Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Psychometric Validation of the KDQoL-36 Using Item Response Theory and Computer Adaptive Testing among Patients Receiving Dialysis

Jasmine Ion Titapiccolo
Fresenius Medical Care, Global Medical Office, Renal Research Institute
Jenny E Flythe
University of North Carolina (UNC) Kidney Center, Division of Nephrology and Hypertension, Department of Medicine, UNC School of Medicine
Sugandha Saxena
Fresenius Medical Care, Global Medical Office
Francesco Bellocchio
Fresenius Medical Care, Global Medical Office, Renal Research Institute
Abraham Rincon Bello
Fresenius Medical Care, Global Medical Office
Maria Eva Baro Salvador
Fresenius Medical Care, Global Medical Office
Christophe Ridel
Fresenius Medical Care, Global Medical Office
Mark Gilchrist
Fresenius Medical Care, UK Renal Services Limited
Alessandro Pizzo
NephroCare Italy
Mario Cioffi
Fresenius Medical Care, Global Medical Office
Adrian Guinsburg
Fresenius Medical Care, Global Medical Office
Berit Goethel-Paal
Fresenius Medical Care, D GmbH
Katarina Nehezova
Fresenius Medical Care - DS, S.R.O
Diana Samaniengo Toro
Fresenius Medical Care, Global Medical Office
Milind Nikam
Fresenius Medical Care, Global Medical Office
Ana Ferreira Ribeiro
Fresenius Medical Care, D GmbH
Len Usvyat
Fresenius Medical Care, Global Medical Office, Renal Research Institute
Stefano Stuard
Fresenius Medical Care, Global Medical Office
Luca Neri
luca.neri@freseniusmedicalcare.com

Fresenius Medical Care, Global Medical Office, Renal Research Institute

Research Article

Keywords: Health-Related Quality of Life (HRQoL), Kidney Disease Quality of Life-36 (KDQoL-36), Dialysis Patients, Item Response Theory (IRT), Computerized Adaptive Testing (CAT), Patient-Reported Outcomes (PROMs)

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9071006/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: Competing interest reported. All authors, except J.E. Flythe, are employees of Fresenius Medical Care. In the last two years, J. E. Flythe has received consulting fees from Fresenius Medical Care not related to this work and Aquapass and compensation for serving on a Data and Safety Monitoring Board for the Veterans Affairs.

Abstract

Background: Health-Related Quality of Life (HRQoL) is a critical outcome in end-stage kidney disease care. The Kidney Disease Quality of Life-36 (KDQoL-36) survey is used widely to measure HRQoL but is limited by length and lack of personalization. We evaluated the psychometric performance of a multidimensional Item Response Theory (IRT) model for KDQoL-36 and simulated a Computerized Adaptive Testing (CAT) version in a large European dialysis cohort.

Methods: We included 8,325 KDQoL-36 responses from dialysis patients in 170 NephroCare clinics across six European countries and used confirmatory factor analysis and bifactor IRT modeling to assess psychometric properties. We conducted CAT simulations to evaluate reliability and potential for burden reduction. Construct and criterion validity were assessed using hospitalization outcomes and known-group comparisons.

Results: Bifactor models demonstrated superior fit for both Short Form-12 (SF-12) and disease-specific KDQoL-36 domains, with excellent reliability ($\omega > 0.90$). IRT-based scores showed high agreement with classical test theory (R^2: 0.77–0.94) and differentiated patient groups by hospitalization risk (logistic regressions: OR = 0.90-0.94; $p < 0.05$). CAT simulations showed strong agreement with full-length scores (R^2: 0.93–0.99) and reduced item count by 25%. CCAT-based administrations averaged 7.40 ± 2.53 items per patient for the SF-12 domain and 18.75 ± 4.09 items per patient for the disease-specific domains while preserving measurement precision.

Conclusion: Multidimensional IRT modeling supports the KDQoL-36 as a reliable, valid measure of HRQoL in dialysis care. CAT implementation may offer substantial efficiency gains with minimal loss of information, reducing patient burden, thus enhancing the meaningful integration of HRQoL into routine hemodialysis care.

Introduction

Patients with end-stage kidney disease receiving chronic dialysis experience significant disruptions to their daily lives and overall well-being, as both the disease and its treatment are profoundly intrusive[1]. Given that more than 3.5 million people worldwide and 540,000 individuals in the United States receive maintenance hemodialysis or peritoneal dialysis, the impact of dialysis on patients' day-to-day functioning and well-being is a major public health and care-quality concern[2]. The dialysis population is typically older and medically complex, with a high symptom burden (e.g., fatigue, cramps, pruritus) and frequent functional impairment/frailty, all of which can make completion of burdensome patient reported outcomes (PRO) instruments difficult[3]. Aligned with research showing that patients with chronic kidney disease (CKD) often prioritize how they feel and function over clinical outcomes as well as the growing recognition of the importance of integrating PROs with traditional biochemical and technical metrics to assess therapy adequacy[4], Kidney Disease Improving Global Outcomes (KDIGO) 2024 CKD Clinical Practice Guideline recommend routine use of validated PRO measures in routine CKD care[5,6]. The Kidney Disease Quality of Life-36 (KDQoL-36) is one of the most widely used instruments for HRQoL assessment in dialysis worldwide. In the United States, HRQoL assessment is mandated by Centers for Medicare & Medicaid Services and the KDQoL-36 is widely implemented in routine care, with documented use across >1,300 dialysis facilities[7].

The KDQoL-36 is a 36-item questionnaire that assesses both generic (physical functioning, emotional well-being) and kidney disease-specific (disease burden, symptoms, and effects on activities) domains[8–12]. Despite its widespread use, the KDQoL-36 is limited by its length and lack of personalization, possibly contributing to survey fatigue[5,13]. Across >413,000 administration opportunities at a U.S. dialysis organization, KDQOL-36 non-response rates were ~20%, and response distributions suggested ceiling effects in at least one domain, indicating potential limitations of fixed-length administration at scale[7]. Together, these findings support the need for refined and more efficient administration approaches that preserve interpretability while reducing respondent burden[7].

Computer-adaptive testing (CAT), a method that uses Item Response Theory (IRT) to tailor questionnaires in real time to each person's responses, reduces survey burden while preserving accuracy[14,15]. CAT-based PROs (e.g., PROMIS CATs) are increasingly used and validated in other clinical areas such as orthopedics and oncology, but their adoption in nephrology remains comparatively limited[16,17]. To address this knowledge gap, we sought to (a) evaluate the psychometric properties of a multidimensional IRT model for the KDQoL-36 in a large dialysis cohort, and (b) conduct CAT simulations to evaluate the impact of adaptive administration on reliability and respondent burden.

METHODS

Setting and Study Design

Since 2021, NephroCare clinics in several European countries have used electronic Patient Reported Outcome Measures (ePROMs) to individualize clinical care as part of their continuous quality improvement (CQI) program[5]. The program entails periodic administration of selected ePROMs that map to outcomes such as itch, fatigue, and quality of life[14]. Data are collected in the local language via a dedicated application during dialysis sessions, with real-time integration into patients' clinical records to support timely clinical decision-making A total of 170 NephroCare clinics of Fresenius Medical Care (FME) has participated in the ePROM CQI program since January 2021 in France, Ireland, Italy, Spain, United Kingdom and Czech Republic. For the presented analyses, we included all adult patients receiving chronic dialysis at participating centers, who completed the ePROMs between September 2021 and February 2023. KDQoL-36 completion was voluntary, and patients provided informed consent for their data to be used for secondary research purposes.

