of 3-month change in CKD status

Measure	Mean (SD) Change Score			F Statistic	$\eta^2$ <sup>a</sup>	RV <sup>b</sup>
	Better, n=28	Same, n=53	Worse, n=14			
KDQOL-36						
Burden (+)	2.68 (19.2)	-4.24 (21.3)	-12.64 (25.4)	2.48	0.05	1.00
Effects (+)	2.71 (17.6)	1.54 (10.1)	-2.00 (8.00)	0.67	0.01	0.27
Symptoms (+)	2.26 (10.8)	0.83 (8.3)	-4.40 (10.2)	2.44	0.05	0.98
CKD-QOL						
CKD-QOL-6 (-)	-0.40 (5.5)	1.05 (4.5)	3.40 (4.9)	2.84	0.06	1.14
CKD-QOL-CAT (-)	-1.40 (6.1)	1.07 (4.7)	4.90 (5.6)	6.65 <sup>c</sup>	0.13	2.68
Three- or six-item CAT (-)	-1.69 (6.6)	0.97 (4.8)	4.92 (4.8)	7.17 <sup>c</sup>	0.14	2.89
SF-12v2						
PCS (+)	-0.15 (7.1)	0.52 (5.6)	-7.55 (7.7)	9.11 <sup>c</sup>	0.16	3.67
MCS (+)	-1.06 (5.9)	-1.08 (7.7)	-5.10 (7.8)	3.39 <sup>d</sup>	0.07	1.37

Means are unadjusted.

<sup>a</sup> $\eta^2$  estimate of variance explained by groups (see text); significance is same as F ratio.

<sup>b</sup>Very small samples and other factors precluded estimation of useful confidence intervals using the bootstrap method.

<sup>c</sup>P<0.01 (see text).

<sup>d</sup>P<0.05 (see text).

**Table 7.** Comparison of mean scores and RV by number of comorbid conditions<sup>a</sup>

Measure	Mean (SD) Score				F Statistic	$\eta^{2b}$	RV (95% CI)
	0, n=137	1, n=140	2, n=83	3+, n=101			
KDQOL-36							
Burden (+)	62.8 (29.1)	62.5 (30.3)	57.5 (27.9)	61.6 (30.3)	0.6	0.00	0.09 (0.02 to 0.27) <sup>c</sup>
Effects (+)	72.3 (22.6)	73.8 (22.3)	72.5 (20.9)	70.8 (23.4)	0.4	0.00	0.05 (0.00 to 0.12) <sup>c</sup>
Symptoms (+)	79.1 (17.5)	76.2 (16.8)	73.4 (16.3)	73.6 (16.2)	3.0 <sup>c</sup>	0.02	0.39 (0.07 to 1.06)
CKD-QOL							
CKD-QOL-6 (-)	47.0 (9.4)	47.6 (9.5)	46.8 (8.2)	47.6 (9.4)	0.2	0.00	0.03 (0.00 to 0.05) <sup>c</sup>
CKD-QOL-CAT (-)	46.2 (10.7)	46.6 (10.5)	47.2 (8.8)	46.8 (10.0)	0.2	0.00	0.02 (0.00 to 0.04) <sup>c</sup>
Three- or six-item CAT (-)	46.7 (10.3)	47.1 (10.1)	47.1 (9.0)	47.0 (9.8)	0.1	0.00	0.01 (0.00 to 0.01) <sup>c</sup>
SF-12v2							
PCS (+)	43.0 (11.8)	41.3 (10.7)	40.3 (9.9)	36.3 (10.2)	7.6 <sup>d</sup>	0.05	1.0
MCS (+)	50.9 (9.4)	50.2 (10.6)	49.4 (11.2)	49.2 (10.7)	0.7	0.01	0.09 (0.00 to 0.33) <sup>c</sup>

Means are unadjusted. 95% CI, 95% confidence interval.

<sup>a</sup>Count of number of comorbid conditions abstracted from medical record (amputation above ankle, amputation below ankle, angina, cardiomegaly, chronic fatigue syndrome, congestive heart failure, chronic obstructive pulmonary disease, diabetes, fibromyalgia, foot ulcers for >3 months, hepatitis, HIV/AIDS, migraine headaches, myocardial infarction in past year, neuropathy/nerve damage, osteoarthritis, restless leg syndrome, rheumatoid arthritis, sleep disorders/insomnia/sleep apnea, and stroke).

