Prepare; before starting dialysis : outcomes in patients with CKD stage 4-5

Voskamp, P.W.M.

3

MEASURING QUALITY OF LIFE IN PRE–DIALYSIS PATIENTS; COMPARING THE SF–12 AND EQ–5D TO THE SF–36

Pauline WM Voskamp, Merel van Diepen, Chloe CA Grim, Ellen K Hoogeveen, Friedo W Dekker

Abstract

Introduction: The influence of chronic kidney disease on quality of life (QoL) is usually measured with the Short Form-36 (SF-36) questionnaire. Frequently used shorter QoL questionnaires are the Short Form-12 (SF-12) and the EuroQol (EQ-5D), but these have not yet been validated in pre-dialysis patients. The aim of this study was to validate the SF-12 and the EQ-5D questionnaire in pre-dialysis patients.

Methods: In a multi-center cohort study, incident pre-dialysis patients (>18y) were included between 2004-2011 and followed until renal replacement therapy, death, or October 2016. These patients completed QoL questionnaires every 6 months. Pearson correlation coefficients between the SF-12 and the SF-36, and between the EQ-5D and the SF-36 were calculated. Bland-Altman plots were used to assess limits of agreement, using Z-scores. Associations between the different measurement tools with time to start of dialysis and death were assessed with Cox regression analyses.

Results: The correlation between the Mental Component Score (MCS)-12 – MCS-36 and the Physical Component Score (PCS)-12 – PCS-36 were both 0.95, for EQ-5D – MCS-36 this was 0.37, and for EQ-5D – PCS-36 this was 0.57. All mean differences were 0, limits of agreement ranged from -0.60 to 0.60 for MCS-12 – MCS-36 to -2.05 to 2.05 for EQ-5D – MCS-36. The adjusted HRs for start of dialysis and for death were similar using the SF-36 and the SF-12. The EQ-5D HR for start of dialysis was similar to the PCS-36 association. The EQ-5D HR for death was similar to the MCS-36.

Conclusion: The SF-12 has better agreement with the SF-36 compared to the EQ-5D and can be used as a substitute for the SF-36 in pre-dialysis patients.

Introduction

Chronic kidney disease (CKD) is a leading cause of premature death and also an important disabling factor in daily life.[1-3] Patients with Chronic Kidney Disease, especially those with an estimated glomerular filtration rate (eGFR) below 30 ml/min/1.73m^2 suffer from a great range of physical and psychological symptoms and are known to have a lower quality of life.[4, 5] Currently, focus is shifting from only attempting to improve prognosis in CKD patients towards also improving these so called patient reported outcomes.

One of the most essential patient reported outcomes is the health related quality of life (HRQoL). In CKD patients a decrease in HRQoL is associated with an increased risk of end stage renal disease (ESRD) and mortality.[6, 7] In dialysis patients a lower HRQoL is associated with hospitalization, dialysis vintage, lower residual kidney function and mortality.[8, 9]

The most frequently used questionnaire to assess HRQoL is the Short Form 36 (SF-36). The SF-36 contains 36 items, which measure health on eight dimensions, as well as overall physical and mental component scores (PCS-36 and MCS-36).[10] Unfortunately, the SF-36 can take up to 12 minutes for a patient to complete and is therefore often considered too long for inclusion in large scale monitoring efforts. Therefore, shorter questionnaires are used when patients are asked to fill in multiple questionnaires.

Two frequently used shorter HRQoL questionnaires are the Short Form-12 (SF-12) and the EuroQol questionnaire (EQ-5D).[11-15] The SF-12 is an abbreviated version of the SF-36, reducing the number of items down to 12. In contrast to the SF-36, which measures eight health domains, the SF-12 measures health only by physical and mental component scores (PCS-12 and MCS-12). The EQ-5D consists of a health valuation questionnaire and a visual analogue scale (VAS). The EQ-5D questionnaire results in a single index score based on five questions covering; mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The VAS records the patient's self-reported health valuation on a vertical axis ranging from 0 to 100. The questionnaire of the EQ-5D is the part that is generally used as a HRQoL measurement tool.[13, 16, 17]

Ideally, a short questionnaire improves efficiency while remaining almost as valid as a long questionnaire, such as the standard SF-36. The SF-12 and EQ-5D have been validated for multiple chronic diseases in multiple populations, including dialysis and kidney transplant patients, but have not yet been validated in pre-dialysis patients.[13, 18, 19] Since pre-dialysis patients differ from the aforementioned patients and have a specific medical treatment, perceived physical and mental health may be dissimilar. Therefore, the aim of this study is to validate the SF-12 and the EQ-5D questionnaire in pre-dialysis patients using the SF-36 as a reference.