Medical and Patient-reported Data

Medical and HRQoL data were retrospectively abstracted from the EuCliD® database, the electronic health information system used in participating clinics. The most recent biochemical results up to 12 months before survey administration were abstracted. Active vascular access and dialysis modality were determined based on the most frequently used access type in the three months preceding survey completion. Comorbidities were identified using pre-survey ICD-10 codes. Chronic kidney disease etiology was classified as diabetes, hypertension, glomerulonephritis, polycystic kidney disease, or other. All-cause hospitalization within 12 months post-survey was extracted from routinely collected data. The KDQoL-36[18–20] was used to assess generic and kidney disease-specific quality of life. The KDQoL-36 includes the SF-12 (used to derive physical and mental health composite scores - Physical Health Composite ($SF\text{-}12_{PHC}$) and Mental Health Composite ($SF\text{-}12_{MHC}$) scores) and three kidney disease-specific domains: Burden (perceived burden of kidney disease - BKD), Symptoms/Problems (symptom frequency/severity - SKD), and Effects of Kidney Disease (EKD - impact on daily life and activities). Domain scores were converted to a 0-100 scale, with higher values indicating better quality of life (i.e., lower burden/fewer symptoms/less impact). Theoretical details on the KDQoL-36 are provided in Appendix 1 (Table A1).

The SF-12 and kidney disease-specific domains were analyzed separately for both IRT modelling and CAT implementation. We described patient characteristics and HRQoL data using descriptive statistics.

Psychometric Analysis

The sample was randomly partitioned into training set (70%) - used for IRT model estimation - and validation set (30%) - used for invariance tests, CAT simulation, and construct validity analysis.

a) IRT model validation

We followed the COSMIN/ISOQOL framework to evaluate measurement properties, with the assessed properties, analyses[21], and decision rules summarized in Appendix 1 (Table A1.2). Item-level properties, including floor and ceiling effects, inter-item and item-total correlations, and internal consistency (Cronbach's alpha), were calculated to assess basic psychometric properties. We conducted a set of confirmatory IRT-based factor analysis on complete cases to assess the factor structure of the SF-12 and KDQoL-36 disease-specific scales, based on previous research findings[22].

We tested the following alternative models (a detailed description is presented in Appendix 1.2):

- SF-12: Four IRT-based models were compared to evaluate the structure underlying $SF\text{-}12_{PHC}$ and $SF\text{-}12_{MHC}$ scores: uncorrelated two-factor (*Model 1a*), correlated two-factor (*Model 2a*), an alternative correlated two-factor (*Model 3a*), and a bifactor model (*Model 4a*);
- Kidney Disease-Specific scales: Three models were compared: correlated three-factor model (*Model 1b*), uncorrelated three-factor model (*Model 2b*), and bifactor model (*Model 3b*).

Model fit was evaluated using RMSEA, SRMSR, CFI, and TLI; model comparisons used AIC, BIC, and SABIC (acronyms are defined in the caption of Table 4a). For bifactor models, we examined loadings/item parameters, reliability (Explained Common Variance, ECV; Omega Hierarchical, ω_H; and Omega coefficients, ω), and test information functions (TIFs), retaining the best-fitting structures. Model performance was then tested on the validation set. Composite scale scores were computed as ω-weighted averages of general and specific factor scores under the selected IRT models.

Construct validity was assessed using predefined hypotheses about expected associations. Pearson correlations and linear regression were used to test hypotheses. Convergent and divergent validity were examined using a multi-trait multi-method (MTMM) matrix: convergent validity was supported by within-factor correlations ≥0.30, and divergent validity by within-factor correlations meaningfully exceeding cross-factor correlations. Known-groups validity was assessed by relating HRQoL scores to a hospitalization risk score stratified into quintiles.

In the test set, we derived a hospitalization risk score (HRS) from a logistic regression model predicting 12-month hospitalization using baseline clinical variables. Construct validity was assessed by modeling 12-month hospitalization as a function of each KDQoL-36 score, adjusting for HRS in separate logistic regressions. We hypothesized stronger associations with the SF-12 general factor ($SF\text{-}12_G$) than with the General Kidney Disease Scale (KDS_G), and stronger associations between PHC and hospitalization risk severity classes than for MHC and KDQoL-36 disease-specific scales.

For each scale we computed the standardized T-score (mean=50, SD=10) relative to the NephroCare network population using the following equation where θ is the computed score, μ and σ correspond to the average and standard deviation values of the computed scores in training dataset:

$$(1) \qquad Tscore = 50 + 10\,\frac{(\theta-\mu)}{\sigma}$$

b) Adaptive Testing Simulations

CAT stopping parameters (SE_θ threshold and Δ_θ change stopping rule) were iteratively tuned to minimize the number of administered items while preserving agreement with full-length scores; specifically, thresholds were chosen such that the Bland-Altman 95% limits of agreement (and their confidence intervals) for the CAT vs. full-form score comparison remained within a tolerance of <0.5 SD of the full-form score[23].

Agreement between Classical Test Theory (CTT)-, IRT-, and CAT-derived scores was quantified using the intraclass correlation coefficient (ICC) from a two-way absolute-agreement, average-measure model (ICCagreement)[21]. The association between IRT- and CTT-based scores was additionally assessed using linear regression, reporting slope, intercept, and R^2; evidence for linear association was evaluated by testing slope difference from zero (two-sided p-value).

While an anchor-based minimum clinical importance difference (MCID) was not available, the SDC offers a conservative reference threshold for change interpretation, indicating the MCID that can be considered beyond measurement error. The smallest detectable change (SDCagreement) was then calculated for each scale as $1.96*\text{sqrt}(2)*SEM_{agreement,}$where $SEM_{agreement}=SD_{pooled}*\text{sqrt}(1-ICC_{agreement})$ and SD_{pooled} is the pooled standard deviation across methods[21,24]. Analyses were conducted in R (version 4.4.2) using the mirt and mirtCAT packages (Chalmers, 2012, 2016).

RESULTS

Cohort Description

Among 9,840 patients invited to complete the KDQoL-36 in routine care, 8,325 returned a completed questionnaire; 1,102 (11.2%) declined and 477 (4.8%) did not complete it. A nurse assisted 2,639 patients (26.5%) during completion. Table 1 reports study patient demographic characteristics and KDQoL-36 scores computed using CTT.

Table 1

Characteristics	Mean (SD) or N (%)
	N = 8325
Age	*68.1 (14.2)*
Female	*5278 (63.4)*
Ethnicity	
Caucasian	*4455 (53.5)*
Missing	*3603 (43.3)*
Other	*267 (3.2)*
Vintage > 1 year	*6840 (82.2)*
Etiology	
Diabetes	*1375 (16.5)*
Hypertension	*940 (11.3)*
Glomerulonephritis	*790 (9.5)*
Polycystic Kidney Disease	*487 (5.8)*
Tubulointerstitial nephritis and related renal tubulointerstitial disorders	*414 (5.0)*
Missing/Unknown/Other causes	*4319 (51.9)*
Tobacco Use	
Missing	*4697 (56.4)*
Current/Past smoker	*1549 (18.6)*
Non-smoker	*2079 (25.0)*
Body Mass Index (BMI)	*26.96 (6.10)*
Charlson Comorbidity Index Score	*3.81 (1.94)*
Dialysis Modality	
Hemodialysis (HD)	*2931 (36.4)*
Hemodiafiltration (HDF)	*5125 (63.6)*
Vascular Access	
Arteriovenous Fistula	*5328 (66.1)*
Central Venous Catheter	*2521 (31.3)*
Graft or other	*476 (5.7)*
Hemoglobin (g/dL)	*11.34 (1.32)*
Calcium (mg/dL)	*9.01 (0.66)*
Phosphate (mg/dL)	*4.49 (1.42)*
Albumin (g/dL)	*3.80 (0.75)*
Parathyroid Hormone (PTH) (pg/mL)	*322.6 (315.5)*
Dialysis adequacy (Kt/V)	*1.72 (0.41)*
Diabetes	*3024 (37.5)*
Hypertension	*4194 (52.0)*
Congestive Heart Failure	*2111 (26.2)*
Chronic Obstructive Pulmonary Disease	*427 (5.3)*
Peripheral Vascular Disease	*1462 (18.1)*
Cerebrovascular Disease	*916 (11.4)*
Cancer (%)	*1140 (14.1)*

Burden of Kidney Disease (BKD)	*43.8 (26.3)*
Effect of Kidney Disease (EKD)	*69.7 (20.9)*
Symptom of Kidney Disease (SKD)	*79.8 (15.0)*
Physical Health Composite Score (PHC)	*38.0 (10.1)*
Mental Health Composite Score (MHC)	*46.5 (10.6)*

Table 1: Baseline characteristics of the study population.
Values are presented as mean (standard deviation: SD) for continuous variables and as n (%) for categorical variables.

