Supplemental Material

Summary of Chronic Kidney Disease (CKD) Health-related Quality of Life (QOL) Item Bank Development and Psychometric Evaluations

Table of Contents

Background

Prior to the clinical validation study, a measurement development project was fielded among independent samples of adults varying across stages of CKD (Stages 3-5), with the following aims:
(1) Expand the item pool and evaluate the appropriateness of new items for CKD-specific applications;
(2) Identify the most important and empirically useful QOL domains to be represented;
(3) Use psychometric methods to construct and evaluate an item bank sufficient for static and computerized adaptive test (CAT) forms and evaluate their performance in terms of psychometric properties underlying scale construction and scoring.

New Instrument Development

To develop an item pool appropriate for QOL impact assessment across a wide range of CKD severity levels and throughout the course of CKD treatment, this effort proceeded in five steps: 1) review of existing HRQOL instruments and their items; 2) focus group studies, 3) Clinical Advisory Board input; and 4) cognitive testing; and 5) pilot tests of items. An important consideration was feasibility of item administrations across multiple data collection modalities (paper-pencil, electronic data collection).

Review of Existing QOL Instruments and Item Pools

The review indicated that although multiple instruments had been used to measure health-related quality of life (QOL) in studies of CKD [1-4], the literature suggests that the most widely-used questionnaires in large scale nephrology studies are the SF-36® Health Survey [5], a generic measure, and the Kidney Disease Quality of Life (KDQOL™), long or short form survey [6]. The KDQOL-SF™ includes the SF-36® Health Survey (KDQOL-36 includes its SF-12 subset) as its generic core and additional content areas that are not represented in the SF-36 (e.g., kidney effects and burden, symptoms) but are of particular relevance to the dialysis population. The comprehensive review of our item pools and existing HRQOL measures used in CKD research examined domain content coverage and item content [7]. Overlap and gaps in domain coverage and item content was observed across tools. Concepts common across CKD-specific and generic measures included mental health, social and role functioning, pain, and fatigue (and to a lesser extent sleep, diet/appetite, social support, appearance, and spirituality). Most of the domains were well-covered by existing item pools but some areas were not (i.e., sexual functioning, sleep, appearance), suggesting content areas for additional consideration.

Focus Group Studies

Prior focus group research [6, 8] with dialysis patients identified the following domain content areas: fatigue, trouble focusing, loss of time/freedom/control, social relationships, body image, role functioning, sexual relations, mental health, and sleep. However, patients in earlier stages of CKD were not well represented in these studies. Accordingly, four focus groups were conducted with CKD patients recruited from Boston-area clinics and dialysis units (see Table 1 for sample characteristics). Men and women in stages 3-5 not on dialysis (n=20) and Stage 5 on dialysis (n=20) were interviewed separately by gender and use of dialysis to understand common and unique domains of CKD impact experienced [9]. Participants defined “quality of life,” discussed the impact of CKD on their lives, and commented on the clarity and

appropriateness of sample disease-specific HRQOL items. Sessions were audio taped and transcribed. Qualitative content analyses were conducted to illuminate and confirm important areas of CKD impact and to identify potential problems with specific item content. Participants defined "quality of life" as life before diagnosis, living a healthy and "normal" life, independence, and freedom. They identified mental health, intimacy/sexuality, fatigue, role functioning related to work, social relationships/support, independence, and finances as key areas affected by CKD. When responding to HRQOL items that asked them to think specifically about their kidney disease (e.g., In the past 4 weeks, how much of the time has your kidney disease left you too tired to do work or daily activities?), many but not all participants felt that they could attribute health states to their kidney disease; patients not on dialysis questioned whether they should focus on the disease itself or on the disease and its treatment; while dialysis patients did not typically distinguish between the two. Suggestions for survey refinement were provided.

Clinical Advisory Board Meeting

We held a full-day meeting with a Clinical Advisory Board at Tufts Medical Center (TMC) in Boston, MA to share the results of the focus group study and solicit feedback on domain-specific content coverage, item content wording, and desired characteristics of the application. Board members identified domains of importance to practitioners and their patients from their perspective (prior to learning the results of the focus groups); discussed the pros and cons of asking HRQOL items with and without disease attributions; and critiqued a prototype computer adaptive testing (CAT) administration method by offering suggestions for interface improvement and reporting features. Prioritized domains included mental health, family, work, and social functioning, sleep, intimacy/sexuality, and fatigue. Clinicians expressed ambivalence toward items with disease-specific attributions, noting the multitude of co-morbidities associated with CKD and the cognitive complexity of pinpointing the main cause of a problem (e.g. poor sleep), while others noted the inherent value of combining generic and disease-specific measurement by applying disease-specific attribution to generic item content. Based on these findings we added a study component to examine how people respond to identical items with and without attribution to kidney disease. Upon completion of these three steps [10], we determined that we had comprehensive generic item banks for several of the domains (e.g., role functioning, vitality/fatigue, social functioning, mental health, pain) identified as important to CKD researchers, patients, and clinicians. We expanded item content for the development of a Kidney Disease Impact Survey (KDIS) CAT, specifically, and examined items for two concepts (Role and Sleep) with and without an attribution to kidney disease. The 58 developmental KDIS items included 37 global disease impact items revised to include an attribution to kidney disease and 21 additional items to expand content coverage (e.g., sexuality, body image/appearance, social stigma). In addition, 41 selected items from the KDQOL-SF™ [6], 16 role and sleep items with an attribution to health, the SF-12v2® Health Survey [11], and a set of background items also were programmed for on-line administration. [Note: The item bank and resulting survey measure at this stage of development was labeled Kidney (K) Disease Impact Scale (DIS) and is referred to as KDIS below and in tables of results reported at that time. These items are now referred to as "CKD-QOL" in the manuscript].

