Comparison of health-related quality of life measures for chronic renal failure: quality of well-being scale, short-form-6D, and the kidney disease quality of life instrument

**Karen L. Saban · Kevin T. Stroupe · Fred B. Bryant ·
Domenic J. Reda · Margaret M. Browning ·
Denise M. Hynes**

Abstract

Background Studies have found that health-related quality of life (HRQOL) measurements with different conceptual bases yield widely varying results within the same study sample. Using data from a cohort of patients with chronic kidney failure, the purpose of this study was to compare the Quality of Well-Being Scale-Self-Administered (QWB-SA), the Short-Form-6D (SF-6D), and the Kidney Disease Component Summary (KDCS).

Methods Baseline data from a multi-site prospective observational study of 322 veterans receiving hemodialysis were analyzed. Descriptive statistics were calculated. Confirmatory factor analysis was conducted to determine how closely the three HRQOL tools reflected the same underlying construct.

Results Our confirmatory factor analysis offered strong evidence that the subscales of the QWB-SA, SF-6D, and 7-subscale KDCS measured more than one factor in this study sample. In the three-factor model, the SF-6D and 7-subscale KDCS correlated .911 ($P < .05$), indicating 83% of the variance in the 7-subscale KDCS was correlated with the SF-6D. However, a two-factor model, in which the highly correlated SF-6D and 7-subscale KDCS were combined, fit the data almost as well as the three-factor model.

Conclusion The three HRQOL measures addressed different underlying HRQOL constructs in this sample. The QWB-SA was significantly different from the SF-6D and KDCS.

K. L. Saban (✉) · K. T. Stroupe · M. M. Browning ·
D. M. Hynes
Center for Management of Chronic Complex Care, Edward
Hines Jr. VA Hospital, P.O. Box 5000 (151H), Hines,
IL 60141, USA
e-mail: Karen.Saban@va.gov

K. L. Saban · D. M. Hynes
Loyola University Chicago Niehoff School of Nursing,
Chicago, IL, USA

K. T. Stroupe · M. M. Browning · D. M. Hynes
Veterans Affairs Information Resource Center, Hines,
IL, USA

K. T. Stroupe
Northwestern University Feinberg School of Medicine, Chicago,
IL, USA

F. B. Bryant
Department of Psychology, Loyola University Chicago,
Chicago, IL, USA

D. J. Reda
Veterans Affairs Cooperative Studies Program Coordinating
Center, Hines, IL, USA

D. M. Hynes
Loyola University Chicago Stritch School of Medicine,
Maywood, IL, USA

Abbreviations

CFA	Confirmatory factor analysis
CKD	Chronic kidney disease
CRF	Chronic renal failure
CFI	Comparative fit index
EQ-5D	EuroQol instrument
GFI	Goodness-of-fit index
HRQOL	Health-related quality of life
HUI	Health utilities index

ICED	Index of coexistent disease
IPI	Index of physical impairment
KDCS	Kidney disease component summary
KDQOL	Kidney disease quality of life instrument
KDQOL-SF	Kidney disease quality of life instrument-short form
NFI	Normed fit index
NNFI	Non-normed fit index
QALYs	Quality-adjusted life years
QOL	Quality of life
QWB	Quality of well-being scale
QWB-SA	Quality of well-being scale-self-administered
RMSEA	Root-mean-square error of approximation
SF-6D	Short-form-6D
SF-12	Short-form-12
SF-36	Short-form-36
SRMR	Standardized root-mean-square residual

Introduction

According to the United States Renal Data System, the prevalence of chronic kidney disease (CKD) continues to grow each year, with the incidence of patients receiving hemodialysis reaching 310 per million population in 2004 [1].[1] In addition, annual Medicare expenditures for outpatient hemodialysis cost nearly US$6.7 billion in 2004, a 9.9% increase from 2003 [1]. As the prevalence and costs of hemodialysis continue to increase, accurate evaluation of treatment outcomes in CKD patients becomes increasingly important, not only in terms of economic burden but also on how this complex chronic illness affects individuals' quality of life.

The measurement of health-related quality of life (HRQOL) has become increasingly common in recent years as an important indicator of health and well-being. Health-related quality of life outcome data are frequently used to determine healthcare effectiveness, including medication and procedural treatment effects as well as resource allocation and policy development [2, 3]. For example, utility HRQOL measures generate a single summary score that can be translated into quality-adjusted life years (QALYs). Theoretically, QALYs provide a standard metric by which results can be compared across studies using any utility or preferenced-based measure [4]. QALYs can have significant implications in clinical practice in that they allow comparison analysis of various treatments regardless of the HRQOL outcome measure used. However, despite the significant role

that HRQOL measurement has in health care and economic evaluation, a lack of clarity remains in how HRQOL is measured and results interpreted.

Health-related quality of life is an important issue for patients with CKD receiving hemodialysis. Hemodialysis, although not a cure for CKD, helps prolong and improve patients' quality of life [5]. Hemodialysis requires that patients be connected to a dialysis machine several hours a day at least three times a week, during which time they are essentially immobile. Social activities, physical functioning, and mental health are impacted due to the constraints of hemodialysis as well as from the effects of the CKD, which can include fatigue and nausea. A number of studies have demonstrated that perceived HRQOL of patients receiving hemodialysis is significantly impaired [6–9].

Although many HRQOL tools have well-documented validity and reliability, each takes a different approach to measuring the highly complex construct of HRQOL. These differences may lead to conflicting results, depending upon which tool is used. In a review of the literature over the past 10 years, several studies found that different HRQOL measurements, both profile and utility tools, yielded widely varying results within the same study sample (Appendix). Relatively few studies have compared HRQOL measurements in patients receiving hemodialysis. One study compared the EQ-5D and the SF-6D in the CKD population [10]. In this study, the authors did not find any significant differences between the two utility measures; however, they concluded that the EQ-5D was the preferred measurement tool because it had a higher response rate than the SF-6D. Another study compared HRQOL utility measures with the disease-specific HRQOL measures, the Kidney Disease Quality of Life Instrument (KDQOL), and found moderate correlation between the HUI3 and KDQOL [11]. Although the Quality of Well-Being Scale-Self-Administered (QWB-SA), SF-6D, and the Kidney Disease Quality of Life Instrument-Short Form (KDQOL-SF) have been used in a number of studies of patients receiving hemodialysis [12–14], no studies were found that compared these HRQOL measures in this population. Given the significant impact that hemodialysis has on HRQOL, as well as the important role that measurement of HRQOL has in health-care decision making and allocation of resources, further evaluation of HRQOL measures in this chronically ill population is important.

The primary objective of this study was to test the hypothesis, using confirmatory factor analysis, that the QWB-SA, SF-6D, and the Kidney Disease Component Summary (KDCS; a subscale of the KDQOL-SF) measured the same construct, HRQOL, in a sample of patients receiving hemodialysis for CKD. Based upon previous research comparing other HRQOL measures, we expected to find these tools measured different aspects of HRQOL.

---

[1] The data reported here have been supplied by the United States Renal Data System. The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy or interpretation of the U.S. government.

We will explore how different HRQOL measures used within the same sample may produce conflicting results and how this issue may be addressed in future studies. Lastly, because the KDCS is a fairly new reported measure and was revised for this study, a secondary objective of our study was to examine the components of the KDCS and refine as necessary.