Psychometric Properties and IRT modelling

Items mean, floor and ceiling effect, and missing percentages for SF-12 and KDQoL-36 disease-specific scales are reported in Table 2a and Table 2b respectively.

For the SF-12 items, we observed no notable floor or ceiling effects, whereas the KDQoL-36 specific scales showed marked floor effects. Item-level missingness was generally very low, except for Item 35 (sexual activity: 24.5%).

Table 2a

Item	N levels	Minimum %	Maximum %	Item Mean (SD)	Missing %
SF12-Item 1	*5*	*4.53*	*15.47*	*3.47 (1.01)*	*0.86*
SF12-Item 2	*3*	*32.45*	*27.77*	*1.95 (0.77)*	*0.91*
SF12-Item 3	*3*	*36.46*	*26.23*	*1.9 (0.79)*	*1.00*
SF12-Item 4	*2*	*58.42*	*41.58*	*1.42 (0.49)*	*1.18*
SF12-Item 5	*2*	*64.43*	*35.57*	*1.36 (0.48)*	*1.15*
SF12-Item 6	*2*	*49.89*	*50.11*	*1.5 (0.5)*	*1.00*
SF12-Item 7	*2*	*45.2*	*54.8*	*1.55 (0.50)*	*1.72*
SF12-Item 8	*5*	*33.17*	*6.23*	*2.39 (1.26)*	*0.88*
SF12-Item 9	*6*	*22.63*	*3.48*	*2.77 (1.42)*	*0.71*
SF12-Item 10	*6*	*9.35*	*10.59*	*3.68 (1.47)*	*0.72*
SF12-Item 11	*6*	*4.11*	*28.18*	*4.4 (1.42)*	*0.6*
SF12-Item 12	*5*	*7.98*	*32.71*	*3.54 (1.29)*	*1.09*

Table 2a. Descriptive psychometric characteristics of SF-12 questionnaire data. *Minimum % and Maximum % indicate floor and ceiling effects, respectively; N levels = number of response categories. Item mean value of answers is reported with standard deviation values (SD).*

Table 2b

Item	N levels	Minimum %	Maximum %	Item Mean (SD)	Missing %
KDQoL-36-Item 13	*5*	*29.63*	*7.77*	*2.29 (1.24)*	*0.72*
KDQoL-36-Item 14	*5*	*25.18*	*9.11*	*2.48 (1.29)*	*0.67*
KDQoL-36-Item 15	*5*	*18.97*	*15.83*	*2.84 (1.39)*	*0.9*
KDQoL-36-Item 16	*5*	*12.8*	*34.26*	*3.41 (1.46)*	*1.27*
KDQoL-36-Item 17	*5*	*36.89*	*5.28*	*2.19 (1.2)*	*0.61*
KDQoL-36-Item 18	*5*	*78.56*	*0.87*	*1.33 (0.74)*	*0.55*
KDQoL-36-Item 19	*5*	*46.66*	*3.07*	*1.91 (1.08)*	*0.48*
KDQoL-36-Item 20	*5*	*51.07*	*3.5*	*1.86 (1.11)*	*0.55*
KDQoL-36-Item 21	*5*	*40.45*	*5.1*	*2.11 (1.20)*	*0.52*
KDQoL-36-Item 22	*5*	*57.69*	*2.58*	*1.73 (1.05)*	*0.44*
KDQoL-36-Item 23	*5*	*55.5*	*1.65*	*1.71 (0.96)*	*0.47*
KDQoL-36-Item 24	*5*	*60.82*	*2.6*	*1.68 (1.03)*	*0.62*
KDQoL-36-Item 25	*5*	*31.71*	*6.28*	*2.3 (1.21)*	*0.52*
KDQoL-36-Item 26	*5*	*50.3*	*2.85*	*1.86 (1.08)*	*0.61*
KDQoL-36-Item 27	*5*	*68.69*	*1.44*	*1.51 (0.9)*	*0.61*
KDQoL-36-Item 28	*5*	*72.4*	*1.64*	*1.46 (0.88)*	*3.59*
KDQoL-36-Item 29	*5*	*32.97*	*9.08*	*2.38 (1.31)*	*0.86*
KDQoL-36-Item 30	*5*	*40.25*	*4.95*	*2.12 (1.19)*	*0.67*
KDQoL-36-Item 31	*5*	*35.09*	*7.86*	*2.31 (1.28)*	*1.13*
KDQoL-36-Item 32	*5*	*29.43*	*18.87*	*2.8 (1.51)*	*1.51*
KDQoL-36-Item 33	*5*	*51.3*	*5.47*	*1.96 (1.22)*	*1.19*
KDQoL-36-Item 34	*5*	*37.87*	*6.02*	*2.17 (1.21)*	*0.86*
KDQoL-36-Item 35	*5*	*53.16*	*10.97*	*2.1 (1.42)*	*24.47*
KDQoL-36-Item 36	*5*	*57.14*	*3.87*	*1.81 (1.14)*	*2.74*

Table 2b. Descriptive psychometric characteristics of KDQoL-36 disease-specific scales data. *Minimum % and Maximum % indicate floor and ceiling effects, respectively; N levels = number of response categories. Item mean value of answers is reported with standard deviation values (SD).*

In construct validity analyses using a MTMM approach, inter-item correlations demonstrated solid convergent validity for the SF-12_{PHC} items (values higher than 0.4 for most correlations) and acceptable convergent validity for the SF-12_{MHC} items (most of the correlation values higher than 0.3). Inter-item correlations among the 24 items belonging to kidney disease-specific scales ranged from low (0.10) to moderate (0.62), reflecting diverse but related facets. Within-factor correlations were stronger for the Burden scale (range: r = 0.35-0.62), while few items in both the Symptoms and Effects scale demonstrated suboptimal convergent validity (SKD: Item-20 and Item-28; EKD: Item-35 and Item-29). Divergent validity was generally weaker with several cross-factor r > 0.30; however, the predefined expectation that within-factor correlations would be meaningfully higher than cross-factor correlations was largely met for the SF-12 and consistently met for the KDQoL-36 disease-specific scales. Tables 3a and 3b display complete results.

Table 3a

Dimension	Inter-item correlation coefficients											
	Items	$SF\text{-}12_{PHC}$						$SF\text{-}12_{MHC}$				
$SF\text{-}12_{PHC}$		SF-12-1	SF-12-2	SF-12-3	SF-12-4	SF-12-5	SF-12-8	SF-12-6	SF-12-7	SF-12-9	SF-12-10	SF-12-11	SF-12
	SF12-1	-	-	-	-	-	-	-	-	-	-	-	-
	SF12-2	0.33	-	-	-	-	-	-	-	-	-	-	-
	SF12-3	0.32	0.59	-	-	-	-	-	-	-	-	-	-
	SF12-4	0.28	0.39	0.37	-	-	-	-	-	-	-	-	-
	SF12-5	0.29	0.42	0.41	0.59	-	-	-	-	-	-	-	-
	SF12-8	0.27	0.31	0.28	0.48	0.47	-	-	-	-	-	-	-
$SF\text{-}12_{MHC}$	SF12-6	0.16	0.24	0.25	0.31	0.35	0.37	-	-	-	-	-	-
	SF12-7	0.35	0.39	0.39	0.35	0.38	0.34	0.26	-	-	-	-	-
	SF12-9	0.30	0.21	0.21	0.21	0.23	0.30	0.20	0.36	-	-	-	-
	SF12-10	0.40	0.35	0.36	0.30	0.32	0.27	0.19	0.39	0.43	-	-	-
	SF12-11	0.28	0.25	0.23	0.27	0.26	0.43	0.22	0.32	0.42	0.28	-	-
	SF12-12	0.26	0.34	0.30	0.34	0.33	0.38	0.25	0.39	0.31	0.29	0.48	-

Table 3a: Multi-trait multi-method matrix (MTMM) of SF12 questionnaire: Item convergent and divergent validity.