Cognitive Testing

Cognitive interviews to test the KDIS items with CKD patients (on dialysis and not on dialysis) probed on interpretation and comprehension of the KDIS items, memory recall of relevant information, decision processes, response processes, and content coverage. In addition, specific probes were used to ask about disease attribution to kidney disease. In general, patients had very few problems comprehending the questions correctly. Several

patients did not understand the word “productivity” in KDIS18, while in KDIS36 (avoid traveling) several people included local travel, while others thought the item only focused on long distance travel. One person reported that in answering one item (KDIS6) she took her allergies into consideration as well as her CKD, but she was able to make the distinction between kidney disease and allergies.

Pilot Study of Sampling and Data Collection

After testing methods for collecting data augmented with Glomerular Filtration Rate data (for severity staging) from large samples via Internet-based *physician panels* without sufficient success, we collaborated with the National Kidney Foundation (NKF) to access their large *patient panel* developed through the Kidney Early Evaluation Program (KEEP) and initiate outreach to consumers. On-line data collection limited participants to those who self-reported the month and result of their most recent serum creatinine test. Recruitment letters were e-mailed to ListServs (e.g., KEEP participants, NKF Patient & Family Council, People Like Us Patient Advocates) to drive people to the NKF website (www.kidney.org) where they could link to the study site as required for pilot testing. After the first 100 cases were collected, sample characteristics were evaluated in relation to target enrollment criteria (e.g., severity stage). The sample was well represented in terms of the numbers of people on dialysis (56%) versus earlier stages of the disease (44%); age (18-76), gender (58% female); and other sociodemographics. Several participants (n=9) did not have serum creatinine (SCr) results within the past six months, an initial screener we used for participation. This time period for reporting SCr data was then increased to 12 months in the item bank calibration study, with the goal of increasing recruitment without affecting staging accuracy. Overall, results of the pilot test suggested that the survey methods and resulting item-level data were sufficient to launch the full item bank development and calibration.

Item Bank Calibration Study

Item bank development began with consideration of the underlying conceptual framework. How much simplification is possible without loss of information? Thorough tests addressed whether a single overall CKD impact score is psychometrically sound and likely to be clinically useful in understanding how CKD severity, symptoms, and treatment status impact upon HRQOL. This work proceeded in four steps: 1) participant sampling and data collection; 2) development of an item bank; 3) item bank analyses (data quality evaluation, factor analyses, evaluation of item characteristic curves and differential item functioning); and 4) IRT parameter estimation and model fit.

Sampling

Large-scale internet-based data collection occurred through the National Kidney Foundation website [12]. Recruitment methods were staggered over time to achieve target enrollment. Recruitment letters were mailed to 22,215 National Kidney Foundation (NKF) Patient & Family Council members, People Like Us Patient Advocates, and Kidney Early Evaluation Program (KEEP) participants with eGFR < 60 mL/min/1.73m2. After observing a lull in enrollment approximately 3 months into data collection, a recruitment article was published in *Kidney Beginnings*, a publication of the American Association of Kidney Patients (AAKP). Respondents who visited the NKF website were provided a brief description of the study and were asked to input their KEEP ID number provided in the recruitment letter, if they had one. Those who were not KEEP members were assigned a unique ID. Then participants were linked to a website where they read consent forms on-line and indicated their willingness to participate

in the study by selecting an “I agree” button, prior to the launching of the survey. Many people showed an interest in the study, as evidenced by the substantial number of clicks on the survey. Although 58% of unique cases clicking on the survey consented to participate, 16% of these dropped out prior to the screener. Subject loss may be due in part to technical issues encountered (e.g., several calls were received regarding problems accessing or maintaining access to the survey through the Internet browser). Of those reporting a Serum Creatinine (SCr) test date within the last 12 months (approximately 98%), 69% (n=1,121) were able to provide an actual SCr result. A total of 1,636 respondents initiated the survey, and 76% (n=1,236) of these completed the full survey (median response time =24 minutes). As shown in Table 2, this sample was diverse in terms of age [range 18-90 (M=49)], gender [female (62%)], race, educational level, income, and employment status. Hispanics were underrepresented relative to the CKD and U.S. general populations. African Americans and older respondents (65+) were somewhat underrepresented relative to the CKD population, yet prevalence was comparable to the U.S. general population. As expected, those with higher educational attainment and socioeconomic status were over-represented. The sample was well distributed across CKD severity stages, with good representation across dialysis, non-dialysis, and transplant groups.

Data Collection

Traditional methods of data collection can be costly when large samples are involved or required. This study is among the first, to the best of our knowledge, to report on the feasibility of collecting HRQOL data from people with CKD via the Internet, including those who are older and on dialysis. Driving people to the NKF website may have influenced the positive response to this study. Respondents participated without monetary incentive and no additional steps were required for data entry, suggesting that this may be a less expensive option for collecting nationwide data. Although this sample was diverse and demographics were generally comparable to the U.S. general population, it became clear that going forward targeted recruitment efforts may be needed to secure a sample closely representative of the CKD population.

Item Bank Data Quality

Based on inclusion criteria, the analytic sample was comprised of CKD Stage 3-5 (non-dialysis, non-transplant) [N=463 (Stage 3, n=218; Stage 4, n=180; Stage 5, n=65)], Dialysis [N=392 (Hemodialysis, n=318; Peritoneal Dialysis, n=74] and Transplant (N=313) participants, who were administered an initial 58-item KDIS bank and the KDQOL-SF. Each KDIS item was missing less than 5% of the time. Items most missed included sexual functioning and stigma-related items. All other items were missed by very few respondents. Within each key group of interest (CKD Stage 3-5, Dialysis, Transplant), all item response options were used. As expected, response options indicative of greater disease impact were used less frequently by Transplant and CKD Stage 3-5 respondents than by Dialysis respondents. None of the items showed extreme skewness (95% of responses in one category). Similarly, excluding the two sexual functioning items that were skipped when respondents reported that they were not sexually active, 97% of the sample completed all KDQOL items. In each subgroup, for each item, all item response options were used, and none of the items showed extreme skewness. Therefore, all KDIS and KDQOL items were retained for further evaluation. Ninety-six percent of those administered the item bank completed all 58 KDIS bank items (N=1,123).