<sup>c</sup> $P < 0.05$  (see text).

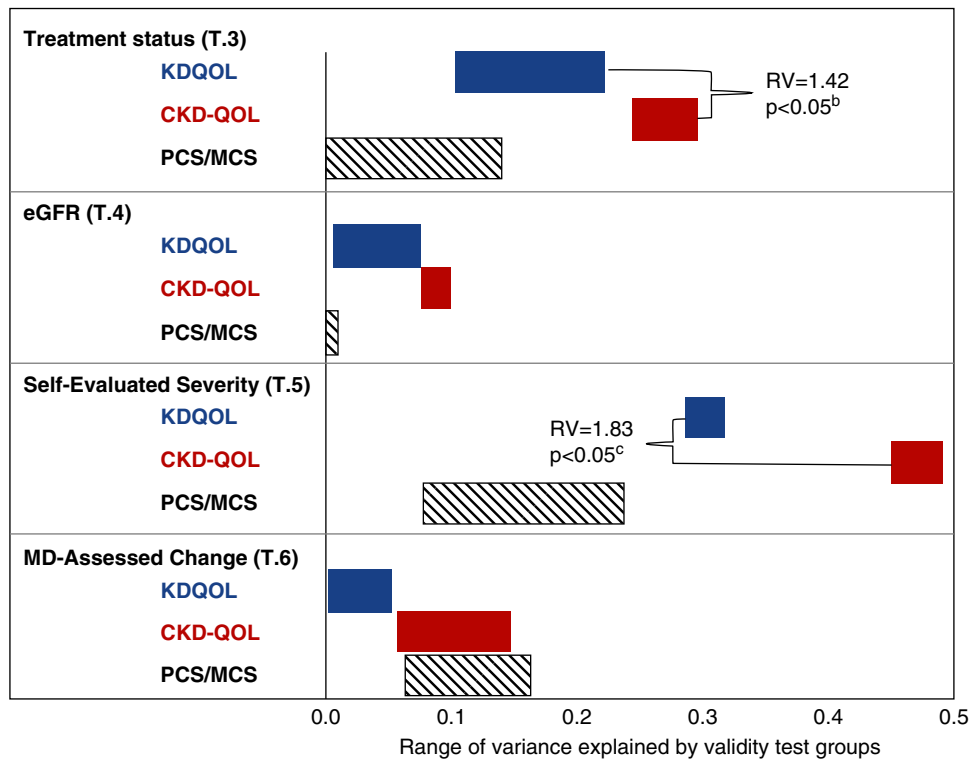
<sup>d</sup> $P < 0.01$  (see text).

As part of a large, internet-based study of chronically ill United States adults,<sup>41</sup> samples with CKD and each of eight comorbidities (diabetes, osteoarthritis, hypertension, seasonal allergies, chronic back problems, hip/knee joint problems, and anemia) were sufficient (sample size  $\geq 50$  and matched methods) for such tests. The methods and tables of CKD-specific results from this study are summarized in the Supplemental Appendix. In summary, for pairs of CKD and comorbid conditions, correlational tests compared the convergent (same disease, different methods) and discriminant (different diseases, same method) validity of CKD-specific and comorbidity-specific QOL impact ratings. Discriminant tests correlated measures of different diseases using the same method, namely CKD in the presence of each comorbid condition. In summary, CKD-specific convergent correlations among three different methods (eGFR, severity, and CKD-specific QOL impact) were significant and two were substantial ( $r=0.39$ – $0.72$ , median: 0.45). Discriminant correlations between CKD QOL impact and QOL impact attributed to each comorbid condition were significantly lower ( $r=0.02$ – $0.49$ , median: 0.13) in magnitude. Across 51 tests of discriminant validity, 71% of correlations between matched methods for measuring different diseases (CKD versus each comorbid condition), which should be lower for valid measures, were significantly lower than their convergent correlations. The great majority of exceptions, and the only two substantial exceptions, involved comorbid anemia and back problems. Such exceptions warrant further study to better understand their implications for the interpretation of CKD-specific QOL ratings, particularly with larger samples. Overall, our study findings and published convergent and discriminant test results appear to be sufficient to warrant further applications and continued testing of the new approach to improving CKD-specific QOL impact attributions.