Inter-item and item-subscale Pearson's correlation coefficients for the Physical Health Composite ($SF\text{-}12_{PHC}$) and Mental Health Composite Score ($SF\text{-}12_{MHC}$) dimensions. Values below the diagonal represent inter-item correlation values, while the last two columns show item-total correlations with $SF\text{-}12_{PHC}$ and $SF\text{-}12_{MHC}$ scales, respectively.

Table 3b

Item-subscale correlation coefficients

Table 3b Inter-item correlation coefficients

DIMENSION		*KDQoL-36$_{BKD}$*					*KDQoL-36$_{SKD}$*											*KDQoL*
ITEM																		
I13	*I14*	*I15*	*I16*	*I17*	*I18*	*I19*	*I20*	*I21*	*I22*	*I23*	*I24*	*I25*	*I26*	*I27*	*I28*	*I29*		*I30*	*I31*
KDQoL-36$_{BKD}$	*I13*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	I14	*0.62*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	I15	*0.55*	*0.57*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	I16	*0.37*	*0.35*	*0.48*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
KDQoL-36$_{SKD}$	*I17*	*0.23*	*0.22*	*0.24*	*0.24*	-	-	-	-	-	-	-	-	-	-	-	-	-	
	I18	*0.12*	*0.12*	*0.17*	*0.21*	*0.27*	-	-	-	-	-	-	-	-	-	-	-	-	
	I19	*0.14*	*0.16*	*0.17*	*0.17*	*0.31*	*0.25*	-	-	-	-	-	-	-	-	-	-	-	
	I20	*0.16*	*0.16*	*0.17*	*0.18*	*0.24*	*0.21*	*0.26*	-	-	-	-	-	-	-	-	-	-	
	I21	*0.20*	*0.18*	*0.17*	*0.20*	*0.29*	*0.19*	*0.26*	*0.49*	-	-	-	-	-	-	-	-	-	
	I22	*0.19*	*0.18*	*0.20*	*0.24*	*0.30*	*0.39*	*0.31*	*0.28*	*0.27*	-	-	-	-	-	-	-	-	
	I23	*0.21*	*0.19*	*0.26*	*0.22*	*0.30*	*0.30*	*0.27*	*0.25*	*0.28*	*0.37*	-	-	-	-	-	-	-	
	I24	*0.17*	*0.14*	*0.19*	*0.24*	*0.22*	*0.20*	*0.15*	*0.20*	*0.24*	*0.26*	*0.28*	-	-	-	-	-	-	
	I25	*0.32*	*0.30*	*0.33*	*0.34*	*0.41*	*0.27*	*0.31*	*0.29*	*0.35*	*0.42*	*0.42*	*0.39*	-	-	-	-	-	
	I26	*0.18*	*0.16*	*0.18*	*0.22*	*0.39*	*0.26*	*0.31*	*0.28*	*0.32*	*0.31*	*0.33*	*0.24*	*0.36*	-	-	-	-	
	I27	*0.17*	*0.17*	*0.20*	*0.23*	*0.26*	*0.29*	*0.25*	*0.25*	*0.26*	*0.32*	*0.35*	*0.36*	*0.36*	*0.30*	-	-	-	
	I28	*0.12*	*0.12*	*0.15*	*0.14*	*0.16*	*0.19*	*0.17*	*0.15*	*0.15*	*0.17*	*0.18*	*0.14*	*0.18*	*0.19*	*0.19*	-	-	
KDQoL-36$_{EKD}$	*I29*	*0.26*	*0.23*	*0.24*	*0.19*	*0.19*	*0.13*	*0.16*	*0.17*	*0.20*	*0.19*	*0.20*	*0.10*	*0.24*	*0.18*	*0.18*	*0.13*	-	
	I30	*0.23*	*0.22*	*0.24*	*0.23*	*0.19*	*0.16*	*0.16*	*0.18*	*0.17*	*0.17*	*0.21*	*0.15*	*0.26*	*0.19*	*0.21*	*0.15*	*0.54*	
	I31	*0.33*	*0.30*	*0.33*	*0.39*	*0.37*	*0.24*	*0.22*	*0.24*	*0.32*	*0.32*	*0.31*	*0.29*	*0.50*	*0.30*	*0.28*	*0.18*	*0.27*	
	I32	*0.34*	*0.34*	*0.28*	*0.26*	*0.23*	*0.13*	*0.16*	*0.17*	*0.22*	*0.21*	*0.21*	*0.16*	*0.34*	*0.20*	*0.21*	*0.15*	*0.29*	
	I33	*0.25*	*0.27*	*0.30*	*0.31*	*0.24*	*0.22*	*0.19*	*0.21*	*0.23*	*0.27*	*0.24*	*0.23*	*0.34*	*0.21*	*0.28*	*0.18*	*0.27*	
	I34	*0.39*	*0.38*	*0.47*	*0.40*	*0.32*	*0.24*	*0.25*	*0.23*	*0.28*	*0.26*	*0.31*	*0.26*	*0.43*	*0.26*	*0.30*	*0.23*	*0.33*	
	I35	*0.21*	*0.19*	*0.22*	*0.22*	*0.16*	*0.15*	*0.11*	*0.13*	*0.16*	*0.15*	*0.20*	*0.13*	*0.23*	*0.16*	*0.17*	*0.16*	*0.19*	
	I36	*0.26*	*0.27*	*0.31*	*0.32*	*0.23*	*0.21*	*0.17*	*0.18*	*0.23*	*0.23*	*0.27*	*0.23*	*0.33*	*0.21*	*0.29*	*0.21*	*0.27*	

Table 3b: Multi-trait multi-method (MTMM) analysis of of KDQoL-36 disease-specific scales questionnaire: Item convergent and divergent validity.
Inter-item and item-subscale Pearson's correlation coefficients for the Burden Kidney Disease (BKD) (I13–I16), Symptom Kidney Disease (I17–I28), and Effect Kidney Disease (I29–I36) dimensions. Values below the diagonal represent inter-item correlation values, while the last four columns show item-total correlations of each item with each dimension (BKD, SKD, EKD) and Crombach's α values indicating internal consistency reliability.

For SF-12 structural validity analyses, 5,245 and 2,325 complete questionnaires were included in the training and testing datasets, respectively. IRT model fit results from the confirmatory factor analyses are reported in Table 4a, and model specifications are provided in Appendix 2. Overall, the bifactor structure showed the best fit (RMSEA = 0.052, 95% CI 0.047-0.056; TLI/CFI = 0.957/0.979) and offered a more parsimonious representation than the alternative models tested. In addition, bifactor model fit was evaluated in the validation dataset and was consistent with the training results (RMSEA = 0.050, 95% CI 0.043–0.058; TLI/CFI = 0.958/0.979). The general factor (SF-12$_G$) accounted for most of the common variance (Explained Common Variance-ECV=77.0%), whereas the specific factors explained substantially less after controlling for SF12$_G$ (PHC: 16.5%; MHC: 31.5%). This pattern is reflected in the Test Information Curve of the SF-12 shown in Appendix 2. Excellent reliability was observed for the bifactor SF12 scores (ω_{SF12G}=0.92, ω_{PHC}=0.90, ω_{MHC}=0.84).