Psychometric Evaluation

Summary of Psychometric Methods

Data analyses were performed in three major steps: (1) evaluation of response categories for items in each domain; (2) categorical factor analysis to confirm the unidimensionality of items hypothesized to measure each domain; and (3) IRT calibration of confirmed homogeneous items [13]. As in previous studies, analysis began with tests of the ordering of item response categories using nonparametric methods using TestGraf software. Our goal was to retain as many items as possible to comprehensively represent QOL content and cover its full range. The steps and criteria in these evaluations included: factor analysis of items, examination of item trace lines in relation to the total score, robustness of item parameters across subgroups, and evaluation of residuals indicating item fit. Factor analyses were performed to confirm the unidimensionality of items. Because distributions of responses to many items were highly skewed, a matrix of polychoric correlations among items was also factor analyzed using Mplus software. IRT modeling was performed for each unidimensional item bank. All items in a given bank were calibrated on a common metric using the generalized partial credit model (GPCM), which represents, respectively, the more parsimonious Rasch model with its distinct advantages and the more general IRT model, which allows each item to have a separate slope and better fits the data. Although the goal was to develop item calibrations on the pooled data, we tested whether, and the extent to which, distinct patient groups demonstrated differential item functioning (DIF) of item characteristics using the logistic regression methods as in previous studies.

Summary of Results-Item Characteristic Curves

Examination of item response curves using nonparametric methods with the program TestGraf [14] showed that each response option curve for retained KDIS bank items had one clear maximum; some CKD-specific KDQOL items not included in the KDIS bank were exceptions. No KDIS item categories required collapsing to meet the assumption or ordinal responses.

Summary of Results-Factor Analyses

To evaluate item bank dimensionality, we conducted confirmatory factor analytic (CFA) [15] and reviewed: 1) eigenvalues associated with each factor extracted; 2) item loadings on the primary factor; and 3) results from overall model fit tests. Stringent testing of a single-factor (unidimensional) global QOL impact model in a CKD population was the priority. Therefore, in each of the three treatment groups (CKD Stage 3-5, Dialysis, Transplant) and in all groups combined, we conducted a series of CFAs, culminating in a 1-factor confirmatory model for a bank of 37 KDIS items across multiple QOL domain content areas Including: role functioning (work, family, usual activities), social functioning, mental health, fatigue, and cognitive functioning. In contrast to very similar versions of these items with attributions to health in general, all KDIS 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”). Although not all traditional fit statistic estimates were satisfactory in each smaller treatment group, overall (n=1123) estimates were satisfactory for establishing unidimensionality and conducting IRT analyses (CFI=0.968, TLI=0.966, RMSEA=0.111 (90% CI=0.109, 0.113, n=1123). No residual correlations were >0.20; the largest residual correlation was 0.1670, mean=0.0429, median= 0.0350.

In support of the 1-factor model, across the groups, 37 KDIS items all loaded highly (>.70) on a single factor, with the exception of one item (avoid traveling) that was slightly lower

in the Dialysis group (.61). Although the CFA fit statistics did not meet some published guidelines for determining good fit, using traditional cutoffs and standards for CFA fit statistics alone is not recommended for establishing the unidimensionality of an item bank [16]. Item banks require items that measure a wide range of the latent trait and CFA fit values are sensitive to data distribution and number of items. In follow-up exploratory factory analyses, the first factor explained the great majority of score variance in each subgroup (76% for both CKD Stage 3-5 and Transplant groups, 65% for Dialysis group). Because of the strength of the size of the first factor, pattern of high factor loadings and very low residual correlations the bank was considered sufficiently unidimensional.

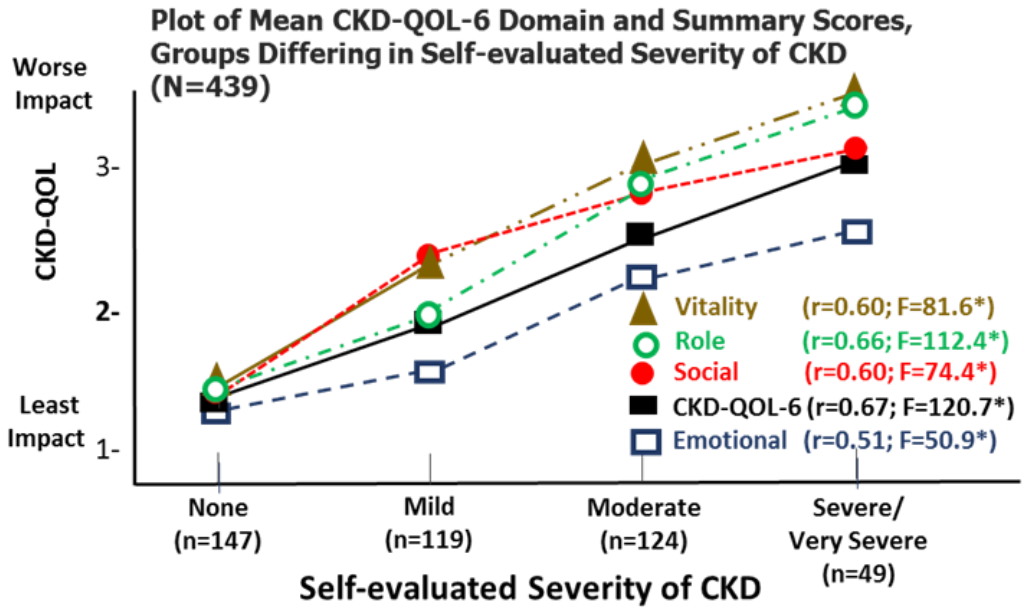
Prior to running the CFA on the 37-item bank, to explore a more comprehensive item bank, a 1-factor confirmatory factor analysis (CFA) was run on a broader set of items [initial KDIS (k=37), additional KDIS items (k=21), KDQOL (k=38), generic role (k=11) and sleep (k=5)] using the aggregate sample. As expected, the full set of items was not unidimensional (CFI=.625; TLI=.953; RMSEA=.205). Symptoms, fluid restriction, dietary restriction, social interaction, and stigma-related items had relatively lower factor loadings (<.60) and accounted for many noteworthy residual correlations (> .20) with other items. As expected, there were some noteworthy residual correlations between the generic and CKD-specific (attribution to kidney disease) versions of role and sleep items. Compared to the generic role and sleep items, higher factor loadings were found for the disease-specific role and sleep items. These results are consistent with the hypothesized distinctions between generic items and their CKD-specific counterparts. We examined several multi-factor solutions that improved fit but were not parsimonious; but they were helpful in evaluating global impact items and identifying and understanding items with more complicated factor content. Items with the best “general factor” impact included those with role, social, mental, fatigue, and cognitive content. These were prioritized and items that did not load highly on the first factor, violated assumptions of unidimensionality, and/or showed local dependence were dropped from the 1-factor model or were evaluated as separate factors. To explore and evaluate construct validity, after investigating various factor solutions, a 5-factor model that included a CKD-specific general QOL impact factor, and factors for symptoms, sleep, sexual functioning, and body image best fit the more comprehensive item set (CFI=.868; TLI=.988; RMSEA=.116).