In contrast to KDQOL-36, CKD-QOL items are aggregated into a single QOL summary impact score. Such aggregate QOL impact scoring is consistent with analogous one-factor models and similar approaches used successfully in other therapeutic areas.<sup>14,22,23</sup> This approach is in sharp contrast to the physical and mental summary scores for SF-12/Short Form-36 and other comprehensive generic surveys asking very similar QOL impact questions, but with attributions to health in general. Psychometric evaluations of the latter have consistently yielded distinct higher-order physical and mental factors.<sup>37,42,43</sup> In contrast, it appears that adults asked to focus on their CKD make QOL ratings on the basis of overall CKD severity and their experience of treatment status, and less on the basis of divergent patterns of social, emotional, and other aspects of QOL. Charts of group means confirming this empirically for domain-specific and the summary CKD-QOL measures in this study are documented in the Supplemental Appendix.

This new approach to CKD-specific measurement warrants further evaluation. Because a single score takes fewer items to estimate than multiple scores, and such scores are more reliable, the brevity and simplicity of a 1-minute summary short form makes such tests practical. On the basis of face and content validity and empirical performance, the new approach appears to yield information that is both CKD-specific and about QOL. The face and content validity of the Burden scale may also explain its better performance over the other two KDQOL measures. We recommend replication of head-to-head comparisons of forms with improvements over our study limitations using larger and more diverse samples and lengthening the follow-up period for validity tests of responsiveness.

However, even if all CKD-specific measures performed equivalently across tests, there are practical reasons to consider, if not favor, an approach that substantially (75% or more)



<sup>a</sup>Ranges of variance ( $\eta^2$ ) shown graphically and RV estimates are from Tables 3-6.  
<sup>b</sup>RV test favors CKD-QOL-CAT comparator over referent KDQOL Burden scale,  $p < 0.05$ , see Table 3  
<sup>c</sup>RV test favors three CKD-QOL static and CAT comparators over referent KDQOL Burden scale,  $p < 0.05$ , see Table 5.

**Figure 3.** Non-overlapping ranges of variances explained by groups and RV estimates across tests confirm greater CKD-QOL validity.

reduces respondent burden. CAT-based survey administrations should be further evaluated to observe if gains in precision and reduced respondent burden achieved in our study, and for similar disease-specific QOL measures in other conditions,<sup>31,44,45</sup> are replicated. The implications of very short CAT forms (e.g., three to six items) for individual patient monitoring warrant further study. Because some patient surveys must be limited to static forms, our study tested a six-item static form that was evaluated in parallel with electronic data capture using CAT as in previous studies. The correlation between static and CAT-based scores was very high ( $r=0.94$ ) and their averages were nearly identical for most groups compared, and most often led to the same conclusions; the exception being the responsiveness test.

To facilitate further scholarly research, all CKD-QOL forms tested in this study are available from the authors. Also available is the newer Quality of Life Disease Impact Scale,<sup>14</sup> which incorporates noteworthy additional improvements including even broader QOL content representation (e.g., adding physical functioning) and IRT-based scoring that has been standardized across CKD and other chronic conditions and normed in the United States chronically ill population. In addition, results from qualitative and other research may lead to further improvements and, because this study was

conducted before the Standardized Outcomes in Nephrology initiative,<sup>46</sup> the conclusions from the initiative should be considered in any revision to the CKD-QOL.

The KDQOL-36 Symptoms/Problems scale, which is an aggregation of 12 relatively heterogeneous symptoms, was the least valid CKD-specific scale in four CKD-specific tests, and it worsened significantly in the presence of comorbid conditions. Whether this is a relative strength or weakness depends on purpose, measuring total disease or CKD-specific QOL burden. Regardless, this pattern of results should be considered when interpreting such aggregations of symptoms, QOL, and other effects. With improvements in short-form surveys such as those reported here, comprehensive patient-based assessment throughout the continuum of CKD-specific and generic QOL outcomes (see Figure 1) may be feasible. In place of the whole KDQOL-36, the best-performing of its scales (Burden) or a new six-item (1-minute) CKD-specific QOL impact measure could be used to monitor CKD-specific QOL outcomes with survey space sufficient to add to the “dashboard” specific prevalent symptoms (e.g., cramping, itching) of importance to patients, to be interpreted separately to achieve more clinically useful and actionable information. Very short improved tools also facilitate a more practical migration from legacy surveys such as KDQOL-36 in parallel with improved tools. CAT

surveys also can pursue additional questioning; for example, with respect to fatigue or mental health, depending on whether they substantially affect the patient. Indeed, a CAT is only the first component of an interactive system. By reducing respondent burden in measuring core scales, it frees respondent time and attention to respond to individualized items.