Table 4a

	RMSEA [$RMSEA_5$-$RMSEA_{95}$]	SRMSR	TLI	CFI	AIC	BIC	SABIC
Model 1a - Uncorrelated Two-Factor structure	0.064 [0.059-0.068]	0.053	0.934	0.966	126984.1	127340.4	127168.8
Model 2a - Correlated Two-Factor structure	0.065 [0.061-0.07]	0.053	0.931	0.966	126986.1	127349	127174.2
Model 3a - Alternative Correlated Two-Factor structure	0.072 [0.068-0.076]	0.067	0.915	0.939	127824	128114.4	127974.5
Model 4a - Bifactor structure	0.052 [0.047-0.056]	0.051	0.957	0.979	126585.5	126948	126773.6
Model 4a - Bifactor structure on the validation dataset	0.050 [0.043-0.058]	0.051	0.958	0.979	54418.4	54734.7	54560

Table 4a: Fit statistics for SF-12 IRT models. *RMSEA = root mean square error of approximation with 90% CI; SRMSR = standardized root mean square residual; TLI = Tucker–Lewis Index; CFI = comparative fit index; AIC = Akaike information criterion; BIC = Bayesian information criterion; SABIC = sample-size–adjusted BIC. Lower RMSEA, SRMSR, AIC, BIC, and SABIC and higher TLI and CFI indicate better model fit. Details on the structure and coefficients of the models in the table are reported in Appendix 1*

For KDQoL-36 disease-specific scales structural validity analyses, 4,098 and 1,756 complete questionnaires were included in the training and testing datasets, respectively. IRT model fit results from the confirmatory factor analyses are reported in Table 4b, and model specifications are provided in Appendix 2. Overall, the bifactor structure showed the best fit (RMSEA = 0.037, 95% CI 0.035-0.040; TLI/CFI = 0.960/0.969) and offered a more parsimonious representation than the alternative models tested. In addition, bifactor model fit was evaluated in the validation dataset and was consistent with the training results (RMSEA = 0.042, 95% CI 0.039–0.046; TLI/CFI = 0.952/0.963). The general factor (KDS_G) accounted for most of the common variance (ECV=64.3%), whereas the specific factors explained substantially less after controlling for KDS_G (BKD: 43.7%; SKD: 31.5%, EKD: 21.6%). TICs are shown in Appendix 1.3. KDS_G had excellent reliability (ω = 0.94), while the disease-specific scales demonstrated slightly lower unadjusted reliability (ω_{BKD} = 0.87, ω_{SKD} = 0.87, and ω_{EKD} = 0.88).

	RMSEA [$RMSEA_5$-$RMSEA_{95}$]	SRMSR	TLI	CFI	AIC	BIC	SABIC
Model 2a - Correlated Two-Factor structure	0.047 [0.045-0.049]	0.044	0.937	0.945	230568.3	231345.4	230954.6
Model 2b - Uncorrelated Two-Factor structure	0.079 [0.077-0.081]	0.192	0.82	0.841	234310.7	235068.9	234687.6
Model 3b - Bifactor structure	0.037 [0.035-0.04]	0.033	0.96	0.969	229367.1	230276.9	229819.3
Model 3b - Bifactor structure on the validation dataset	0.042 [0.039-0.046]	0.04	0.952	0.963	100092.6	100880.4	100422.9

Table 4b: Fit statistics for Kidney disease-specific scales IRT models. *RMSEA = root mean square error of approximation with 90% CI; SRMSR = standardized root mean square residual; TLI = Tucker–Lewis Index; CFI = comparative fit index; AIC = Akaike information criterion; BIC = Bayesian information criterion; SABIC = sample-size–adjusted BIC. Lower RMSEA, SRMSR, AIC, BIC, and SABIC and higher TLI and CFI indicate better model fit. Details on the structure and coefficients of the models in the table are reported in Appendix 1.*

In construct validity analyses, IRT-based PHC and MHC scores closely aligned with their CTT counterparts (PHC: $R^2 = 0.774$, $p < 0.001$; MHC: $R^2 = 0.773$, $p < 0.001$), with regression equations $PHC_{CTT} = 4.44 + 0.88 \cdot PHC_{IRT}$ and $MHC_{CTT} = 5.58 + 0.87 \cdot MHC_{IRT}$.

IRT-based disease-specific scores also aligned with CTT scores (BKD: $R^2 = 0.916$, $p < 0.001$; SKD: $R^2 = 0.941$, $p < 0.001$; EKD: $R^2 = 0.774$, $p < 0.001$), with regression equations $BKD_{CTT} = 8.28 + 0.89 \cdot BKD_{IRT}$, $SKD_{CTT} = 10.83 + 0.87 \cdot SKD_{IRT}$, and $EKD_{CTT} = 11.83 + 0.84 \cdot EKD_{IRT}$. All regression lines showed a strong linear association between IRT-based and CTT-based scores, with little bias. High correlations indicated good construct validity, as good convergence was observed between the IRT-based scores and traditional scoring methods.

A bridge table between IRT-based raw scores θ and T-scores is reported in Appendix 3.

Computerized Adaptive Testing

Technical details on computed CAT are reported in Appendix 2.

The Bland-Altman plots from the tuning of the stopping rule for the SF12 are shown in Figures 1a–c. The plots showed negligible bias and acceptable imprecision between CAT and full-length IRT scores across general and sub-dimensions. CAT accuracy was evaluated through its agreement with full IRT scores. CAT scores closely mapped into full-length $SF12_{G\text{-}IRT}$, PHC_{IRT} and MHC_{IRT} scores ($SF12_{G\text{-}IRT}$: $R^2 = 0.93$, $p < 0.001$; PHC_{IRT}: $R^2 = 0.95$, $p < 0.001$; MHC_{IRT}: $R^2 = 0.96$, $p < 0.001$) and regression analysis showed a strong linear association for all the three components ($SF12_{G\text{-}IRT} = 0.0086 + 0.93 * SF12_{G\text{-}CAT}$; $PHC_{IRT} = -0.0017 + 0.94 * PHC_{CAT}$–; $MHC_{IRT} = -0.0128 + 0.96 * MHC_{CAT}$).

The number of items administered per patient applying the identified CAT stopping rule varied between 3 and 12, with the majority of participants receiving between 5 and 7 items (Figure 1d): only 9% of patients needed to complete the entire survey. Overall, there was a 38.4% reduction in items administered.

The Bland-Altman plots resulting from the tuning of the stopping rule for the kidney disease-specific scales are shown in Figures 2a-d. The plots showed negligible bias and acceptable imprecision between CAT and full-length IRT scores across general and sub-dimensions. CAT accuracy was evaluated through its agreement with full IRT scores. CAT scores closely mapped into full-length $KDS_{G\text{-}IRT}$, BKD_{IRT}, SKD_{IRT} and EKD_{IRT} scores ($KDS_{G\text{-}IRT}$: $R^2 = 0.97$, $p < 0.001$; BKD_{IRT}: $R^2 = 0.93$, $p < 0.001$; SKD_{IRT}: $R^2 = 0.94$, $p < 0.001$; EKD_{IRT}: $R^2 = 0.99$, $p < 0.001$) and regression analysis showed a strong linear association for all the four components ($KDS_{G\text{-}IRT}$ =-0.003+0.97*$KDS_{G\text{-}CAT}$; BKD_{IRT} = -0.001+0.93*BKD_{CAT}; SKD_{IRT} = -0.006+0.94*SKD_{CAT}; EKD_{IRT} = -0.0002+0.99*EKD_{CAT}).

The number of items administered per patient applying the identified CAT stopping rule to KDQoL-36 disease-specific scales varied between 12 and 24, with the majority of participants receiving between 13 and 20 items (Figure 2e): 75% of patients would be asked to complete a reduced number of items and a reduction of 22% items administered was gained, overall.

Agreement across scoring approaches was consistently excellent across all domains, with $ICC_{agreement}$ ranging from 0.961 (MHC; 95% CI 0.958–0.963) to 0.994 (KDS_G; 95% CI 0.993–0.994), and high values for all the scales: $SF12_G$ (0.982; 95% CI 0.981–0.983), PHC (0.968; 95% CI 0.965–0.970), BKD (0.986; 95% CI 0.985–0.987), SKD (0.989; 95% CI 0.989–0.990), and EKD (0.970; 95% CI 0.967–0.972).

The corresponding $SDC_{agreement}$ values were small, ranging from 2.2 T-score points (KDS_G) to 5.5 T-score points (MHC), with intermediate thresholds for SKD (2.9), BKD (3.3), SF12G (3.7), EKD (4.8), and PHC (5.0). These values reflect the minimum score changes that are likely to be perceived as meaningful by respondents within each respective domain.

Construct Validity and Known Group Analysis

Forty-one percent of patients experienced at least one hospital admission within 12 months after survey administration. The HRS moderately discriminated across hospital admission likelihood (AUC = 0.67). The coefficients of the models and their associated significance are summarized in Appendix 4. Construct validity, assessed through known-groups validity, was evaluated using HRS quintiles.

Both full-length IRT-based and CAT-based $SF\text{-}12_G$, PHC, and MHC scores declined significantly across HRS quintiles ($SF12_G$: both η^2=0.01, p for trend<0.001; PHC: both η^2=0.02, p for trend<0.001; MHC: both η^2=0.01, p for trend<0.001), indicating an association between HRS-used as an objective indicator of declining health-and poorer perceived general quality of life.