Methods

Study design

Data for this investigation were from a large prospective observational study comparing outcomes and costs of care of veterans dialyzing at U.S. Department of Veterans Affairs facilities or in the private sector [15]. Any patient who had received care at a Veterans Affairs facility within the prior 3 years and was receiving hemodialysis for CKD was eligible for enrollment. Patients were excluded if they: (1) had a live kidney donor identified; (2) required care in a skilled nursing facility; (3) had a life expectancy of less than 1 year, determined by a nephrologist; (4) were cognitively impaired; (5) had a severe speech or hearing impairment; (6) were not fluent in English; and (7) had no access to a telephone for follow-up contact.

Patients were recruited from eight Veterans Affairs medical centers with outpatient dialysis facilities. Institutional review board approval was obtained from all Veterans Affairs sites. Coordinators at each site explained the study and obtained written informed consent from patients who were interested in participating. A total of 364 patients consented to participate in the study with 322 subsequently completing baseline questionnaires.

Measures

The three HRQOL measures used in this study—QWB-SA, SF-6D, and KDQOL—are summarized in Table 1.

Quality of well-being scale-self-administered 1.4

The QWB-SA [16] is a preference-based measurement of quality of life and is derived from the longer, more complex, Quality of Well-Being Scale (QWB) [17]. The QWB-SA, a 76-item questionnaire, assesses objective level of functioning in three areas: mobility, physical activity, and social activity. In addition, the QWB-SA measures the presence of 58 acute and chronic symptoms. The scores are combined to form one preference score on a continuum scale between .30 (for death) to 1.0 (for perfect health) [18]. In addition to one overall preference score, scores for the following four, preference-weighted subscales can also be calculated: (1) mobility scale, (2) physical activity scale, (3) social activity scale, and (4) the symptom/problem complexes scale. However, few studies have reported the subscales of the QWB. The QWB-SA has been used in several studies with a wide variety of disease entities, including migraines, diabetes, and posttraumatic stress disorder [19–21]. Test–retest reliability of the original QWB ranged from .83 to .98 [18]. Reliability and validity of the QWB-SA are similar to the original QWB [16].

Table 1 Characteristics of the QWB-SA, SF-6D, and KDQOL instruments

Characteristics	QWB-SA	SF-6D	KDCS disease-specific subscales[a]
Domains	Symptoms	Physical functioning	Symptoms/problems
	Problems	Role limitation	Effects of kidney disease on daily life
	Mobility	Social Pain functioning	Quality of social interaction
	Physical activity	Mental health	Burden of kidney disease
	Social activity	Vitality	Cognitive function
			Work status
			Sexual function
			Social support
			Sleep
			Staff encouragement
			Patient satisfaction
Severity levels	Four levels indicating which days (if any) problem was experience—e.g., "no days" to "3 days ago"	Four to six ordinal or categorical levels—e.g., "no limitation" to "limited a lot"	Five ordinal levels—e.g., "not bothered at all" to "extremely bothered"
Minimum/maximum score	.09–1.00	.30–1.00	0–100

[a] KDQOL also includes the SF-36 from which the SF-6D was calculated

Short-form-6D

The SF-6D is a preference-based HRQOL measure that can be derived from either the Short-Form-36 (SF-36) or Short-Form-12 (SF-12) [22–24]. Both the SF-36 and SF-12 are measures of quality of life and comprise eight, health-related dimensions: (1) physical functioning, (2) role limitations related to physical functioning, (3) bodily pain, (4) general health perceptions, (5) vitality, (6) social functioning, (7) role limitation related to emotional problems, and (8) mental health [25, 26]. Using an algorithm based upon preference weights, these eight dimensions are condensed into six dimensions (general health is omitted, and role limitation related to emotional problems and role limitations related to physical functioning are combined) and a single score is calculated to produce the SF-6D index score [22, 23, 27]. The SF-6D index score ranges on a continuum from .296 (most impaired) to 1.0 (full health) and can be translated into QALYs [28]. Although fairly new, the SF-6D has been used in a number of studies with a wide range of patient populations [29–32]. Studies have demonstrated good test–retest reliability and discriminant validity of the SF-6D [32, 33]. For this study, reliability was good (Cronbach's alpha = .702).

Kidney disease component summary

The KDCS is a subscale of the KDQOL questionnaire. The KDQOL was developed as a self-report, health-related quality of life measurement tool designed specifically for patients with CKD [34]. The 134-item KDQOL was later condensed into the 80-item Kidney Disease Quality of Life Instrument-Short Form (KDQOL-SF) [35]. The questionnaire consists of the generic SF-36 [25] as well as multi-item scales focused on quality of life issues specific to patients with kidney disease. The kidney-disease-specific subscales are listed in Table 1. All subscales are scored on a 0–100 scale, with higher numbers representing better HRQOL. A few studies were found in which scores from the 11 kidney-disease-specific subscales were averaged to form a KDCS [36–39]. However, no literature was found describing the psychometrics of the KDCS. The KDQOL-SF has been widely used in several studies of patients with kidney disease, including the ongoing, international Dialysis Outcomes and Practice Patterns Study (DOPPS) [13, 36, 40–44], and has demonstrated good test–retest reliability on most dimensions.

Statistical analyses

Baseline data from the QWB-SA and SF-6D were converted into utility scores based upon the scoring algorithms from Kaplan et al. [16] and Brazier and Roberts [24],

respectively. Preference-weighted single attribute scores were used as indicators of the domains of the SF-6D and QWB-SA. Missing data (<5%) were replaced in the SF-6D with values using a multiple imputation procedure [45]. All other data were complete (n = 322), except for the KDCS subscale for sexual activity, which was completed by only 55 participants. Therefore, the sexual activity subscale of the KDCS was not included in the calculation of the KDCS, resulting in 10 kidney-disease-specific subscales of the KDQOL-SF generating the KDCS for this study. Descriptive statistics (mean, 95% confidence intervals, median, and range) and Pearson correlations of the QWB-SA, SF-6D, and KDCS were calculated using SAS version 9.1 (SAS Institute, Cary, NC, USA). Ceiling and floor effects for each instrument were compared by determining the highest and lowest scores for each tool.

One way to assess the degree of commonality and uniqueness across the three HRQOL instruments is merely to inspect the Pearson correlations among scores on the three instruments. However, because unreliability in measurement attenuates (i.e., dilutes) the strength of observed correlations [46, 47], differences in the reliabilities of the three instruments will produce differences in the apparent degree of relationship among the instruments, thereby yielding spurious conclusions. Confirmatory factor analysis overcomes this statistical problem by adjusting correlations among the instruments for measurement error (i.e., unreliability), thereby disattenuating correlations and providing unbiased estimates of the degree of interrelationship.

Maximum-likelihood confirmatory factor analysis was completed using LISREL 8.8 [48] in three phases: (1) analysis of each instrument (QWB-SA, SF-6D, and KDCS) separately using a one-factor model representing HRQOL containing each instruments' respective subscales; (2) consideration of a one-factor model combining the subscales of the QWB-SA, SF-6D, and KDCS, in which the one factor represented HRQOL (Fig. 1); and (3) analysis of a three-factor model containing the subscales of the QWB-SA, SF-6D, and KDCS, in which each factor represented its corresponding HRQOL measure (Fig. 2). Models were refined as necessary. Factors were allowed to intercorrelate.