The correlations between the total bank KDIS QOL impact scale and scales constructed using non-overlapping sleep, sexual functioning, and body image items were substantial as were correlations between the total KDIS scale and KDQOL symptom scales (using developers’ scoring). However, they were clearly best described as separate factors based on content and an evaluation of factor loadings. A set of 34 KDIS items was selected for further item bank development and evaluation, excluding sleep, sexual functioning, and body image items that did not fit a single QOL impact factor. Further, these domains were represented by too few items to evaluate them further as distinct scales. As a result, they were earmarked for study as separate variables in the clinical validation study. Finally, KDQOL symptom items were clearly not appropriate for IRT analysis.

Summary of Results-Comparisons of Domain-specific and Summary Scales Across CKD Severity Groups

In response to concerns regarding information loss and interpretability of a single summary of KDIS (CKD-QOL) scores, mean scores for domain-specific QOL impact ratings and the 6-item/domain summary score were compared across groups differing in self-evaluated severity of CKD symptoms,

Evidence that multiple items cutting across distinct QOL domains can be summarized to achieve a reliable and valid disease-specific QOL impact measure has been published independently for multiple medical conditions, including asthma, Cushing's syndrome, inflammatory bowel disease and Parkinson's disease, in addition to the investigators' findings in headache, asthma, and multiple other medical conditions. As evidence that multiple items cutting across distinct QOL domains can be summarized to achieve a reliable and valid disease-specific QOL impact score, a plot of KDIS item means (1-5 scoring, higher worse) for four KDIS domains (vitality, role, social and emotional) and the 6-item aggregate/summary score (scored on same 1-5 scale; labeled CKD-QOL-6) across groups differing in self-evaluated CKD symptom severity is shown below. Labeling includes correlation of each item with the severity criterion and the F-ratio for mean differences.



Each of the domain-specific items showed the same pattern of differences observed for the summary. Results for other domains, not shown, follow the same pattern. The group separations across domains and the summary are very similar. Correlations with the severity criterion were all very high although slightly lower for items in comparison with the summary, as would be expected for less and more reliable scores. The substantial differences in F-ratios favoring the summary were largely due to greater within group variances for domain items in comparison with the more reliable summary; for example, nearly two-fold greater within group variance for vitality versus summary score. The domain-specific information and contribution to the summary was not lost. Consistent with the literature, Vitality clearly measured worse and the Emotional domain the better average impact levels for the underlying QOL summary variable at all levels of severity; the other domain levels were distributed in between.

These results support the 1-factor model. It should also be noted that with broad content representation disaggregated analyses of specific domains is possible.

Differential Item Functioning

Tests of differential item functioning (DIF) explored systematic errors due to a group bias across subgroups, including: treatment status, dialysis type, age, gender, and race. Ordinal logistic regression models were used and items with statistical significance and an R^2 difference of .03 (proportion of variation explained by model, combining uniform and non-uniform DIF) using Nagelkerke's R^2 [17] was considered a demonstration of DIF. A significant effect of the independent variable on the item response, when controlling for the total sum score, is an indication of uniform DIF. A significant interaction between the independent variable and sum score indicates non-uniform DIF. No DIF was found for age, gender, or race. Uniform DIF was found for KDIS36 (avoid travel). Dialysis patients and those in the most severe stages of CKD reported greater impact across levels of the general impact factor However. we retained the KDIS travel item as it discriminated between the subgroups very well, and was included in other global disease impact item banks. Uniform DIF was also found for KDQOL44 (difficulty concentrating and thinking), with Dialysis respondents reporting less impact across levels of the latent trait.

Item Parameters and Item Fit

IRT statistical models show the relationship between a person's score on the latent construct measured and the probability of choosing each response on each item as a function of the underlying latent variable. Items were calibrated on a common metric using the generalized partial credit model (GPCM) [18] with marginal maximum likelihood estimation [19]. This two-parameter logistic model allows items to vary in their difficulty and in their ability to discriminate between respondents with varying levels of the latent trait. For each item, the GPCM was used to establish slope and threshold parameters, estimated with the Parscale program [20]. The slope parameter is an indication of an item's ability to discriminate between patients with more or less kidney disease impact. The threshold parameters of each item are a marker of kidney disease impact measured by an item. Each threshold represents the point on the kidney disease impact continuum at which there is an equal probability of choosing the two adjacent response choice categories. Item fit to the model was assessed by comparing expected and observed item frequency distributions at varying score levels, and calculating overall fit statistics [21]. IRT Fit analyses were conducted using IRTFIT software [22].