Methods

Study design and population

The PREPARE-2 study is a prospective cohort study of incident pre-dialysis care patients (≥18 y) who had an estimated glomerular filtration rate (eGFR) of less than 30 ml/min/1.73m^2 and progressive renal function loss. Patients with a failing kidney transplant who were transplanted at least one year ago were also eligible for inclusion. For this study all patients with complete SF-36 component scores were included. The PREPARE-2 study has been described in detail elsewhere.[20] In brief, patients were recruited in one of 25 nephrology specialized pre-dialysis outpatient clinics in the Netherlands between July 2004 and June 2011. All patients were treated by their nephrologist in accordance with the treatment guidelines of the Dutch Federation of Nephrology, guidelines partly based on the K/DOQI and EBPG guidelines.[21-24] Patients were followed from the start of pre-dialysis care until start of dialysis, kidney transplantation, death or censoring. Censoring was applied in case of: refusal for further participation, recovery of kidney function, moving to an outpatient clinic not participating in the PREPARE-2 study, loss to follow up or October 2016 (end of follow up), whichever came first. This study was approved by the medical ethics committee or institutional review boards (as appropriate) of all participating centers. Written informed consent was obtained from all patients.

Health Related Quality of Life Questionnaires

Both the SF-36 and the EQ-5D were part of the HRQoL questionnaire in the PREPARE-2 study, and were measured every six months. The SF-36 Health Survey consists of 36 multiple choice questions with standardized responses, of which 35 are organized into eight multi item domains (supplemental table 1).[25] The 36th question asks how health has changed over a specified time period (in this study: three months), this question is not included in the eight domains. These eight domains are used to form two component scores, referred to as the PCS-36, consisting of physical functioning, role limitations due to physical problems and bodily pain dimensions, and the MCS-36, consisting of social functioning, role limitations due to emotional problems and mental health dimensions. The component scores share the domains general health and vitality. At least half of the items per domain have to be complete for a domain to be scored. The component scores are a weighted summary score of the domains.[26] The PCS-36 and MCS-36 are calculated using norm based scoring, which employs linear transformation to achieve standardized scores with a mean of 50 and a standard deviation of 10.[27] A score over 50 can be interpreted as above average. We used the United States (US) population as a reference group.[28] We used the first version of the SF-36 questionnaire.[25] The SF-12 is obtained by extracting 12 questions from the SF-36 (supplemental table 1). The SF-12 does not have enough items to assess eight domains but solely calculates the two component scores. In the SF-12 all items have to be complete in order to calculate physical and mental component scores.

[29] The PCS-12 and MCS-12 regression weights were calculated with norm based scoring to achieve standardized scores with a mean of 50 and a standard deviation of 10, using the US population as the reference group.[29, 30]

The EQ-5D questionnaire measures quality of life by five dimensions (supplemental table 2a).[31] The five dimensions are scored as follows: no, some, or extreme problems. The five dimensions are summarized in a single score by applying a weight to each dimension score according to the time trade-off valuation technique (TTO), explained in detail elsewhere.[32] In short, the TTO creates a set of values based on the opinion of a group Dutch citizens. This value set results in a so-called tariff score ranging from -0.33 to 1.00. Higher scores indicate a better quality of life. A score of zero indicates that the quality of life is equal to death.

Statistical analyses

Baseline characteristics were presented as mean ± standard deviation (SD) for normally distributed continuous variables, skewed continuous variables as median with interquartile range (IQR). Multiple imputation was used to minimize the risk of bias and obtain correct standard errors.[33] Missing values of PCS-12 and MCS-12 scores, EQ-5D scores at baseline, and of possible confounders were imputed (using 10 repetitions).

Both construct validity and external validity were investigated. We defined construct validity as the extent to which the short questionnaires actually measure QoL. We defined external validity as the performance of the short questionnaires when measuring external constructs. Construct validity was measured by comparing the SF-12 and EQ-5D with our reference, the SF-36. External validity was assessed by using symptoms, the VAS part of EuroQol, the health change question from the SF-36, start of dialysis and death as external constructs.