In addition to the overall test of fit (the χ^2 statistic), six fit indices were used to assess model adequacy: (1) the goodness-of-fit index (GFI) [49]; (2) the root-mean-square error of approximation (RMSEA) [50]; (3) the standardized root-mean-square residual (SRMR) [51]; (4) the comparative fit index (CFI) [52]; (5) the normed fit index (NFI) [53]; and (6) the non-normed fit index (NNFI) [53]. Acceptable standards of fit for the absolute fit indices, which indicate how well the data fit the theoretically proposed model, are: GFI >.90 [53], RMSEA ≤.08 [54, 55], and SRMR ≤.05 [54, 55]. Browne and Cudeck [56] suggest that RMSEA values ≤.05 indicate close fit, and RMSEA
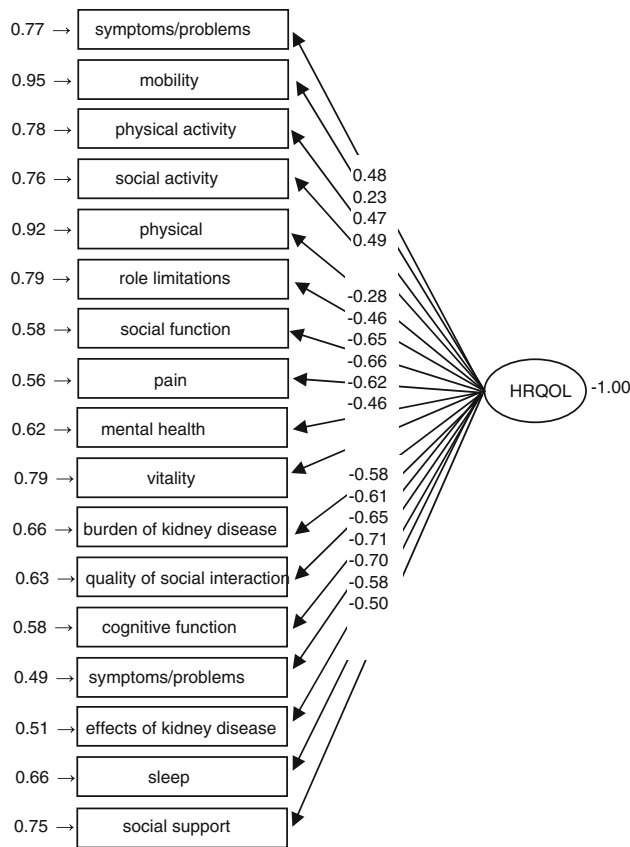
Fig. 1 Completely standardized solution path diagram of one-factor model of HRQOL measures. Standardized factor loadings of one-factor model. Chi-square = 364.86, df = 119. *Boxes* represent HRQOL subscales from the QWB-SA, SF-6D, and KDCS. *Long arrows* include factor loadings and *short arrows* represent measurement error
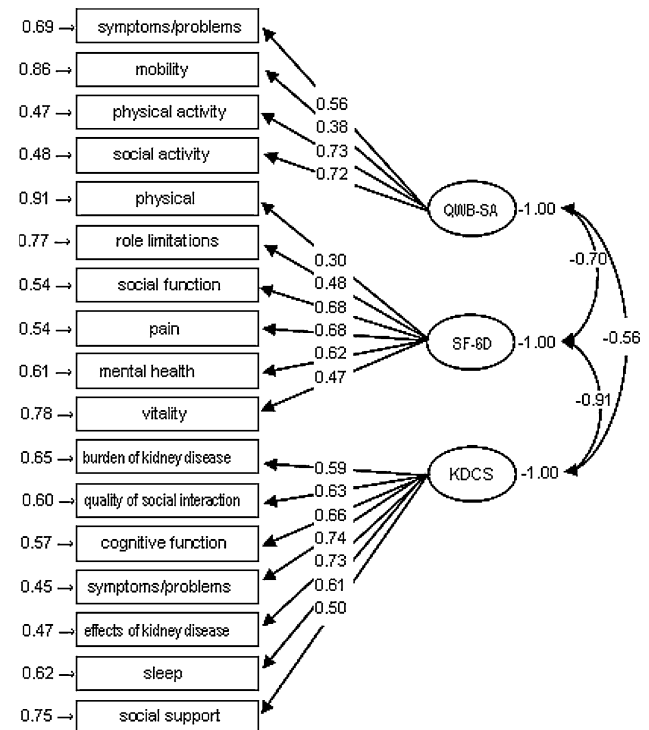

Fig. 2 Completely standardized solution path diagram of three-factor model of HRQOL measures. Standardized factor loadings of three-factor model. Chi-square = 241.53, df = 116. *Boxes* represent HRQOL subscales from the QWB-SA, SF-6D, and KDCS. *Long arrows* include factor loadings and *short arrows* represent measurement error. *Curved arrows* represent factor intercorrelations

values between .05 and .08 indicate adequate fit. The CFI, NFI, and NNFI are relative or incremental fit indices and compare the theoretical model to the null model in which all observed variables are uncorrelated [53]. A fit index above .90 for the CFI, NFI, and NNFI is considered to be acceptable [53].

Results

Table 2 summarizes the demographic and clinical characteristics of the sample. Consistent with the veteran population, the majority of participants were male. Participants had been receiving hemodialysis for an average of 2.50 years, and most had at least one comorbidity with the most common ones being hypertension (84%), diabetes (63%), and congestive heart failure (46%).

There were 322 QWB-SA, SF-6D, and KDCS forms available for analysis. Summary statistics for each HRQOL measure are reported in Table 3 and correlations between instruments are reported in Table 4. None of the HRQOL measures revealed ceiling or floor effects.

Confirmatory factor analysis

Phase 1: one-factor analyses of each individual instrument

In the first phase of the confirmatory factor analysis, the subscales of each instrument—QWB-SA, SF-6D, and KDCS—were considered in three separate one-factor models representing HRQOL. As expected, for both the QWB-SA and SF-6D, the one-factor structures fit the data very well with all six fit index thresholds met (Table 5). However, the one-factor structure was a poor model for the initial set of 10 KDCS subscales, with all fit indices outside recommended thresholds of acceptability (Table 5). Completely standardized factor loadings, in which both the items and the factors were standardized to have variances of 1.0, were strong for the 10-item KDCS, ranging from .494 to .740, except for three subscales—patient satisfaction (.312), work status (.233), and dialysis staff encouragement (.205)—suggesting that these subscales measured something other than HRQOL. Work status may have loaded weakly because most participants (91.6%) in this study were not working. Patient satisfaction and dialysis staff encouragement seem to measure satisfaction with