Table 3 summarizes IRT results for KDIS items. Slopes and thresholds ranged from 1.20 to 4.16 and -1.03 to +2.21, respectively. Items with high slopes are more discriminating items; and examination of the threshold parameters provide information on the range of scores covered by the item. For example, KDIS18 (restrict you in performing your usual activities) has a very high slope (4.16) and covers a wide score range (-0.71 to +1.92). For most items, tests of item fit were in the acceptable range, although some showed fit problems and were dropped from the KDIS item bank. The end product was a 34-item Kidney Disease Impact Scale (KDIS) item bank appropriate for IRT-based CAT administrations (CKD-CAT) and also static short-form construction, all transformed (using norm-based scoring) to have a mean of 50 and a standard deviation of 10 in the developmental study (KEEP) sample.

Validity of CKD-specific QOL ratings among adults with comorbid conditions

At the heart of this new approach using QOL impact item content improvements that are closer to that of the SF-36 and other widely-used generic health surveys, is common content with attribution of QOL impact to kidney disease, as opposed to health in general. A crucial assumption is that individuals can make valid attributions to CKD in the presence of multiple chronic conditions (MCC). Because about 2/3 of study participants had 2 or more chart-confirmed MCC, this was a challenge for all CKD-specific measures studied. A weakness of the discriminant tests based on a count of multiple chronic conditions is that they are limited to evidence that the *presence* of MCC did not appear to worsen CKD-specific QOL of ratings. Stronger discriminant tests, which are very rare in the QOL literature, require analyses of concurrent ratings of multiple comorbid conditions and PRO methods for each.

As part of a large Internet-based study of chronically-ill US adults, cited in the table below, samples with CKD and each of 8 comorbidities (diabetes, osteoarthritis, hypertension, seasonal allergies, chronic back problems, hip/knee joint problems, and anemia) were sufficient (sample size>=50 and matched methods) for such tests. The methods and a summary of previously-published results for CKD patients with comorbid conditions are presented here. 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.

As summarized in Table 4, CKD convergent correlations among three different measurement methods (eGFR, severity, kidney-specific QOL impact) were significant and two were substantial (r = 0.39 to 0.72, median = 0.45) (top part of Table 4). Discriminant correlations between CKD QOL impact and QOL impact attributed to each comorbid condition were significantly lower (r=0.02 to 0.49, median = 0.13) in magnitude (bottom part of Table 4). 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 (discriminant correlations bolded in Table 4) 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, current 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.

Summary

The development and psychometric evaluation project yielded a bank of CKD-specific QOL impact items with improved content validity and items with more similar operational definitions to SF-36 and other widely-used generic health surveys. The main differences between the new items and such generic measures are the new item attributions to CKD, as opposed to health in general. Also, in contrast to KDQOL, developmental project results support aggregating CKD-QOL items across content areas into a single QOL summary impact score. Such aggregate QOL impact scoring is consistent with similar approaches used successfully in other therapeutic areas. The new approach is in sharp contrast to the physical and mental

summary scores for SF-12/SF-36 and other comprehensive generic surveys asking very similar QOL impact questions, but with attributions to health in general.

It was clear, after the developmental and psychometric evaluation project, that the new KDIS, in various forms, should be evaluated in comparisons with KDQOL generic and CKD-specific scales using original KDQOL developer scoring. To go beyond psychometric-only evidence, such evaluations should include independent clinical criteria such as treatment group status and CKD severity, to more closely approximate the kind of clinical applications that might use patient-reported CKD-specific QOL outcomes.

Items Used in 6-Item Static and Computerized Adaptive Test (CAT) Administrations

The verbatim content of six items from the original KDIS item bank used to score the static KDIS-6 scale, later published as the CKD-QOL-6 are noted (*) and bolded below. Five other KDIS/CKD-QOL items frequently selected in CAT administrations are also documented below. Five different sets of item response categories (labeled a-e for each item) are documented in the footnote below.

Verbatim Item Content

***In the past 4 weeks, how much did your kidney disease restrict you in performing your usual daily activities? (a)**

In the past 4 weeks, how much of the time did you have difficulty in performing work or other daily activities because of your kidney disease? (b)

***In the past 4 weeks, how much of the time has your kidney disease left you too tired to do work or daily activities? (b)**

In the past 4 weeks, how often did you lie down and rest because of your kidney disease? (c)

***In the past 4 weeks, how often did your kidney disease keep you from enjoying your social activities? (c)**

How often in the past 4 weeks did you miss family, social, or leisure activities because of your kidney disease? (c)

***In the past 4 weeks, how much of the time have you felt fed up or frustrated because of your kidney disease? (b)**

In the past 4 weeks, how much of the time has your kidney disease interfered with how well you dealt with family, friends, and others who are close to you? (b)

***In the past 4 weeks, how much did your kidney disease limit your usual activities or enjoyment of everyday life? (d)**

***In the past 4 weeks, how much of the time did your kidney disease limit your ability to concentrate on work or daily activities? (b)**

In the past 4 weeks, how often did your kidney disease make it difficult for you to focus your attention on other things? (e)

* CKD-QOL-6 and KDIS-6 item

Note: 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. (e) Never, rarely, sometimes, often, very often.