A CAT may not administer the same items to a patient over time. When the underlying domain is what is important and items use IRT-based scoring, administering the same items is not necessary to determine the respondent's level for that concept and, in some instances, may not be appropriate. For example, if an interactive physical functioning module were administered to a patient immediately after repair of a lower extremity fracture, questions about the ability to transfer to a chair might be relevant, whereas a year later relevant questions might include the distance the patient could run, but not questions about the ability to transfer. There are, however, individual items that are considered so intrinsically important that they should be conserved across administrations of an instrument. For patients on hemodialysis, the "recovery question" may be such an item, as may be a question about whether the patient ever experiences muscle cramps during or between treatments. Instruments used in clinical practice should allow clinicians responsible for content to "force" inclusion of items they consider important for all respondents at all times.

Improvements in QOL surveys must balance practicality with the precision needed for clinical research and practice, particularly for scores at the individual patient level. Comprehensive static measures with enough items for precision across higher and lower score levels, as required across multiple settings and for patients with diverse clinical presentations, requires many more time-consuming items and can be cost-prohibitive. Repeated administrations of such lengthy forms are often unacceptable to patients. Static surveys shortened by restricting the range of measurement often lack the precision required for interpretation for individual patients at the earlier CKD stages experiencing smaller but still noteworthy QOL impairments, as well as for those at CKD stage 5 about to begin dialysis, and those treated by dialysis. Whether CAT-based administrations that match items to patients' severity levels can improve QOL assessment, as observed in this study, and are also more clinically useful in improving patient care warrants further study.

Our study was limited by the use of a convenience sample that overrepresented white patients. Age, sex, and educational attainment distributions are similar to their respective populations and the different CKD treatment groups are what would be expected clinically. However, future work should explore CKD-QOL or similar measures in a more representative population.

In conclusion, this new approach to expanding and summarizing CKD-specific QOL impact in a single score promises to better capture the QOL effects of differences in CKD treatment status and severity on patient-reported outcomes. Focusing on QOL impact attributed specifically to CKD appears to improve clinical validity across multiple tests, including responsiveness. It is also likely that CAT-based administrations

that match survey items to patient severity levels will enable more efficient QOL surveys, with reductions in respondent burden and clinically useful gains in score precision.

## ACKNOWLEDGMENTS

The National Kidney Foundation supported early development of this work by encouraging participants in its Kidney Early Evaluation Program to complete online questionnaires enabling development of CKD-specific quality-of-life impact scale measures. Members of the Tufts Medical Center Division of Nephrology made valuable comments on study design and allowed recruitment of patients from the Tufts Medical Center Kidney and BP Center and from community-based practices. Dr. Ajay Singh facilitated recruitment of patients from the Brigham and Women's Hospital Division of Nephrology. We gratefully acknowledge patients, staff, and medical directors of Dialysis Clinic, Inc., and study coordinators Kimberly Clayton Lindsey, Alice Martin, and Alison Taubes for study recruitment, data collection, and management.

J.E.W., M.M.R., and K.B.M. designed the study. M.M.R., K.B.M., and J.E.W. supervised or participated in data acquisition. J.E.W. and B.G. analyzed the data. All authors interpreted the data. J.E.W. and B.G. made the figures. All authors drafted and revised the paper, and approved the final version of the manuscript.

The Functional Health Computer Adaptive Test in the CKD study was funded by the National Institutes of Health (grant 2 R44-DK062555-03; coinvestigators: J.E.W. and K.B.M.). A research grant donation from the Amgen Foundation (J.E.W.) supported project planning and data analysis. John Ware Research Group and Tufts Medical Center supported data analysis and reporting of this validation study out of their own research funds.

## DISCLOSURES

J.E.W. reports grants from the National Institutes of Health, a research donation from the Amgen Foundation, and other support from the John Ware Research Group during the conduct of the current study; being an original copyright holder of generic SF-36 and SF-12 used in short and long forms of KDQOL; and developer and copyright holder of new CKD-specific forms used in the current study and developed subsequently.

## SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at <http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2018080814/-/DCSupplemental>.

Supplemental Appendix. Summary of CKD health-related quality of life (QOL) item bank development and psychometric evaluations.