Table 2 Demographics of sample

Variable	n = 322
Age (years)	
Mean	62.12
Range	28–85
Standard deviation	11.31
Gender n (%)	
Female	8 (2.5)
Male	314 (97.5)
Marital status n (%)	
Married	156 (48.7)
Single	38 (11.9)
Divorced/separated	89 (27.8)
Widowed	37 (11.6)
Race n (%)	
White	156 (48.4)
Black	155 (48.1)
Other	7 (2.2)
Education n (%)	
Less than high school	60 (18.7)
Completed high school/trade school	74 (23.1)
Some college	141 (44.1)
Completed college	35 (10.9)
Graduate work	10 (3.1)
Employed n (%)	27 (8.4)
Annual income n (%)	
$0–10,000	76 (23.6)
$10,000–20,000	104 (32.3)
$20,000–30,000	66 (20.5)
>$30,000	65 (20.2)
Not reported	11 (3.4)
Years since beginning dialysis	
Mean	2.50
Median	1.76
Range	0.01–22.99
Selected comorbidities n (%)	
Congestive heart failure	147 (45.8)
Diabetes	202 (62.7)
Complications of diabetes	136 (42.2)
Hypertension	270 (83.9)
Coronary artery disease	82 (25.5)
Drug dependence	25 (7.8)

care rather than HRQOL. Therefore, the three subscales of work status, patient satisfaction, and dialysis staff encouragement were dropped from the KDCS, resulting in 7 subscales. Subsequently, the 7-subscale KDCS one-factor model fit the data reasonably well, with 4 out of the 6 fit index thresholds met (Table 5). The 7-subscale KDCS one-factor model fit the data significantly better than the 10-subscale KDCS in this sample. Therefore, the

7-subscale KDCS was used in subsequent phases of the confirmatory factor analysis.

We computed the reliability of both the 10-subscale and 7-subscale versions of KDCS total score using Mosier's [57] formula, which requires a reliability estimate for each subscale that is combined to construct the composite total score. For each multi-item subscale, at least two reliability estimates were available: the Cronbach's alpha (α) reliability coefficient and the squared multiple correlation (R^2) from the one-factor CFA solution. Whereas the former reflects the average inter-item correlation for each subscale as an index of internal consistency, the latter represents the proportion of variance in the subscale that the underlying KDCS factor explains as an index of measurement reliability. Weighting each subscale equally, we computed separate Mosier reliability estimates using both α and R^2, although we were able to use only R^2 as a reliability estimate for the single-item measure of patient satisfaction (which is part of the 10-subscale version of KDCS total score). These Mosier reliability estimates for KDCS total score were as follows: 10-subscale version (using α: .88; using R^2: .77); 7-subscale version (using α: .91; using R^2: .81). Note that, although both KDCS total scores achieved acceptable reliability, Mosier reliabilities were noticeably higher for the 7-subscale version of KDCS total score, compared to the 10-subscale version; and in each case, the estimated reliability of the composite total score was lower when using R^2 versus α, reflecting the fact that R^2 was lower than α for each multi-item subscale.

Phase 2: one-factor analysis combining subscales of QWB-SA, SF-6D, and 7-subscale KDCS

The second phase of the confirmatory factor analysis comprised evaluating a one-factor model imposed on the data for the subscales of the three HRQOL measures: QWB-SA, SF-6D, and 7-subscale KDCS (Fig. 1). The one-factor structure proved to be a poor fit to the data (Table 6). Although the CFI, NFI, and NNFI suggested reasonably close fit, the GFI, RMSEA, and SRMR were outside recommended thresholds of acceptability. These findings suggested that, although the one-factor model came close to fitting the data, the combined subscales of the three HRQOL instruments measured more than one factor.

Phase 3: three-factor analysis combining QWB-SA, SF-6D, and 7-subscale KDCS

Because the originally proposed one-factor model did not demonstrate uniformly acceptable fit, a three-factor model was imposed on the data for subscales of the three HRQOL measures (Fig. 2). The three-factor model fit the data well, with 5 of the 6 fit indices within acceptable ranges (Table 6), and explained between 30 and 74% of the

Table 3 Summary statistics for each instrument ($n = 322$)

	QWB-SA	SF-6D	KDCS-10 subscales[a]
Mean (95% confidence intervals)	.49 (.48–.51)	.69 (.68–.71)	.65 (.64–.67)
Median	.50	.69	.65
Range	.13–.93	.37–1.0	.25–.95

[a] Sexual activity subscale not included in calculation of KDCS because of missing data. The KDCS scores were transformed to a 0–1.0 scale from a 0–100 scale for ease of comparison

Table 4 Pearson correlation coefficients of QOL variables and index of physical impairment

	QWB-SA	SF-6D	10-subscale KDCS	7-subscale KDCS	IPI
QWB-SA	1	–	–	–	–
SF-6D	.500**	1	–	–	–
10-subscale KDCS	.455**	.696**	1	–	–
7-subscale KDCS	.467**	.722**	.939**	1	–
IPI	−.460**	−.227**	−.120*	−.122*	1

* Correlation is significant at the .05 level (2-tailed)

** Correlation is significant at the .01 level (2-tailed)

The 10-subscale KDCS includes subscales of patient satisfaction, burden of kidney disease, quality of social interaction, cognitive function, symptoms and problems associated with renal disease, effects of kidney disease on daily life, sleep, social support, work status, and dialysis staff encouragement

The 7-subscale KDCS includes all of the above subscales except for patient satisfaction, work status, and dialysis staff encouragement

Table 5 Fit statistics for each individual instrument ($n = 322$)

Model	χ^2	df	GFI	RMSEA	SRMR	CFI	NFI	NNFI
Threshold			>.90	≤.080	≤.050	>.90	>.90	>.90
One-factor model with 4 subscale QWB-SA	2.33	2	**.996**	**.024**	**.017**	**.999**	**.991**	**.996**
One-factor model with 6 subscale SF-6D	7.709	9	**.992**	**.00**	**.0259**	**1.000**	**.983**	**1.005**
One-factor model with 10-subscale KDCS	207.64	35	.889	.121	.081	.876	.855	.840
One-factor model with 7-subscale KDCS	87.89	14	**.925**	.130	.059	**.935**	**.924**	**.903**

Bold figures represent goodness-of-fit index within acceptable range

Table 6 Fit statistics for HRQOL measurement models ($n = 322$)

Model	χ^2	df	GFI	RMSEA	SRMR	CFI	NFI	NNFI
Threshold			>.90	≤.080	≤.050	>.90	>.90	>.90
One-factor model with QWB-SA, SF-6D, and 7-subscale KDCS	364.86	119	.872	.086	.064	**.940**	**.914**	**.932**
Three-factor model with QWB-SA, SF-6D, and 7-subscale KDCS	241.53	116	**.920**	**.056**	.052	**.970**	**.943**	**.964**
Two-factor model with QWB-SA and combined SF-6D and 7-subscale KDCS	264.65	118	**.911**	**.062**	.056	**.964**	**.938**	**.959**

Bold figures represent goodness-of-fit index within acceptable range

variance in the individual subscales (Fig. 2). The SRMR of .052 was near the indicated threshold of acceptability of ≤.05. Thus, the three-factor model fit the data significantly better than the one-factor model, as evidenced by a difference in the χ^2 of 123.3 between the two models as well as more acceptable goodness-of-fit indices.

In the completely standardized solution for this three-factor model (Fig. 2), the SF-6D and 7-subscale KDCS correlated .911 ($P < .05$), indicating that 83% (i.e.,

$.91^2 = .83$) of the variance in the 7-subscale KDCS was correlated to the SF-6D, and vice versa. As expected, factor correlations between the HRQOL measures were higher than their corresponding Pearson correlations (Table 4).