References

1. Cagney KA, Wu AW, Fink NE, Jenckes MW, Meyer KB, Bass EB et al. (2000). Formal literature review of quality-of-life instruments used in end-stage renal disease. American Journal of Kidney Diseases, 36, 327-36.
2. Devine EB, Smith KL, Stehman-Breen C, & Patrick DL (2003). Health-related quality of life assessment in chronic kidney disease. Expert Review of Pharmacoeconomics and Outcomes Research, 3, 89-100.
3. Edgell ET, Coons SJ, Carter WB, Kallich JD, Mapes D, Damush TM et al. (1996). A review of health-related quality-of-life measures used in end-stage renal disease. Clinical Therapy, 18, 887-938.
4. Butt Z, Yount SE, Caicedo JC, Abecassis MM, & Cella D (2008). Quality of life assessment in renal transplant: review and future directions. Clin Transplant, 22, 292-303.
5. Ware JE Jr, Snow KK, Kosinski M, & Gandek B (1993). SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: The Health Institute, New England Medical Center.
6. Hays RD, Kallich JD, Mapes DL, Coons SJ, & Carter WB (1994). Development of the kidney disease quality of life (KDQOL) instrument. Quality of Life Research, 3, 329-38.
7. Saris-Baglama RN, Taubes AL, Richardson MM, Anatchkova M, Turner-Bowker DM, Stevens LA et al. (2007). Health-related quality of life concepts measured in chronic kidney disease research. American Journal of Kidney Diseases 49[4].
8. Bass EB, Jenckes MW, Fink NE, Cagney KA, Wu AW, Sadler JH et al. (1999). Use of focus groups to identify concerns about dialysis. Choice Study. Med Decis Making, 19, 287-295.
9. Richardson MM, Saris-Baglama RN, Anatchkova MD, Stevens LA, Miskulin DC, Turner-Bowker DM et al. (2007). Patient experience of chronic kidney disease (CKD): Results of a focus group study. American Journal of Kidney Diseases 49[4].
10. Saris-Baglama RN, Richardson MM, Anatchkova MD, Turner-Bowker DM, Stevens LA, Miskulin DC et al. (2007). Preliminary steps to develop a computerized adaptive test for measuring health-related quality of life in chronic kidney disease. Quality of Life Research [16], S1.
11. Ware JE Jr, Kosinski M, Turner-Bowker DM, & Gandek B (2002). How to Score Version 2 of the SF-12® Health Survey (With a Supplement Documenting Version 1). Lincoln, RI: QualityMetric Incorporated.
12. Saris-Baglama RN, Clayton KA, Richardson MM, Gannon M, Politoski G, Anatchkova MD et al. (2008). Feasibility of Internet-based data collection with people who have kidney disease. American Journal of Kidney Diseases 51[4].
13. Embretson SE & Reise SP (2000). Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.
14. Ramsay JO (1995). TestGraf - A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data [Computer software]. Montreal: McGill University.
15. Muthen B & Muthen L (1998). MPlus User's Guide. Los Angeles, CA: Muthen & Muthen.
16. Cook KF, Kallen MA, & Amtmann D (2009). Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. Quality of Life Research, 18, 447-460.
17. Nagelkerke NJD (1991). A Note on a General Definition of the Coefficient of Determination. Biometrika, 78, 691-692.
18. Muraki E (1997). A generalized partial credit model. In Handbook of modern item response theory (pp. 153-164). Berlin, Springer.

19. Bock RD & Aitkin M (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. Psychometrika, 46, 443-459.
20. Muraki E & Bock RD (1996). Parscale - IRT based Test Scoring and Item Analysis for Graded Open-ended Exercises and Performance Tasks (Version 3) [Computer software]. Chicago: Scientific Software Inc.
21. Orlando M & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement, 24, 50-64.
22. Bjorner JB, Smith KJ, Stone C, & Sun X (2007). IRTFIT: A macro for item fit and local dependence tests under IRT models. Lincoln, RI: QualityMetric Incorporated.

Tables

Table 1. Focus Group Study Participant Characteristics

Variables	Total N=40 (%)
Gender	
Female	21(52.5%)
Male	19(47.5%)
Age	
18-44	9(22.5%)
45-64	19(47.5%)
65-75	10(25.0%)
>75	2(5.0%)
Ethnicity	
Hispanic or Latino	1(2.5%)
Not Hispanic or Latino	39(97.5%)
Race	
Asian	2(5.0%)
Black or African American	12(30.0%)
White	24(60.0%)
Multi-Racial	1(2.5%)
Not Reported	1(2.5%)
Education Level	
Grade school or some high school	3(7.5%)
High school graduate or GED	10(25.0%)
Some college or technical school	8(20.0%)
College graduate	9(22.5%)
Graduate or professional degree	8(20.0%)
Not Reported	2(5.0%)
Household Income	
<$15,000	6(15.0%)
$15,000-$29,999	9(22.5%)
$30,000-$45,000	5(12.5%)
>$45,000	17(42.5%)

Table 2. Item Calibration Study Sample Characteristics (N=1,236)

Gender	Female	762	62%
	Male	474	38%
Age (years)	18-34	198	16%
	35-44	258	21%
	45-54	328	26%
	55-64	282	23%
	65+	170	14%
Hispanic	Yes	60	5%
Race	African American	141	12%
	White	960	79%
	Asian	45	4%
	American Indian/Alaskan Native	9	1%
	Native Hawaiian/Pacific Islander	5	<1%
	Multi-racial	24	2%
	Other	29	2%
Education Level	8th grade or less	10	1%
	Some high school	22	2%
	High school graduate	149	12%
	Some college or post-HS training	447	36%
	College graduate	341	28%
	Postgraduate education or degree	261	21%
Household Income	Less than $20,000	241	20%
	$20,000 to $45,000	309	26%
	$45,000 to $75,000	304	25%
	More than $75,000	354	29%
Employment	Retired due to age	129	11%
	Retired due to disability	266	22%
	Full-time employed outside home	512	42%
	Part-time employed (19 hours or less)	92	7%
	Unemployed	70	6%
	Full-time homemaker	61	5%
	Full-time student	21	2%
	Other	77	6%
Disease Type	CKD (Non-transplant, Not on Dialysis)	531	43%
	Dialysis	392	32%
	Transplant	313	25%
CKD Staging	Stage 1	18	3%
(Non transplant, Not on Dialysis)	Stage 2	50	9%
	Stage 3	218	41%
	Stage 4	180	34%
	Stage 5	65	12%
Dialysis	Hemodialysis	318	81%
	Peritoneal dialysis	74	19%
CKD Staging Transplant	Stage 1	14	5%
	Stage 2	80	26%
	Stage 3	177	57%
	Stage 4	30	10%
	Stage 5	10	3%