## REFERENCES

1. Perrone RD, Coons SJ, Cavanaugh K, Finkelstein F, Meyer KB: Patient-reported outcomes in clinical trials of CKD-related therapies: Report of a symposium sponsored by the national kidney foundation and the U.S. Food and Drug Administration. *Am J Kidney Dis* 62: 1046–1057, 2013

2. Kusek JW, Greene P, Wang SR, Beck G, West D, Jamerson K, et al.: Cross-sectional study of health-related quality of life in African Americans with chronic renal insufficiency: The African American study of kidney disease and hypertension trial. *Am J Kidney Dis* 39: 513–524, 2002
3. Da Silva-Gane M, Wellsted D, Greenshields H, Norton S, Chandna SM, Farrington K: Quality of life and survival in patients with advanced kidney failure managed conservatively or by dialysis. *Clin J Am Soc Nephrol* 7: 2002–2009, 2012
4. Patrick DL, Deyo RA: Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 27[Suppl 3]: S217–S232, 1989
5. Meyer KB, Espindle DM, DeGiacomo JM, Jenuleson CS, Kurtin PS, Davies AR: Monitoring dialysis patients' health status. *Am J Kidney Dis* 24: 267–279, 1994
6. Unruh M, Benz R, Greene T, Yan G, Beddhu S, DeVita M, et al.: HEMO Study Group: Effects of hemodialysis dose and membrane flux on health-related quality of life in the HEMO Study. *Kidney Int* 66: 355–366, 2004
7. Wu AW, Fink NE, Marsh-Manzi JV, Meyer KB, Finkelstein FO, Chapman MM, et al.: Changes in quality of life during hemodialysis and peritoneal dialysis treatment: Generic and disease specific measures. *J Am Soc Nephrol* 15: 743–753, 2004
8. 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
9. Ware J Jr., Kosinski M, Keller SD: A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Med Care* 34: 220–233, 1996
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. 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
12. 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
13. Naik N, Hess R, Unruh M: Measurement of health-related quality of life in the care of patients with ESRD: Isn't this the metric that matters? *Semin Dial* 25: 439–444, 2012
14. Ware JE Jr., Gandek B, Guyer R, Deng N: Standardizing disease-specific quality of life measures across multiple chronic conditions: Development and initial evaluation of the QOL Disease Impact Scale (QDIS®). *Health Qual Life Outcomes* 14: 84, 2016
15. Centers for Medicare & Medicaid Services: End-Stage Renal Disease Patient-Reported Outcomes Technical Expert Panel: Summary Report, In-Person Meeting. ESRD Quality Measure Development, Maintenance, and Support, Contract Number HHSM-500-2013-130171, Baltimore, MD, Centers for Medicare & Medicaid Services, 2017
16. Mujais SK, Story K, Brouillette J, Takano T, Soroka S, Franek C, et al.: Health-related quality of life in CKD patients: Correlates and evolution over time. *Clin J Am Soc Nephrol* 4: 1293–1301, 2009
17. Bjorner JB, Kosinski M, Ware JE Jr.: Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the headache impact test (HIT). *Qual Life Res* 12: 913–933, 2003
18. Haley SM, McHorney CA, Ware JE Jr.: Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol* 47: 671–684, 1994
19. Fisher WP Jr., Eubanks RL, Marier RL: Equating the MOS SF36 and the LSU HSI physical functioning scales. *J Outcome Meas* 1: 329–362, 1997
20. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al.: PROMIS Cooperative Group: The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol* 63: 1179–1194, 2010
21. Wainer H, Dorans NJ, Eignor D, Green B.F., Flaugher R., Mislevy R.J., et al.: *Computerized Adaptive Testing: A Primer*, 2nd Ed., New York, Routledge, 2014
22. de Boer AG, Spruijt RJ, Sprangers MA, de Haes JC: Disease-specific quality of life: Is it one construct? *Qual Life Res* 7: 135–142, 1998
23. Webb SM, Badia X, Barahona MJ, Colao A, Strasburger CJ, Tabarin A, et al.: Evaluation of health-related quality of life in patients with Cushing's syndrome with a new questionnaire. *Eur J Endocrinol* 158: 623–630, 2008
24. Bayliss MS, Espindle DM, Buchner D, Blaiss MS, Ware JE: A new tool for monitoring asthma outcomes: The ITG asthma short form. *Qual Life Res* 9: 451–466, 2000
25. Wilson IB, Cleary PD: Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 273: 59–65, 1995
26. Wu AW, Fink NE, Cagney KA, Bass EB, Rubin HR, Meyer KB, et al.: Developing a health-related quality-of-life measure for end-stage renal disease: The CHOICE health experience questionnaire. *Am J Kidney Dis* 37: 11–21, 2001
27. Richardson MM, Saris-Baglana RN, Anatchkova MD, et al.: Patient experience of chronic kidney disease (CKD): Results of a focus group study. Presented at Annual National Kidney Foundation Spring Clinical Meeting, Orlando, FL, April 2007
28. Kosinski M, Bjorner JB, Ware JE Jr., Sullivan E, Straus WL: An evaluation of a patient-reported outcomes found computerized adaptive testing was efficient in assessing osteoarthritis impact. *J Clin Epidemiol* 59: 715–723, 2006
29. Turner-Bowker DM, Saris-Baglana RN, Derosa MA, Paulsen CA, Bransfield CP: Using qualitative research to inform the development of a comprehensive outcomes assessment for asthma. *Patient* 2: 269–282, 2009
30. Turner-Bowker DM, Saris-Baglana RN, Derosa MA, Paulsen CA: Cognitive testing and readability of an item bank for measuring the impact of headache on health-related quality of life. *Patient* 5: 89–99, 2012
31. Ware JE Jr., Kosinski M, Bjorner JB, Bayliss MS, Batenhorst A, Dahlöf CG, et al.: Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res* 12: 935–952, 2003
32. Ware JE Jr., Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A: Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: Summary of results from the medical outcomes study. *Med Care* 33[Suppl 4]: AS264–AS279, 1995
33. Ware JE Jr., Kosinski M, Turner-Bowker DM, Gandek B: *How to Score Version Two of the SF-12 Health Survey*, Lincoln, RI, QualityMetric Incorporated, 2002
34. Cronbach LJ: Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334, 1951
35. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al.: PROMIS Cooperative Group: Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 45[Suppl 1]: S22–S31, 2007
36. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al.: Scientific Advisory Committee of the Medical Outcomes Trust: Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 11: 193–205, 2002
37. McHorney CA, Ware JE Jr., Raczek AE: 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, 1993
38. Deng N, Allison JJ, Fang HJ, Ash AS, Ware JE Jr.: Using the bootstrap to establish statistical significance for relative validity comparisons among patient-reported outcome measures. *Health Qual Life Outcomes* 11: 89, 2013
39. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR; Clinical Significance Consensus Meeting Group: Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 77: 371–383, 2002
40. Campbell DT, Fiske DW: Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 56: 81–105, 1959