Refinement of model: two-factor analysis

The strong correlation between the SF-6D and 7-subscale KDCS (.911) suggested that a two-factor model combining
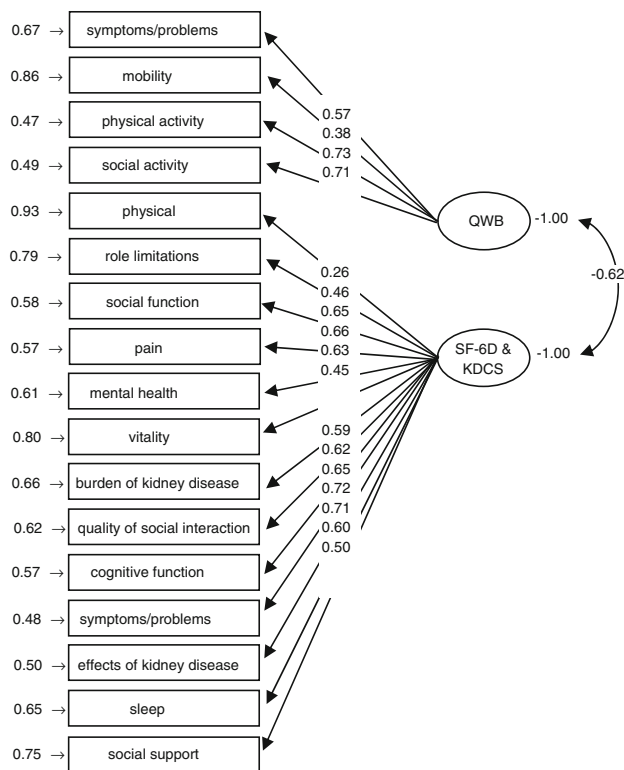
Fig. 3 Completely standardized solution path diagram of two-factor model of HRQOL measures. Standardized factor loadings of two-factor model. Chi-square = 264.65, df = 118. *Boxes* represent HRQOL subscales from the QWB-SA, SF-6D, and KDCS. *Long arrows* include factor loadings and *short arrows* represent measurement error. *Curved arrows* represent factor intercorrelations

the SF-6D and KDCS might fit the data. Therefore, a two-factor model with the combined measures (SF-6D and 7-subscale KDCS) and the QWB-SA was imposed on the data (Fig. 3). This two-factor model came close to fitting the data with 5 of the 6 fit indices being within acceptable ranges (Table 6). The SRMR of .056 was just slightly higher than the recommended threshold of ≤.050. Thus, the two-factor model fit the data reasonably well and was more parsimonious [58].

Discussion

In regards to our hypothesis that the subscales of the QWB-SA, SF-6D, and the KDCS measured the same construct, we found evidence to the contrary. We determined that the QWB-SA measured a related but different construct than the SF-6D and KDCS. This was evidenced by only 38% (i.e., $.62^2 = .38$) of the variance shared between the combined SF-6D and KDCS with the QWB-SA compared to 83% of the variance shared between the SF-6D and KDCS.

Our CFA offered strong evidence that the QWB-SA, SF-6D, and 7-subscale KDCS measured more than one

factor in this study sample. A three-factor model fit the data well. However, a two-factor model, in which the highly correlated SF-6D and 7-subscale KDCS were combined, fit the data almost as well as the three-factor model and was more parsimonious. The QWB-SA and SF-6D, were found to measure different, but similar, constructs in this study.

One would have expected that the QWB-SA and the SF-6D, both generic, preference-weighted tools, would have been more similar to each other than the SF-6D with the KDCS, a disease-specific, non-preference-weighted instrument. The difference we found between the QWB-SA and SF-6D may be related to variances in preference valuation techniques, measured health states, and conceptual bases. Brazier and Roberts [24] reported that, in the SF-6D model, mental health appeared to be the second most important dimension in determining health states, following bodily pain. The findings of this study support those found in a meta-analysis which distinguished quality of life measures and health status in quality of life research [59]. The meta-analysis concluded that tools such as the QWB, that do not measure psychological functioning (deemed an important component of QOL), may be considered to be measures of perceived health rather than QOL [59]. A significant implication of the conceptual differences between these tools is that QALYs derived from these different tools cannot be compared.

Although no other studies that specifically compared the QWB-SA, SF-6D, and the KDCS were found, the finding that these HRQOL measures were not interchangeable is consistent with other studies that have compared various HRQOL measures in different patient populations (Appendix). The plethora of HRQOL measures with little standardization has been repeatedly addressed over the years [60–62]. Suggestions to improve HRQOL studies have included providing details in articles regarding why a particular HRQOL measure was chosen over others to be used in the study [62] and standardization of HRQOL tools for specific populations [63]. Although some of these recommendations have been implemented, such as the coordination of task forces to standardize instrumentation within a specific patient population (for example, the DOPPS [44]), more work is needed to be able to reliably compare results of HRQOL studies and have confidence in the construct validity of the measures. We believe that a large-scale meta-analysis of instruments used to assess HRQOL is required to determine and compare overall levels of reliability and construct validity across the various available measures in the many different health-related areas. This approach would enable researchers to make choices about what instruments to use based on objective rather than subjective preference. Ultimately, researchers need to come to consensus on which HRQOL instruments best capture the concept of HRQOL and end the proliferation of different instruments to measure HRQOL.

A secondary objective of this study was to explore the components of the KDCS and refine as necessary. After eliminating the subscale of sexual activity because of missing data, we found that the subscales of patient satisfaction, work status and dialysis staff encouragement loaded weakly in our one-factor model suggesting that these subscales measured something other than HRQOL. Thus, a 7-subscale KDCS one-factor model fit the data significantly better than the 10-subscale KDCS in this sample. In addition, reliability of the 7-subscale KDCS was stronger than for the 10-subscale KDCS. These findings may have implications for how the KDCS is calculated in future studies. However, it is important to note that a high percentage of participants in our sample did not work (91.6%) as compared to those reported in other studies of hemodialysis patients, which ranged from 70.6 to 81.1% unemployed [13, 64, 65]. The low employment rate in our sample may be one reason why work status had such a low factor loading (.23). A sample with a higher proportion of employed subjects could lead to a choice of using different subscales of the KDCS than we used in our study.

A significant strength of this study is that it is the first known study to utilize confirmatory factor analysis (CFA) to compare the subscales of the HRQOL measures within a study sample. There are several advantages to using CFA over Pearson and intraclass correlations in evaluating how well HRQOL measures within the same sample relate to one another. The first advantage is that, unlike Pearson and intraclass correlations, CFA removes measurement error, which can dilute the observed strength of the associations. The second advantage to CFA is that it allows the researcher to compare how well competing conceptual models fit the data [66]. Thus, in using CFA for this study, we were able to confidently determine that the QWB-SA, SF-6D, and 7-subscale KDCS measured more than one construct.

This study included limitations. The first is that, because this study included mostly males, it is unclear whether or not results can be generalized to females with CKD receiving outpatient hemodialysis. Secondly, the sexual activity subscale of the KDCS was not included in the analysis because of missing data. Missing data for this subscale may have been related to the sensitive nature of the questions or the severity of physical impairment of the patients in this sample. Thirdly, the high unemployment rate in our sample may have influenced our choice of KDCS subscales to use in the CFA. Lastly, the results of the confirmatory factor analysis may have been different if item level scores or total scores, rather than subscales scores, were used.