In contrast, the association between IRT-based and CAT-based KDS_G, BKD, SKD, and EKD scores showed no significant or consistent trends across HRS quintiles. All trend test p-values were statistically non-significant, and the effect sizes were very small (η^2 ranging from 0.001 to 0.007).

.Higher $SF12_G$ score was strongly associated with reduced hospitalization after adjustment for clinical and demographic confounders (Odds Ratio: OR=0.90, p<0.001). Similarly, a 5-point increase in PHC_{IRT} and MHC_{IRT} was associated with 7.1% (p<0.001) and 8.6% (p<0.001) reduction in hospitalization risk, respectively. The use of scores obtained through CAT yielded consistent risk estimates (not shown).

We observed a statistically significant association between KDS_G and hospitalization risk after adjustment for clinical and demographic confounders (OR = 0.94, p<0.05). Among the disease specific sub-scales, only SKD_{IRT}, and its corresponding CAT version, was significantly associated with hospitalization risk (SKD_{IRT}: OR=0.94, p = 0.027; SKD_{CAT}: OR=0.95, p = 0.038).

DISCUSSION

The KDQoL-36 is the most widely used PROM in nephrology, but real-world evidence suggests that its length hinders meaningful implementation in practice. In this large European dialysis cohort, we applied multidimensional IRT to the KDQoL-36 and demonstrated robust psychometric performance and practical benefits.

Our results suggest that a bifactor model provides the best fit for both the SF-12 and the kidney-disease specific scales surveys in patients receiving dialysis. To our knowledge this is the first study highlighting a strong general factor for the SF-12 which aligns with previous findings of a strong correlation between the physical and mental health component of the SF-12. Such an interconnection between physical and mental health may be even stronger among dialysis patients and justify a bifactor structure, given the known effect of uremic toxins on fatigue, sleep disturbances and cognition that may overlap with symptoms of depression and anxiety[25].

We observed that a single overarching QoL dimension captures most response variance in the disease-specific items, with smaller domain-specific contributions. The KDSG factor explained ~64% of common variance in the 24 kidney-targeted items, whereas subdomains (BKD, SKD, EKD) each explained <45% after accounting for KDSG. Reliability was excellent for the general factor ($\omega \approx 0.94$) and remained high for subscales ($\omega \approx 0.87–0.90$). These findings partly align with data from a large U.S. dialysis organization[26] showing that a composite Kidney Summary Score derived from the three kidney-targeted scales improved model fit, suggesting a summary metric may enhance clinical interpretability. However, individual domain scores may still help clinicians pinpoint specific issues (e.g., symptom management vs illness burden), and the KDQoL-36 supports reporting both domain and summary scores as needed.

We provide the calibration parameters of our item bank in appendix to enable and encourage replication studies by independent researchers (Appendix 5). The high correlation between CTT and IRT scores ($R^2 \approx 0.77–0.94$), together with the expected known-groups differences, is consistent with previous findings that Kidney Disease Symptoms and $SF\text{-}12_{PHC}$ domains show a graded association with comorbidity burden, an important correlate of hospitalization risk in CKD[27].

[665]While the content of the existing KDQoL-36 aligns with core outcomes domains prioritized by patients, doctors, and nurses in international consensus exercises[4], the survey's respondent burden poses a challenge to its meaningful implementation in routine clinical care. Evidence on CAT approaches in kidney care is still limited; for example, a Dutch group evaluated the validity and reliability of multiple PROMIS CAT domains in patients with advanced CKD, supporting feasibility while also highlighting that broader implementation is still emerging[28]. In this context, our CAT simulations demonstrate the potential for substantial efficiency gains: on average, patients would need to answer as few as 5–7 items for the SF-12 and 13–20 items for the KDQoL-36 disease-specific domain to achieve precise measurement, corresponding to an estimated 25% reduction in administered items with minimal bias and excellent agreement with full-length scores.

Ware et al. (2019) showed that a brief IRT-based CKD-QoL scale outperformed the KDQoL-36's multiple scales on several validity tests[27] and significantly reduced respondent burden. Their conclusion that a concise, CKD-targeted PRO measure can be more valid and far more efficient than a longer survey, closely parallels with our findings. However, a 6-item CKD-QoL may be less actionable in routine care because it provides limited detail on which specific domains (e.g., symptoms, burden, physical or emotional functioning) are driving a low score. Our study specifically addresses concerns about survey fatigue and clinical workload of the KDQoL-36 while preserving the diversity of content explored by the instrument.

Furthermore, Rincon Bello et al. have previously shown that KDQoL-36 scores are associated with increased risk of hospitalization and mortality after adjustment for clinical and socio-demographic factors, suggesting that PROMs capture elements of health not fully explained by clinical parameter[29]. We partially confirmed these findings in the present study by showing a similar pattern of association with IRT-based KDQoL-36 scores with hospitalization risk highlighting HRQoL as an independent prognostic indicator. In this way, tracking PROs is not only about patient satisfaction but may also provide an early warning for adverse outcomes, enabling preemptive care adjustments.

Finally, for PRO measures to truly impact clinical practice, clinicians must be able to interpret the scores and know how to act on them. To ensure interpretability, we provide a bridge between raw IRT scores into T-scores and explored actionability thresholds (e.g. standard error-based minimal important differences) thus providing reference points to contextualize individual scores[30]. By anchoring KDQoL-36 results to a large European dialysis reference population and by mapping them into published US norms as well, we can define what a "low" or "high" score is, and what magnitude of change might be clinically relevant.

This study has several strengths. To our knowledge, this is the first study to validate a multidimensional IRT model for the KDQoL-36 in a large, international dialysis cohort and to evaluate a simulated CAT for this instrument. The large sample size provided excellent power for psychometric analysis and allowed rigorous split-sample validation. Including multiple countries and languages increases the generalizability of our findings within European dialysis settings. We also quantified the potential real-world impact of CAT implementation; these findings are immediately relevant for quality improvement planning. Finally, our study used real-world clinical data, enhancing its relevance to practice.

However, we acknowledge a few limitations. First, our cohort comes from a single European provider network, which may limit generalizability to other regions or healthcare systems. Future research should assess the cross-cultural validity of the KDQoL-36 IRT model (e.g., via differential item functioning analyses). Second, we could not evaluate responsiveness to change or longitudinal outcomes. Third, while we examined one aspect of construct validity, other known-group tests may further support score interpretability.

In summary, we provided evidence that the KDQoL-36, when calibrated on a large dialysis population, is a reliable and valid PRO instrument that can be streamlined through adaptive testing without sacrificing clinical information. Through the development of rigorously validated, patient-acceptable PROMs, we move closer to judging treatment success not only by laboratory or survival metrics, but also by how patients feel and function in daily life.

Declarations

Funding

REMOVED FOR ANONYMIZATION

Competing Interests

REMOVED FOR ANONYMIZATION

Author Contributions

REMOVED FOR ANONYMIZATION

Ethics approval

This study was performed in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Hospital Clínic de Barcelona (CEIm), Barcelona, Spain (Reg. HCB/2025/0638; meeting acta 12/2025, dated 26 June 2025).

Funding declaration

This research received no external funding.