Table 3. KDIS Item Content, IRT Parameters (Slopes, Thresholds) and Total Bank Loadings

Domain	Label	Item Content	Response Choices	Slope	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Factor Loading
Role	KDIS1	In the past 4 weeks, how much did your kidney disease limit your usual activities or enjoyment of everyday life?	Not at all-Extremely	3.14	-0.74	0.18	0.77	1.59	0.906
Role	KDIS2	In the past 4 weeks, how much of the time did your kidney disease limit you in performing your usual daily activities, including housework, work, school, or social activities?	None-All of the time	3.51	-0.66	0.20	0.87	1.70	0.927
Mental	KDIS3	In the past 4 weeks, how much of the time has your kidney disease interfered with how well you dealt with family, friends, and others who are close to you?	None-All of the time	2.28	-0.26	0.46	1.20	1.90	0.863
Mental	KDIS4	In the past 4 weeks, how much of the time have you felt fed up or frustrated because of your kidney disease?	None-All of the time	1.46	-0.75	0.08	0.69	1.46	0.837
Cognitive	KDIS5	In the past 4 weeks, how much of the time did your kidney disease limit your ability to concentrate on work or daily activities?	None-All of the time	2.07	-0.47	0.41	1.06	1.95	0.855
Role	KDIS6	In the past 4 weeks, how much of the time did you have difficulty in performing work or other daily activities because of your kidney disease?	None-All of the time	3.43	-0.48	0.35	0.89	1.75	0.919
Social	KDIS7	In the past 4 weeks, how much of the time have you felt you should avoid social or family activities because of your kidney disease?	None-All of the time	2.12	0.23	0.06	1.34	2.08	0.868
Fatigue	KDIS8	In the past 4 weeks, how much of the time has your kidney disease left you too tired to do work or daily activities?	None-All of the time	2.49	-0.94	0.01	0.65	1.63	0.883
Burden	KDIS9	In the past 4 weeks, how much of the time have you felt like you were a burden on others because of your kidney disease?	None-All of the time	This item is not currently in the KDIS item bank but is the subject of ongoing study					
Role	KDIS10	In the past 4 weeks, how much of the time did your kidney disease keep you from getting as much done at work or at home?	None-All of the time	2.91	-0.86	0.15	0.68	1.64	0.910
Social	KDIS11	In the past 4 weeks, how much of the time did you avoid being around people because of your kidney disease?	None-All of the time	2.21	0.29	0.65	1.31	2.21	0.881
Role	KDIS12	In the past 4 weeks how often did you cancel work or daily activities because of your kidney disease?	Never-Always	2.85	-0.06	0.36	1.18	2.17	0.895

Domain	Label	Item Content	Response Choices	Slope	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Factor Loading
Mental	KDIS13	In the past 4 weeks, how often did your kidney disease make you angry?	Never-Always	1.20	-0.19	-0.08	1.16	1.75	0.818
Fatigue	KDIS14	In the past 4 weeks, how often did you lie down and rest because of your kidney disease?	Never-Always	1.80	-1.03	-0.45	0.49	1.91	0.823
Role	KDIS15	In the past 4 weeks how often did you stop work or other activities to deal with your kidney disease?	Never-Always	2.29	-0.38	0.15	1.02	2.04	0.864
Social	KDIS16	How often in the past 4 weeks did you miss family, social, or leisure activities because of your kidney disease?	Never-Always	2.91	-0.14	0.34	1.24	2.07	0.895
Mental	KDIS17	In the past 4 weeks, how often did you feel that you were going to lose control because of your kidney disease?	Never-Always	1.42	0.39	0.52	1.32	2.15	0.823
Role	KDIS18	In the past 4 weeks, how much did your kidney disease restrict you in performing your usual daily activities?	Not at all-Could not do activities	4.16	-0.71	0.17	0.84	1.92	0.939
Role	KDIS19	In the past 4 weeks, how often was your ability to engage in non-work related activities reduced because of your kidney disease?	Not reduced-Unable to perform activities	3.40	-0.63	0.21	0.88	1.94	0.914
Role	KDIS20	In the past 4 weeks, how often did you restrict your recreational activities because of your kidney disease?	Never-Very often	2.59	-0.64	-0.07	0.7	1.40	0.895
Mental	KD1S21	In the past 4 weeks, how often did you get tense because of your kidney disease?	Never-Very often	1.50	-0.58	0.08	0.95	1.51	0.857
Cognitive	KDIS22	In the past 4 weeks, how often did your kidney disease make it difficult for you to focus your attention on other things?	Never-Very often	1.83	-0.59	0.19	1.08	1.69	0.848