Conclusion

Using CFA, we were able to provide strong evidence that the SF-6D and KDSC were more closely related to each other than with the QWB-SA. The differences between these utility tools have significant implications for how QALYs derived from the instruments can (or cannot) be compared. More work regarding construct validity of HRQOL instruments is needed to provide future researchers with the information required to make sound decisions when choosing an appropriate HRQOL instruments.

Appendix

Appendix Comparison of HRQOL results from selected studies

Study	Population (sample size)	HRQOL measures	Summary of findings
Barton et al. (2005) [30]	Hearing-impaired patients (n = 915)	EQ-5D HUI3 SF-6D	Estimated utility level varied depending on which quality of life (QOL) tool was used within this sample. Authors suggest that these differences may be related to the tools measuring different constructs
Boonen et al. (2007) [67]	Patients with ankylosing spondylitis (n = 254)	EQ-5D SF-6D RS (well-being rating scale in ankylosing spondylitis)	Agreement among instruments was moderate. Instruments correlated well with health measures but had different psychometric properties

Appendix continued

Study	Population (sample size)	HRQOL measures	Summary of findings
Brazier et al. (2004) [68]	Seven patient groups with a variety of illnesses (n = 2,436)	EQ-5D SF-6D	The mean SF-6D index was higher than the EQ-5D. Intraclass correlation between the two measures was low at .51. Floor effects were found in the SF-6D and ceiling effects in the EQ-5D. The SF-6D had a smaller range than the EQ-5D
Gabriel et al. (1999) [69]	Women with osteoporosis, fractures versus nonfractures (n = 183, 199)	SF-36 HUI Time-trade off	Scores for hypothetical health states from healthy women were 50% lower than those women who actually experienced the health state of a fracture. No significant differences between HUI and TTO scores
Gerard et al. (2004) [10]	Hemodialysis patients (n = 477)	EQ-5D SF-6D	Study found weak agreement between scores on EQ-5D and SF-6D. Utility scores from the EQ-5D were on average 27 unites higher than for the SF-6D. SF-6D distribution of scores was narrower than the EQ-5D. Scores for SF-6D were higher for poorer health states and lower for higher health states than the EQ-5D. EQ-5D and SF-6D evaluated different aspects of QOL
Gorodetskaya et al. (2005) [11]	Chronic kidney disease patients (n = 115)	KDQOL-36 (disease-specific) HUI3 Time trade off	Significant associations were found between estimated glomerular filtration rate and the KDQOL-36 Physical Health Composite Score, the Effects and Burden of Kidney Disease subscales, TTO, and HUI3. However, utility estimates derived from the HUI3 and TTO were not comparable. The calculated QALY for subjects with chronic kidney disease in Stage 4 and Stage 5 from the HUI3 was .58, whereas the QALY derived from the TTO for Stage 5 was .72. Authors suggested that differences might have been due to the HUI3 focusing more on physical health differences, whereas the TTO on functional capacity, disease burden, and risk tolerance. No comparisons of the HUI3 and TTO were made with the KDQOL-36
Hatoum et al. (2004) [27]	Patients undergoing percutaneous coronary intervention (n = 331)	HUI3 SF-6D	Both the SF-6D and HUI3 demonstrated discriminate validity. However, there was significant disagreement between the two tools, and measures are not interchangeable for translating to QALYs. HUI3 focuses on physical and mental dimensions of health, whereas SF-6D includes social functioning
Longworth and Bryan (2003) [31]	Patients undergoing liver transplant (n = 524)	EQ-5D SF-6D	SF-6D did not detect changes over time. In addition, SF-6D did not describe lower levels of health states and had a more narrow range than the EQ-5D. SF-6D proved better at detecting smaller changes than EQ-5D. EQ-5D scores tended to be higher than SF-6D
Marra et al. (2005) [70]	Rheumatoid arthritis patients (n = 239)	EQ-5D HUI2 HUI3 SF-6D RAQoL (disease-specific)	The disease-specific (RAQoL) was the most responsive measure overall. The HUI3 and the SF-6D were the best at detecting improvements in health status; however, the SF-6D detected changes when there were not any. The EQ-5D detected worsening status the best
McDonough et al. (2005) [29]	Patients with spine disease (n = 2,097)	EQ-5D HUI2 HUI3 SF-6D eQWB Visual analog scale (VAS)	In this study, the estimated QWB (eQWB) was estimated from five subscales of the SF-36 based on a regression analysis. Found high correlations between EQ-5D, HUI3, HUI2, and SF-6D and modest correlations among eQWB, HUI3, HUI2, EQ5D, and SF-6D. The lowest correlations were with the VAS and other measures. All of the instruments could discriminate between good, fair, and poor health except for the eQWB. Suggested that eQWB may be limited in describing persons with debilitating disease. Concluded that all tools measured QOL to some degree; however, measured health state values are not interchangeable

Appendix continued

Study	Population (sample size)	HRQOL measures	Summary of findings
O'Brien et al. (2003) [28]	Patients at risk for sudden cardiac death ($n = 246$)	SF-6D HUI3	This study found low intraclass correlation between SF-6D and HUI3. Authors suggested that the low agreement between these tools might be because they measure different concepts or because health states are valued differently
Pickard et al. (2005) [71]	Stroke survivors ($N = 98$)	EQ-5D VAS EQ-index HUI3 HUI2 SF-6D	SF-6D, EQ-Index, and HUI3 more responsive than EQ-5D VAS and HUI2. Changes in EQ-5D VAS and SF-6D more strongly correlated with changes in mental aspects of health, whereas changes in EQ-Index, HUI2, and HUI3 more strongly related to changes in level of disability
Siderowf et al. (2002) [72]	Parkinson's disease ($n = 97$)	EQ-5D HUI2 UPDRS (disease-specific) PDQ $= 39$ (disease-specific)	Scores from the EQ-5D and the HUI2 correlated well with the UPDRS and PDQ-39. However mean scores of the measures differed significantly from each other. The EQ-5D gave the lowest scores of all of the instruments. Authors conclude that tools cannot be used interchangeably
Stavem et al. (2005) [32]	Patients with HIV/AIDS ($n = 60$)	15D SF-6D EQ-5D	The Spearman correlations among the SF-6D, EQ-5D, and 15D were between .74 and .80. The EQ-5D demonstrated a ceiling effect. Authors caution that these QOL measures are not interchangeable
van Stel and Buskens (2006) [73]	Coronary artery disease ($n = 561$)	EQ-5D SF-6D	Although mean utility scores were similar, there were differences in range of scores, agreement, and responsiveness over time. Instruments measure different concepts