First, for construct validity, Pearson correlation coefficients were calculated for the MCS-12 and PCS-12 and the EQ-5D on the one hand with the MCS-36 and PCS-36 on the other hand. This was done at baseline, after six months and after 12 months of follow-up. In all analyses, the MCS-12 was compared to the MCS-36 and the PCS-12 was compared to the PCS-36. The EQ-5D scores were compared with both the MCS-36 and the PCS-36. Changes over time in the aforementioned questionnaire scores were calculated and Pearson correlation coefficients were calculated for these changes. The Pearson correlation coefficient was favored over the intra-class correlation coefficient, since the EQ-5D has a different scoring system as compared with the SF questionnaires. This results in a systematic error which is not of interest for validating the questionnaires, which is ignored when using a Pearson correlation coefficient.[34] Second, agreement between the SF-12 and the EQ-5D on the one hand and the SF-36 on the other was assessed with Bland –Altman plots.[35] In this plot individual differences in scores are plotted against the mean of two different questionnaire measurements (e.g. MCS-12 ***vs*** MCS-36, PCS-12 ***vs*** PCS-36, EQ-5D ***vs*** MCS-36, EQ-5D ***vs*** PCS-36). Horizontal lines show the population average of all individual differ-

ences and the 95% limits of agreement (mean difference ±1.96 standard deviation of the differences). To create comparable scores for all questionnaires, and take the systematic error caused by the different scoring systems into account, Z-scores were calculated for the five different questionnaire components.

Next, for external validation, Pearson or Spearman correlation coefficients were calculated for the different questionnaire components as appropriate for our three numerical and ordinal external constructs, in this case symptom frequency, the VAS part of the EuroQol, and the health change question from the SF-36. Symptom frequency was measured in the patient questionnaires by evaluating the occurrence of 14 different symptoms, which resulted in a symptom frequency score ranging from 0 to 14. The VAS is the second part of the EuroQol, not used to measure HRQoL (supplemental table 2b). This is a thermometer on which patients can mark their own current health status from zero (worst imaginable health state) to 100 (best imaginable health state).[31] The health change question in the SF-36 ranges from 1; large increase in health over the past three months, to 5; large decrease in health over the past three months. The correlations with these three external constructs were calculated at baseline, after six months, and after 12 months of follow-up. The correlation for changes between baseline and six months of follow-up were calculated as well. Correlations with the external constructs should be similar for questionnaire scores measuring the same construct, in this case HRQoL.

Finally, the association between the five questionnaire components (MCS-36, PCS-36, MCS-12, PCS-12, and EQ-5D) and our two time to event external measurements start of dialysis and mortality were estimated with Cox proportional hazard regression analyses. Analyses were adjusted for the confounders age, sex, ethnicity, and primary kidney disease. For these analyses the different questionnaire components were transformed into Z-scores. This should result in similar hazard ratios for questionnaire scores measuring the same construct. Follow-up time was defined as time between baseline visit of the patient and the start of dialysis, renal replacement therapy (RRT), death, withdrawal or end of follow-up (October 2016). To estimate the median follow up time, a reversed Kaplan-Meier was used.

In a sensitivity analysis we repeated all analyses without using multiple imputation for missing values. All analyses were performed using SPSS version 23.0 for Windows.

Results

Of the 502 pre-dialysis patients in PREPARE-2, 433 patients had complete SF-36 component scores at baseline. Table 1 shows the baseline characteristics for the 433 patients with complete component scores and for the 69 patients with incomplete component scores. Patients with complete SF-36 component scores at baseline were more often women, less often had diabetes mellitus as primary kidney disease, suffered more often from cardiovascular disease, and had a higher kidney function as compared with patients with

Table 1. Baseline characteristics of 502 pre–dialysis patients with complete and incomplete SF–36 component scores

	Patients with complete SF-scores (N=433)	Patients with incomplete SF-scores (N=69)
Sex, men	287 (66)	54 (78)
Age, years	69 (56-76)	66 (49-75)
Ethnicity, Caucasian	401 (93)	61 (88)
Marital status, married or living together	304 (70)	-
Education		
Low	97 (23)	-
Intermediate	254 (59)	-
High	60 (14)	-
Other	13 (3)	-
Children, yes	358 (83)	-
Employment, yes	108 (25)	-
Primary kidney disease		
Diabetes Mellitus	57 (13)	15 (22)
Glomerulonephritis	58 (13)	9 (13)
Renal vascular disease	133 (31)	21 (31)
Other	185 (43)	24 (35)
Cardiovascular disease, yes [a]	182 (42)	25 (36)
Diabetes Mellitus, yes [b]	111 (26)	19 (28)
Congestive heart failure, yes [c]	51 (12)	10 (15)
Systolic blood pressure, mmHg	142 (22)	147 (21)
Hemoglobin, mmol/l	7.7 (0.9)	7.6 (0.8)
Serum creatinine, μmol/l	350 (112)	402 (130)
eGFR, ml/min/1.73 m^2 [d]	14.5 (11.4-18.9)	13.2 (10.0-16.3)
Proteinuria, g/24h	1.0 (0.4-2.1)	1.3 (0.6-2.6)

SF; Short Form

Values are given as number (percentage of the total), means ± SD or median (interquartile range).