References

1. Neri L, Brancaccio D, Rocca Rey LA, Rossa F, Martini A, Andreucci VE. Social support from health care providers is associated with reduced illness intrusiveness in hemodialysis patients. *Clin Nephrol*. 2011;75(2):125-134. doi:10.5414/CNP75125
2. Flythe JE, Watnick S. Dialysis for Chronic Kidney Failure: A Review. *JAMA*. 2024;332(18). doi:10.1001/jama.2024.16338
3. Boenink R, Bonthuis M, Boerstra BA, et al. The ERA Registry Annual Report 2022: Epidemiology of Kidney Replacement Therapy in Europe, with a focus on sex comparisons. *Clin Kidney J. Oxford University Press*. 2025;18(2). doi:10.1093/ckj/sfae405
4. Tong A, Manns B, Hemmelgarn B, et al. Establishing Core Outcome Domains in Hemodialysis: Report of the Standardized Outcomes in Nephrology-Hemodialysis (SONG-HD) Consensus Workshop. *American Journal of Kidney Diseases*. 2017;69(1):97-107. doi:10.1053/j.ajkd.2016.05.022
5. Bennett P, Warren M, Aydin Z, et al. Kidney Disease: Improving Global Outcomes (KDIGO) Workshop on the Nurse's Role in Managing the Symptoms of People Receiving Dialysis. *Kidney Int Rep*. 2025;10(2):313-320. doi:https://doi.org/10.1016/j.ekir.2024.11.029
6. Paul E. Stevens SBAJJCBF. . . AL. *KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease*. 2024. doi:10.1016/j.kint.2023.10.018
7. Cohen DE, Lee A, Sibbel S, Benner D, Brunelli SM, Tentori F. Use of the KDQOL-36^{TM} for assessment of health-related quality of life among dialysis patients in the United States. *BMC Nephrol*. 2019;20(1). doi:10.1186/s12882-019-1295-0
8. Tang E, Yantsis A, Ho M, et al. Patient-Reported Outcome Measures for Patients With CKD: The Case for Patient-Reported Outcomes Measurement Information System (PROMIS) Tools. *American Journal of Kidney Diseases*. 2024;83(4):508-518. doi:10.1053/j.ajkd.2023.09.007
9. Hays RD, Kallich JD, Mapes DL, Coons SJ, Carter WB. Development of the Kidney Disease Quality of Life (KDQoL^{TM}) Instrument. *Quality of Life Research*. 1994;3(5):329-338. doi:10.1007/BF00451725
10. Chesnaye NC, Meuleman Y, de Rooij ENM, et al. Health-Related Quality-of-Life Trajectories over Time in Older Men and Women with Advanced Chronic Kidney Disease. *Clinical Journal of the American Society of Nephrology*. 2022;17(2). https://journals.lww.com/cjasn/fulltext/2022/02000/health_related_quality_of_life_trajectories_over.8.aspx
11. Janmaat CJ, van Diepen M, Meuleman Y, et al. Kidney function and symptom development over time in elderly patients with advanced chronic kidney disease: results of the EQUAL cohort study. *Nephrology Dialysis Transplantation*. 2021;36(5):862-870. doi:10.1093/ndt/gfz277
12. Finkelstein FO, Finkelstein SH. Time to Rethink Our Approach to Patient-Reported Outcome Measures for ESRD. *Clinical Journal of the American Society of Nephrology*. 2017;12(11). https://journals.lww.com/cjasn/fulltext/2017/11000/time_to_rethink_our_approach_to_patient_reported.25.aspx
13. Aiyegbusi OL, Roydhouse J, Rivera SC, et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat Commun. Nature Research*. 2022;13(1). doi:10.1038/s41467-022-33826-4
14. Neri L, Ponce P, Matias N, Stuard S, Cromm K. Clinical target achievement is associated with better quality of life among dialysis patients: results from a continuous quality improvement program in a Portuguese healthcare network. *Quality of Life Research*. 2020;29(10):2705-2714. Accessed January 29, 2021. https://link.springer.com/article/10.1007/s11136-020-02543-0
15. Porter AC, Lash JP, Xie D, et al. Predictors and outcomes of health-related quality of life in adults with CKD. *Clinical Journal of the American Society of Nephrology*. 2016;11(7):1154-1162. doi:10.2215/CJN.09990915
16. Wagner LI, Schink J, Bass M, et al. Bringing PROMIS to practice: Brief and precise symptom screening in ambulatory cancer care. *Cancer*. 2015;121(6):927-934. doi:10.1002/cncr.29104
17. Wong LH, Meeker JE. The promise of computer adaptive testing in collection of orthopaedic outcomes: an evaluation of PROMIS utilization. *J Patient Rep Outcomes. Springer Science and Business Media Deutschland GmbH*. 2022;6(1). doi:10.1186/s41687-021-00407-w
18. Hays RD, Kallich J, Mapes D, et al. *Kidney Disease Quality of Life Short Form (KDQoL-SF), Version 1.3: A Manual for Use and Scoring*. RAND Corporation; 1997. https://www.rand.org/pubs/papers/P7994.html
19. Ricardo AC, Hacker E, Lora CM, et al. Validation of the Kidney Disease Quality of Life Short Form 36 (KDQOL-36) US Spanish and English versions in a cohort of Hispanics with chronic kidney disease. *Ethn Dis*. 2013;23(23(2)):202-209. www.FACITtrans.org,
20. Yang F, Wang VW, Joshi VD, Lau TWL, Luo N. Validation of the English Version of the Kidney Disease Quality of Life Questionnaire (KDQOL-36) in Haemodialysis Patients in Singapore. *The Patient - Patient-Centered Outcomes Research*. 2013;6(2):135-141. doi:10.1007/s40271-013-0015-2
21. Mokkink LB, Elsman EBM, Terwee CB. *COSMIN Manual: COSMIN Guideline for Conducting Systematic Reviews of Patient-Reported Outcome Measures Version 2.0 (Including the COSMIN Risk of Bias Checklist Version 3 and the COSMIN Criteria for Good Measurement Properties Version 2.0)*. COSMIN; 2024. https://www.cosmin.nl/wp-content/uploads/COSMIN-manual-V2_final.pdf
22. Forero CG, Vilagut G, Adroher ND, Alonso J, Investigators Esem. Multidimensional item response theory models yielded good fit and reliable scores for the Short Form-12 questionnaire. *J Clin Epidemiol*. 2013;66(7):790-801. doi:10.1016/j.jclinepi.2013.02.007
23. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: A systematic review. *PLoS One*. 2012;7(5). doi:10.1371/journal.pone.0037908
24. Copay AG, Subach BR, Glassman SD, Polly Jr. DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal*. 2007;7(5):541-546. doi:10.1016/j.spinee.2007.01.008
25. Rose M, Fischer FH, Liegl G, et al. The CONVINCE randomized trial found positive effects on quality of life for patients with chronic kidney disease treated with hemodiafiltration. *Kidney Int*. 2024;106(5):961-971. doi:10.1016/j.kint.2024.07.014
26. Peipert JD, Bentler PM, Klicko K, Hays RD. Psychometric Properties of the Kidney Disease Quality of Life 36-Item Short-Form Survey (KDQOL-36) in the United States. *American Journal of Kidney Diseases*. 2018;71(4):461-468. doi:10.1053/j.ajkd.2017.07.020
27. Ware JEJr, Richardson MM, Meyer KB, Gandek B. Improving CKD-Specific Patient-Reported Measures of Health-Related Quality of Life. *Journal of the American Society of Nephrology*. 2019;30(4).

https://journals.lww.com/jasn/fulltext/2019/04000/improving_ckd_specific_patient_reported_measures.16.aspx

28. Van Der Willik EM, Milders J, Bart JAJ, et al. Discussing results of patient-reported outcome measures (PROMs) between patients and healthcare professionals in routine dialysis care: a qualitative study. *BMJ Open*. 2022;12(11). doi:10.1136/bmjopen-2022-067044
29. Rincon Bello A, Ion Titapiccolo J, Berdud Godoy I, et al. Better health-related quality of life is associated with prolonged survival and reduced hospitalization risk among dialysis-dependent chronic kidney disease patients: a historical cohort study. *BMC Nephrol*. 2024;25(1). doi:10.1186/s12882-024-03835-0
30. Chan EKH, Edwards TC, Haywood K, Mikles SP, Newton L. Implementing patient-reported outcome measures in clinical practice: a companion guide to the ISOQOL user's guide. *Quality of Life Research*. 2019;28(3):621-627. doi:10.1007/s11136-018-2048-4