References

1. U.S. Renal Data System. (2007). USRDS 2006 annual data report: Atlas of end-stage renal disease in the united states. Resource document. http://www.usrds.org/adr.htm. Accessed 20 June 2007.
2. Testa, M. A., & Nackley, J. F. (1994). Methods for quality-of-life studies. *Annual Review of Public Health, 15*, 535–559. doi:10.1146/annurev.pu.15.050194.002535.
3. Bryan, S., & Longworth, L. (2005). Measuring health-related utility: Why the disparity between EQ-5D and SF-6D? *The European Journal of Health Economics, 6*(3), 253–260. doi:10.1007/s10198-005-0299-9.
4. Chiang, C. (1965). *An index of health: Mathematical models*. PHS 1000 HRP-0005421 (pp. 1–25). Washington, DC: Public Health Service Publications.
5. Korevaar, J. C., Merkus, M. P., Jansen, M. A., Dekker, F. W., Boeschoten, E. W., Krediet, R. T., et al. (2002). Validation of the KDQOL-SF: A dialysis-targeted health measure. *Quality of Life Research, 11*(5), 437–447. doi:10.1023/A:1015631411960.
6. Berlim, M. T., Mattevi, B. S., Duarte, A. P., Thome, F. S., Barros, E. J., & Fleck, M. P. (2006). Quality of life and depressive symptoms in patients with major depression and end-stage renal disease: A matched-pair study. *Journal of Psychosomatic Research, 61*(5), 731–734. doi:10.1016/j.jpsychores.2006.04.011.
7. Guerini, R. D., Mercieri, A., & Yavuzer, G. (2006). Multidimensional health-status assessment of chronic hemodialysis patients: The impact on quality of life. *Europa Medicophysica, 42*(2), 113–119.
8. Al-Arabi, S. (2006). Quality of life: Subjective descriptions of challenges to patients with end stage renal disease. *Nephrology Nursing Journal: Journal of the American Nephrology Nurses' Association, 33*(3), 285–292.
9. Drayer, R. A., Piraino, B., Reynolds, C. F., I. I. I., Houck, P. R., Mazumdar, S., Bernardini, J., et al. (2006). Characteristics of depression in hemodialysis patients: Symptoms, quality of life and mortality risk. *General Hospital Psychiatry, 28*(4), 306–312. doi:10.1016/j.genhosppsych.2006.03.008.
10. Gerard, K., Nicholson, T., Mullee, M., Mehta, R., & Roderick, P. (2004). EQ-5D versus SF-6D in an older, chronically ill patient group. *Applied Health Economics and Health Policy, 3*(2), 91–102. doi:10.2165/00148365-200403020-00005.
11. Gorodetskaya, I., Zenios, S., McCulloch, C. E., Bostrom, A., Hsu, C. Y., Bindman, A. B., et al. (2005). Health-related quality of life and estimates of utility in chronic kidney disease. *Kidney International, 68*(6), 2801–2808. doi:10.1111/j.1523-1755.2005.00752.x.
12. Fryback, D. G., Lawrence, W. F., Martin, P. A., Klein, R., & Klein, B. E. (1997). Predicting quality of well-being scores from the SF-36: Results from the Beaver Dam health outcomes study. *Medical Decision Making, 17*(1), 1–9. doi:10.1177/0272989X9701700101.
13. Lopes, A. A., Bragg-Gresham, J. L., Goodkin, D. A., Fukuhara, S., Mapes, D. L., Young, E. W., et al. (2007). Factors associated with health-related quality of life among hemodialysis patients in the DOPPS. *Quality of Life Research, 16*(4), 545–557. doi:10.1007/s11136-006-9143-7.
14. Parsons, T. L., Toffelmire, E. B., & King-VanVlack, C. E. (2006). Exercise training during hemodialysis improves dialysis efficacy and physical performance. *Archives of Physical Medicine and Rehabilitation, 87*(5), 680–687. doi:10.1016/j.apmr.2005.12.044.
15. Hynes, D. M., Stroupe, K. T., Kaufman, J. S., Reda, D. J., Peterman, A., Browning, M. M., et al. (2006). Adherence to guidelines for ESRD anemia management. *American Journal of Kidney Diseases, 47*(3), 455–461. doi:10.1053/j.ajkd.2005.11.012.

16. Kaplan, R. M., Sieber, W. J., & Ganiats, T. G. (1997). The quality of well-being scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychology & Health, 12*, 783–791. doi:10.1080/08870449708406739.

17. Kaplan, R. M., & Anderson, J. P. (1988). A general health policy model: Update and applications. *Health Services Research, 23*(2), 203–235.

18. Kaplan, R. M., Ganiats, T. G., Sieber, W. J., & Anderson, J. P. (1998). The quality of well-being scale: Critical similarities and differences with SF-36. *International Journal for Quality in Health Care, 10*(6), 509–520. doi:10.1093/intqhc/10.6.509.

19. Sieber, W. J., David, K. M., Adams, J. E., Kaplan, R. M., & Ganiats, T. G. (2000). Assessing the impact of migraine on health-related quality of life: An additional use of the quality of well-being scale-self-administered. *Headache, 40*(8), 662–671. doi:10.1046/j.1526-4610.2000.040008662.x.

20. Tabaei, B. P., Shill-Novak, J., Brandle, M., Burke, R., Kaplan, R. M., & Herman, W. H. (2004). Glycemia and the quality of well-being in patients with diabetes. *Quality of Life Research, 13*(6), 1153–1161. doi:10.1023/B:QURE.0000031336.81580.52.

21. Mancino, M. J., Pyne, J. M., Tripathi, S., Constans, J., Roca, V., & Freeman, T. (2006). Quality-adjusted health status in veterans with posttraumatic stress disorder. *The Journal of Nervous and Mental Disease, 194*(11), 877–879. doi:10.1097/01.nmd.0000244686.79689.21.

22. Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 health survey. *Journal of Clinical Epidemiology, 51*(11), 1115–1128. doi:10.1016/S0895-4356(98)00103-6.

23. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*(2), 271–292. doi:10.1016/S0167-6296(01)00130-8.

24. Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42*(9), 851–859. doi:10.1097/01.mlr.0000135827.18610.0d.

25. Ware, J., & Sherbourne, C. (1992). The MOS 36-item short-form health survey (SF-36). *Medical Care, 30*(6), 473–483. doi:10.1097/00005650-199206000-00002.

26. Ware, J. E., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). *How to score version 2 of the SF-12 health survey*. Massachusetts: QualityMetric Incorporated.

27. Hatoum, H. T., Brazier, J. E., & Akhras, K. S. (2004). Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. *Value in Health, 7*(5), 602–609. doi:10.1111/j.1524-4733.2004.75011.x.

28. O'Brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P., & Brazier, J. (2003). A view from the bridge: Agreement between the SF-6D utility algorithm and the health utilities index. *Health Economics, 12*(11), 975–981. doi:10.1002/hec.789.

29. McDonough, C. M., Grove, M. R., Tosteson, T. D., Lurie, J. D., Hilibrand, A. S., & Tosteson, A. N. (2005). Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among spine patient outcomes research trial (SPORT) participants. *Quality of Life Research, 14*(5), 1321–1332. doi:10.1007/s11136-004-5743-2.

30. Barton, G. R., Bankart, J., & Davis, A. C. (2005). A comparison of the quality of life of hearing-impaired people as estimated by three different utility measures. *International Journal of Audiology, 44*(3), 157–163. doi:10.1080/14992020500057566.

31. Longworth, L., & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics, 12*(12), 1061–1067. doi:10.1002/hec.787.

32. Stavem, K., Froland, S. S., & Hellum, K. B. (2005). Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in

patients with HIV/AIDS. *Quality of Life Research, 14*(4), 971–980. doi:10.1007/s11136-004-3211-7.

33. Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5d and SF-6D based on hypothetical preferences in a general population. *Health Economics, 14*(11), 1169–1189. doi:10.1002/hec.1006.

34. Hays, R. D., Kallich, J. D., Mapes, D. L., Coons, S. J., & Carter, W. B. (1994). Development of the kidney disease quality of life (KDQOL) instrument. *Quality of Life Research, 3*(5), 329–338. doi:10.1007/BF00451725.