[a] Defined as presence of coronary artery disease, a history of cardiovascular accident, peripheral vascular disease, or myocardial infarction. [b] Defined as the presence of diabetes mellitus as primary kidney disease or a history of diabetes mellitus. [c] Defined as a history of congestive heart failure [d] eGFR (estimated glomerular filtration rate) is calculated with the CKD EPI (Chronic Kidney Disease Epidemiology Collaboration) formula 2009.

Number of missings: Complete SF-36 component scores; marital status 2, children 4, education 9, employment 8, hemoglobin 54, proteinuria 218, eGFR 51, creatinine 51, systolic blood pressure 4. Incomplete SF-36 component scores; marital status 66, children 66, education 66, employment 66 hemoglobin 6, proteinuria 37, eGFR 13, creatinine 13t

an incomplete questionnaire. Education, current employment and having children were incomparable for the two groups, since these questions were part of the questionnaire and therefore missing for almost all patients with an incomplete questionnaire. All further analyses were performed in the 433 patients with complete SF-36 component scores.

Construct validity

At baseline correlations between the SF-12 and the SF-36 were high, with a Pearson correlation coefficient of 0.95 for both the MCS and the PCS (table 2a). The correlations between the EQ-5D and the SF-36 were lower, with a lower correlation between the EQ-5D and the MCS as compared with the correlation with the PCS. After 6 and 12 months of follow-up correlations were comparable (table 2a). The correlations for change over time in component scores were all slightly weaker as compared with the cross-sectional correlations (table 2b). However, the correlations between the SF-12 and SF-36 remained high (0.92 for the MCS, 0.88 for the PCS).

Table 2a. Pearson correlation coefficients of the SF–36 component scores at baseline, after 6 and 12 months with the time corresponding SF–12 component scores and EQ–5D score

	Baseline		6 months follow-up		12 months follow-up	
	SF-36 MCS	SF-36 PCS	SF-36 MCS	SF-36 PCS	SF-36 MCS	SF-36 PCS
SF-12 MCS	0.95	-	0.96	-	0.96	-
SF-12 PCS	-	0.95	-	0.96	-	0.95
EQ-5D	0.37	0.57	0.45	0.62	0.47	0.60

SF; Short Form, EQ-5D; EuroQol 5D, MCS; mental component summary, PCS; physical component summary

Table 2b. Pearson correlation coefficients of the changes in SF–36 component scores with the changes in the SF–12 component scores and EQ–5D score during follow–up

	0-6 months		6-12 months	
	SF-36 MCS	SF-36 PCS	SF-36 MCS	SF-36 PCS
SF-12 MCS	0.92	-	0.91	-
SF-12 PCS	-	0.88	-	0.89
EQ-5D	0.30	0.31	0.39	0.26

SF; Short Form, EQ-5D; EuroQol 5D, MCS; mental component summary, PCS; physical component summary

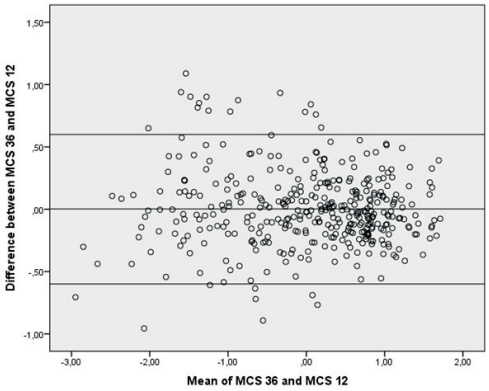
In order to create Z-scores the SD was calculated for each score. This was 9.4 for the MCS-36, 10.2 for the PCS-36, 9.4 for the MCS-12, 9.9 for the PCS-12, and 0.2 for the EQ-5D. In figure 1 the Bland-Altman plots show the individual Z-score differences between the different SF-12 and EQ-5D with the SF-36 components at baseline. Limits of agreement ranged from -0.60 to 0.60 SD for the MCS-12 scores, from -0.58 to 0.62 SD for the PCS-12 scores, from -2.05 to 2.05 SD for the EQ-5D and the MCS-36, and from -1.69 to 1.73 SD for the EQ-5D and the PCS-36. For changes over time in the component scores the limits of agreement were somewhat wider for all comparisons (Supplemental figure 1).