41. Ware JE Jr., Gandek B, Allison J: The validity of disease-specific quality of life attributions among adults with multiple chronic conditions. *Int J Stat Med Res* 5: 17–40, 2016
  42. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK: An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Med Care* 35: 522–537, 1997
  43. Ware JE Jr., Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al.: The factor structure of the SF-36 health survey in 10 countries: Results from the IQOLA project. International quality of life assessment. *J Clin Epidemiol* 51: 1159–1165, 1998
  44. Bjorner JB, Kosinski M, Ware JE Jr.: Using item response theory to calibrate the Headache Impact Test (HIT) to the metric of traditional headache scales. *Qual Life Res* 12: 981–1002, 2003
  45. Kosinski M, Bayliss MS, Bjorner JB, Ware JE Jr., Garber WH, Batenhorst A, et al.: A six-item short-form survey for measuring headache impact: The HIT-6. *Qual Life Res* 12: 963–974, 2003
  46. Tong A, Craig JC, Nagler EV, Van Biesen W; SONG Executive Committee and the European Renal Best Practice Advisory Board; SONG Executive Committee and the European Renal Best Practice Advisory Board: Composing a new song for trials: The Standardized Outcomes in Nephrology (SONG) initiative. *Nephrol Dial Transplant* 32: 1963–1966, 2017
  47. Cohen J: *Statistical Power Analysis for the Behavioral Sciences*, 2nd Ed., New York, Psychology Press, 1988
  48. Armor DJ: Theta reliability and factor scaling. *Sociol Methodol* 5: 17–50, 1973
  49. Thissen D: Reliability and measurement precision. In: Wainer H, et al., *Computerized Adaptive Testing: A Primer*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates 2000: 159–183
- 
- See related editorial, “Patient-Reported Outcomes: Toward Better Measurement of Patient-Centered Care in CKD,” on pages 523–525.