35. Hays, R. D., Kallich, J. D., Mapes, D. L., Coons, H. L., & Carter, W. B. (1995). *Kidney disease quality of life short form (KDQOL-SF)*. Santa Monica, CA: Rand.

36. Mapes, D. L., Bragg-Gresham, J. L., Bommer, J., Fukuhara, S., McKevitt, P., Wikstrom, B., et al. (2004). Health-related quality of life in the dialysis outcomes and practice patterns study (DOPSS). *American Journal of Kidney Diseases, 44*(5 Suppl 2), 54–60.

37. Paniagua, R., Amato, D., Vonesh, E., Guo, A., Mujais, S., & Mexican Nephrology Collaborative Study Group. (2005). Health-related quality of life predicts outcomes but is not affected by peritoneal clearance: The ademex trial. *Kidney International, 67*(3), 1093–1104. doi:10.1111/j.1523-1755.2005.00175.x.

38. Lopes, A. A., Bragg-Gresham, J. L., Satayathum, S., McCullough, K., Pifer, T., Goodkin, D. A., et al. (2003). Health-related quality of life and associated outcomes among hemodialysis patients of different ethnicities in the United States: The dialysis outcomes and practice patterns study (DOPPS). *American Journal of Kidney Diseases, 41*(3), 605–615. doi:10.1053/ajkd.2003.50122.

39. van Janssen, D. K., Heylen, M., Mets, T., & Verbeelen, D. (2004). Evaluation of functional and mental state and quality of life in chronic haemodialysis patients. *International Urology and Nephrology, 36*(2), 263–267. doi:10.1023/B:UROL.0000034653.59183.77.

40. Bakewell, A. B., Higgins, R. M., & Edmunds, M. E. (2002). Quality of life in peritoneal dialysis patients: Decline over time and association with clinical outcomes. *Kidney International, 61*(1), 239–248. doi:10.1046/j.1523-1755.2002.00096.x.

41. Kawauchi, A., Inoue, Y., Hashimoto, T., Tachibana, N., Shirakawa, S., Mizutani, Y., et al. (2006). Restless legs syndrome in hemodialysis patients: Health-related quality of life and laboratory data analysis. *Clinical Nephrology, 66*(6), 440–446.

42. Vazquez, I., Valderrabano, F., Fort, J., Jofre, R., Lopez-Gomez, J. M., Moreno, F., et al. (2005). Psychosocial factors and health-related quality of life in hemodialysis patients. *Quality of Life Research, 14*(1), 179–190. doi:10.1007/s11136-004-3919-4.

43. Manns, B., Johnson, J. A., Taub, K., Mortis, G., Ghali, W. A., & Donaldson, C. (2003). Quality of life in patients treated with hemodialysis or peritoneal dialysis: What are the important determinants? *Clinical Nephrology, 60*(5), 341–351.

44. Arbor Research Collaborative for Health. (2007). Dialysis outcomes and practice patterns study. Resource document.

45. Yuan, Y. (2000). *Multiple imputation for missing data: Concepts and new developments*. P267-25. Rockville, MD: SAS Institute.

46. Carmines, E. G., & Zeller, R. A. (1978). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

47. McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

48. Joreskog, K. G., & Sorbom, D. G. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software.

49. Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48*(4), 493–517. doi:10.1007/BF02293875.

50. Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City. http://www.statpower.net/Steiger%20Biblio/Steiger-Lind%201980.pdf.

51. Bentler, P. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

52. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. doi:10.1037/0033-2909.107.2.238.

53. Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. doi:10.1037/0033-2909.88.3.588.

54. Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453. doi:10.1037/1082-989X.3.4.424.

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

56. Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

57. Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8*, 161–168. doi:10.1007/BF02288700.

58. Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106*(2), 315–317. doi:10.1037/0033-2909.106.2.315.

59. Smith, K. W., Avis, N. E., & Assmann, S. F. (1999). Distinguishing between quality of life and health status in quality of life research: A meta-analysis. *Quality of Life Research, 8*, 447–459. doi:10.1023/A:1008928518577.

60. Coons, S. J., & Kaplan, R. M. (1993). Quality of life assessment: Understanding its use as an outcome measure. *Hospital Formulary, 28*(5), 486–490.

61. Garratt, A., Schmidt, L., Mackintosh, A., & Fitzpatrick, R. (2002). Quality of life measurement: Bibliographic study of patient assessed health outcome measures. *BMJ (Clinical Research Ed.), 324*(7351), 1417. doi:10.1136/bmj.324.7351.1417.

62. Deverill, M., Brazier, J., Green, C., & Booth, A. (1998). The use of QALY and non-QALY measures of health-related quality of life. Assessing the state of the art. *PharmacoEconomics, 13*(4), 411–420. doi:10.2165/00019053-199813040-00004.

63. Naughton-Collins, M., Walker-Corkery, E., & Barry, M. J. (2004). Health-related quality of life, satisfaction, and economic outcome measures in studies of prostate cancer screening and treatment, 1990–2000. *Journal of the National Cancer Institute. Monographs, 33*, 78–101. doi:10.1093/jncimonographs/lgh016.

64. Kutner, N., Bowles, T., Zhang, R., Huang, Y., & Pastan, S. (2008). Dialysis facility characteristics and variation in employment rates: A national study. *Clinical Journal of the American Society of Nephrology; CJASN, 3*(1), 111–116. doi:10.2215/CJN.02990707.

65. Bohlke, M., Marini, S., Gomes, R., Terhorst, L., Rocha, M., Poli De Figueiredo, C. E., et al. (2008). Predictors of employment after successful kidney transplantation—a population-based study. *Clinical Transplantation, 22*, 405–410.

66. Bryant, F. B., Yarnold, P. R., & Michelson, E. A. (1999). Statistical methodology: VIII. Using confirmatory factor analysis (CFA) in emergency medicine research. *Academic Emergency Medicine, 6*(1), 54–66. doi:10.1111/j.1553-2712.1999.tb00096.x.

67. Boonen, A., van der Heijde, D., Landewe, R., van Tubergen, A., Mielants, H., Dougados, M., & van der Linden, S. (2007).How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis?*Annals of the Rheumatic Diseases, 66*, 771–777.

68. Brazier, J. E., Tsuchiya, A., Roberts, J., & Busschbach, J. (2004). A comparison of the EQ-5D and the SF-6D across seven patient groups. *Health Economics, 13*, 873–884.

69. Gabriel, S. E.,Kneeland, T. S., Melton, L. J., Moncur, M. M.,Ettinger, B.,Tosteson, A. N. A. (1999). Health-related quality of life in economic evaluations for osteoporosis:whose values should we use? *Medical Decision Making, 19*, 141–148.

70. Marra, C. A., Rashidi, A. A., Guh, D., Kopec, J. A., Abrahamowicz, M., Esdaile, J. M., et al. (2005). Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Quality of Life Research, 14*(5), 1333–1344.

71. Pickard, A. S., Johnson, J. A., & Feeny, D. H. (2005). Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research, 14*(1), 207–219.

72. Siderowf, A., Ravina, B., & Glick, H. (2002). Preference-based quality-of-life in patients with Parkinson's disease. *Neurology, 59*, 103–108.

73. van Stel, H. F., & Buskens, E. (2006). Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease.*Health Quality of Life Outcomes, 4*, 20.