Domain	Label	Item Content	Response Choices	Slope	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Factor Loading
Fatigue	KDIS23	In the past 4 weeks, how often did your kidney disease keep you in bed for all or most of the day?	Never-Very often	1.85	0.34	0.77	1.50	1.83	0.831
Social	KDIS24	In the past 4 weeks, how often did your kidney disease keep you from socializing?	Never-Very often	2.65	-0.09	0.45	1.18	1.69	0.899
Role	KDIS25	In the past 4 weeks, because of your kidney disease, how much of the time did you need help in handling routine daily tasks? (for example, household chores, shopping, or caring for others)	None-All of the time	1.79	0.03	0.55	1.24	1.96	0.824
Role	KDIS26	In the past 4 weeks, how much of the time has your kidney disease interfered with your leisure time activities, such as reading or exercising?	None-All of the time	2.30	-0.33	0.35	1.00	1.95	0.864
Mental	KDIS27	In the past 4 weeks, how much of the time have you been afraid of letting others down because of your kidney disease?	None-All of the time	1.39	0.14	0.31	0.87	1.55	0.813
Role	KDIS28	How much of the time in the past 4 weeks was your productivity at work, school, or other usual daily activities reduced by half or more because of your kidney disease?	None-All of the time	2.68	0.09	0.46	0.85	1.62	0.906
Mental	KDIS29	In the past 4 weeks, how much of the time did you feel irritable because of your kidney disease?	None-All of the time	1.62	-0.52	0.43	1.05	1.68	0.864
Mental	KDIS30	In the past 4 weeks, how much of the time did you feel frustrated because of your kidney disease?	None-All of the time	This item is not currently in the KDIS item bank but is the subject of ongoing study					
Role	KDIS31	In the past 4 weeks, how much of the time did your kidney disease limit your ability to work, study, or do chores?	None-All of the time	3.48	-0.55	0.31	0.85	1.71	0.927
Role	KDIS32	In the past 4 weeks, how much of the time did your kidney disease make simple tasks hard to complete?	None-All of the time	2.94	-0.16	0.42	1.02	1.82	0.904

Domain	Label	Item Content	Response Choices	Slope	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Factor Loading
Role	KDIS33	How often in the past 4 weeks did you miss work, school, or other daily activities because of your kidney disease?	Never-Always	2.37	0.03	0.43	1.21	1.70	0.880
Social	KDIS34	In the past 4 weeks, how often did your kidney disease keep you from enjoying your social activities?	Never-Always	3.43	-0.33	0.19	1.01	1.82	0.919
Mental	KDIS35	In the past 4 weeks, how often did you feel desperate because of your kidney disease?	Never-Always	1.38	0.29	0.38	1.26	1.71	0.837
Role	KDIS36	In the past 4 weeks, how often did you avoid traveling because of your kidney disease?	Never-Always	1.25	0.47	0.09	0.87	1.25	0.796
Social	KDIS37	In the past 4 weeks, how often did your kidney disease place stress on your relationships with family or friends?	Never-Always	This item is not currently in the KDIS item bank but is the subject of ongoing study					

Table 4. Chronic Kidney Disease (Stages 3-5): Convergent correlations among three CKD-specific measurement methods in comparison with discriminant correlations for 2-3 disease-specific measures for each of eight prevalent comorbid conditions

	Chronic Kidney Disease (CKD)		
Measures and Conditions	CKD1	CKD2	CKD3
CKD (N=147)			
CKD1 Glomerular filtration rate[a]			
CKD2 CKD severity	**-0.45**		
CKD3 CKD QOL impact	**-0.39**	**0.72**	
Obesity (N=50)			
Ob1 Body mass index	0.02	0.20	0.23
Ob2 Obesity severity	-0.05	0.26	**0.40***
Ob3 Obesity QOL impact	0.02	0.22	0.28
Diabetes (N=56)			
Di2 Diabetes severity	-0.13	0.24	0.20
Di3 Diabetes QOL impact	-0.23	0.24	0.34
Osteoarthritis (N=57)			
OA2 OA severity	0.09	-0.01	0.13
OA3 OA QOL impact	0.05	-0.06	0.08
Allergies-seasonal (N=60)			
As2 Allergy-seasonal severity	-0.18	0.21	**0.32***
As3 Allergy-seasonal QOL impact	-0.24	**0.32***	**0.39***
Anemia, History of (N=56)			
An2 Anemia severity	-0.30	**0.49***	**0.48***
An3 Anemia QOL impact	-0.22	**0.37***	**0.46***
Chronic back problems (N=51)			
Bp2 Back severity	**-0.37***	**0.37***	**0.37***
Bp3 Back QOL impact	**-0.33***	**0.41***	**0.48***
Hypertension (N=121)			
Hy2 Hypertension severity	-0.07	0.25	0.12
Hy3 Hypertension QOL impact	-0.02	0.17	0.20
Joint problems, hip/knee (N=59)			
Hk2 Hip/knee severity	-0.04	0.00	0.18
Hk3 Hip/knee QOL impact	-0.08	0.02	0.27

Convergent correlations

Discriminant Correlations

Higher scores on self-evaluated severity and QOL impact items indicate greater disease severity or greater disease-specific QOL impact.

[a] Glomerular filtration rate defined in Table 1. Lower GFR indicates worse kidney function.

***** Discriminant correlations are between CKD-specific methods and measures of comorbid conditions using the same or different methods) r>0.31, absolute value, (CKD1-CKD3 columns) defined failed convergent-discriminant 1-tailed tests. Median convergent
r=-0.45 (across three CKD-specific methods shown in triangle of CKD1-CKD3 correlations) was used as convergent estimate in comparisons with discriminant correlations below it.
Source: Ware JE, Gandek B, Allison J., The validity of disease-specific quality of life attributions among adults with multiple chronic conditions. *International Journal of Statistics in Medical Research* 2016; 5(1):17-40. (Please see Table 1 for definitions of variables and see Methods for explanations measures used in convergent-discriminant tests of validity.
Briefly, severity is self-evaluated (5-category item) disease-specific rating; QOL impact is disease-specific evaluation using 5-category QDIS global QOL impact item, from the improved version of CKD-QOL item bank standardized (content and scoring) normed (mean=50, SD=10) in US chronically-ill population. For latter, please see Table X1 reference.