Figures

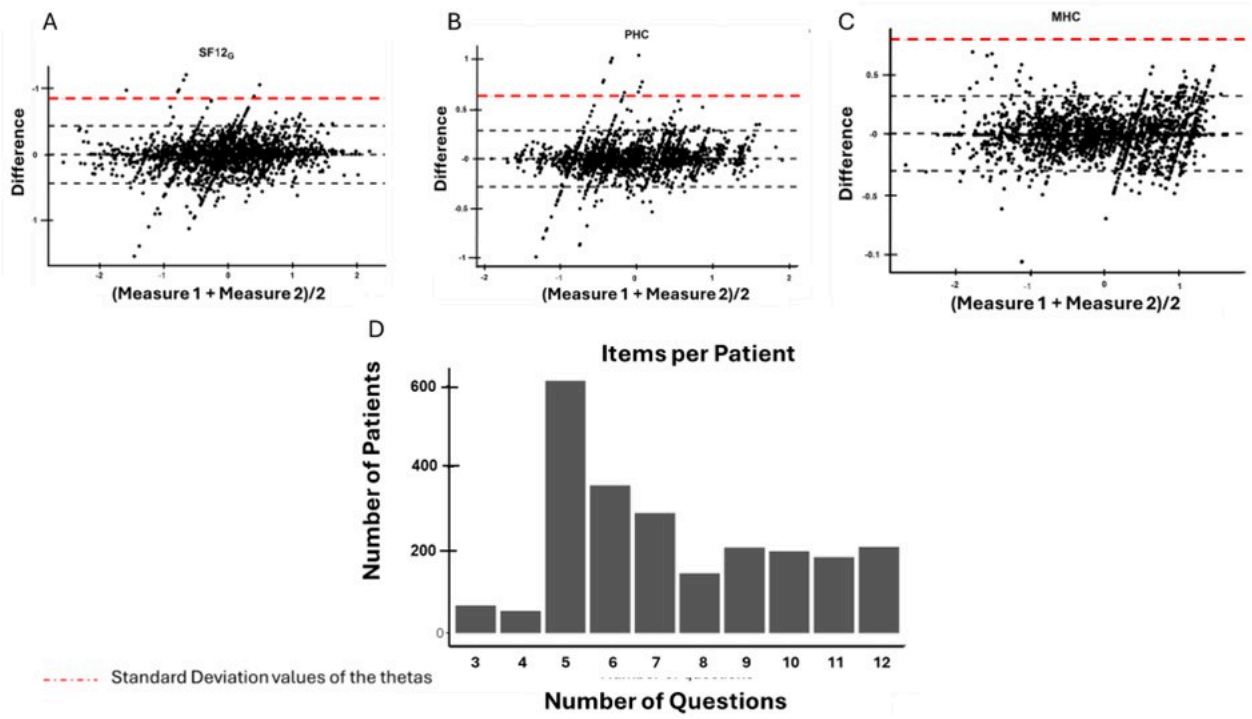


Figure 1

Agreement between IRT and CAT measurements of SF-12 scores and distribution of item burden. *(**A–C**) Bland–Altman plots comparing two administrations/measurements of the same construct ("Measure 1:IRT" and "Measure 2:CAT") for three SF-12 outcomes: (**A**) SF-12$_G$ (General health), (**B**) Physical Health Composite (PHC), and (**C**) Mental Health Composite (MHC). The x-axis shows the average of the pair for each patient, (Measure 1+Measure 2)/2; the y-axis shows the within-patient difference, Measure 1−Measure 2. Each point represents one patient. The solid horizontal line denotes the mean difference (bias). Flanking grey dashed lines depict the limits of agreement (mean difference ± 1.96 SD of the differences). Where shown, a sloped dashed fit illustrates the trend of the difference across the measurement range (assessment of proportional bias). The red dot-dash horizontal reference line marks the standard deviation values of Measure 1, a clinically important difference value for context. (**D**) Histogram of questionnaire length per patient used to derive the SF-12 scores. Bars show the number of patients by the count of items completed (3–12 items), illustrating the distribution of response burden in the sample.*

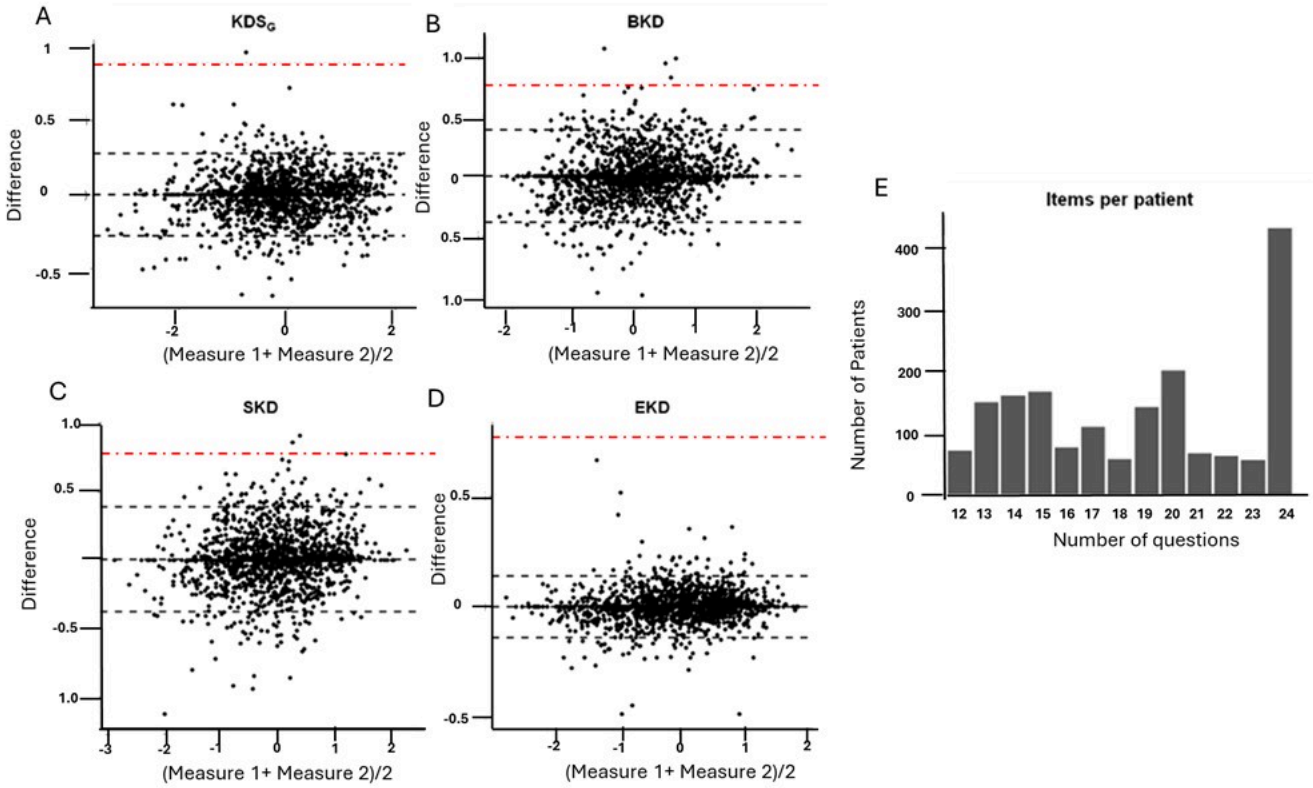


Figure 2

Agreement between IRT and CAT measurements of KDQoL-36 disease-specific scales scores and distribution of item burden.

*(**A–D**) Bland–Altman plots comparing two administrations/measurements of the same construct ("Measure 1:IRT" and "Measure 2:CAT") for the four KDQoL-36 disease-specific scales outcomes: (**A**) Kidney-Disease Specific General score (KDS$_G$), (**B**) Burden of Kidney Disease(BKD), (**C**) Symptoms of Kidney Disease (SKD), and (**D**) Effects of Kidney Disease (EKD). The x-axis shows the within-patient average (Measure 1+Measure 2)/2; the y-axis shows the within-patient difference (Measure 1−Measure 2). Each dot represents one patient. The solid horizontal line is the mean difference (bias); grey dashed lines indicate the 95% limits of agreement (mean difference ±1.96 SD of the differences). The red dot-dash horizontal line marks the standard deviation value of Measure 1. (**E**) Histogram showing the number of disease-specific items completed per patient (range 12–24 items from the KDQOL-36 kidney-disease–targeted sections), illustrating variability in response burden; the tallest bar at 24 items indicates the number of patients needing to complete the full set of items*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendices230226.docx