External validity

Table 3 shows the Pearson and Spearman correlation coefficients for the SF-36, SF-12 components, and the EQ-5D with the external constructs symptom frequency, VAS, and the health change question. The negative correlations indicate that an increase in HRQoL

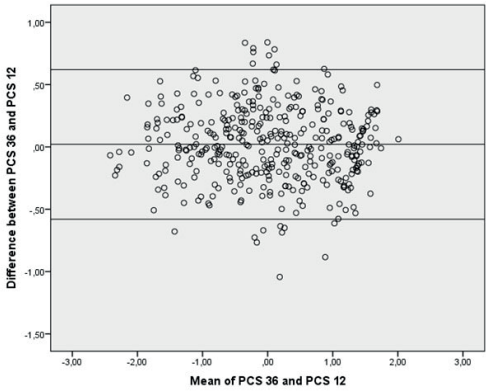
Figure 1. Bland–Altman plots showing mean difference and 95% limits of agreement between Z–scores of the SF–36, SF–12, and EQ–5D

a. MCS-36 and MCS-12 (n=381)



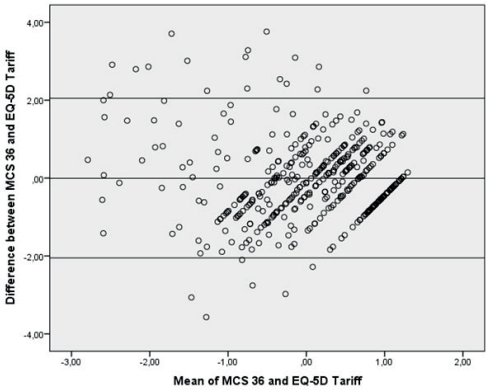
Horizontal lines show the population average of all individual differences (mean: 0.00) and the 95% limits of agreement (mean difference ±1.96 standard deviation of the differences): -0.60, +0.60

b. PCS-36 and PCS-12 (n=381)



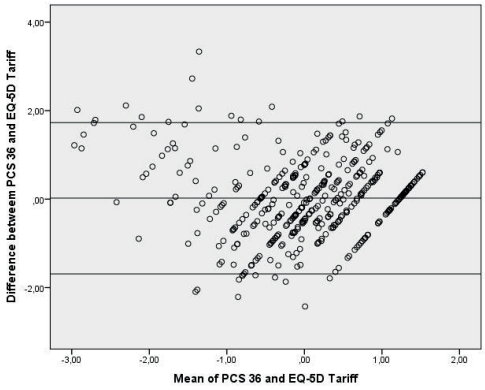
Horizontal lines show the population average of all individual differences (mean: 0.02) and the 95% limits of agreement (mean difference ±1.96 standard deviation of the differences): -0.58, 0.62

c. MCS-36 and EQ-5D score (n=413)



Horizontal lines show the population average of all individual differences (mean: 0.00) and the 95% limits of agreement (mean difference ±1.96 standard deviation of the differences): -2.05, 2.05

d. PCS-36 and EQ-5D score (n=413)



Horizontal lines show the population average of all individual differences (mean: 0.02) and the 95% limits of agreement (mean difference ±1.96 standard deviation of the differences): - 1.69, 1.73

is correlated with a decrease in respectively symptoms and the health change question (in which a low score is the most positive). The positive correlations indicate that an increase in HRQoL is correlated with an increase in the VAS score. The correlations for the SF-36 and SF-12 PCS and MCS were virtually the same. At baseline, the EQ-5D showed a correlation comparable to the PCS-36 correlation with symptom frequency, had a correlation with the VAS in between the PCS-36 and MCS-36 correlations, and had a somewhat lower correlation with the health change question compared to both the PCS-36 and MCS-36 correlations.

Table 3. Pearson and Spearman correlation coefficients between a. symptom frequency, b. EuroQol Visual Analogue Scale, c. the health change question at baseline, after 6 and 12 months with the time corresponding SF–36 component scores, SF–12 component scores, and EQ–5D score

	Baseline	6 months follow-up	12 months follow-up
a. Symptom frequency			
SF-36 MCS	-0.29	-0.32	-0.38
SF-36 PCS	-0.38	-0.33	-0.42
SF-12 MCS	-0.30	-0.33	-0.38
SF-12 PCS	-0.37	-0.32	-0.44
EQ-5D	-0.39	-0.41	-0.43
b. EuroQol Visual Analogue Scale			
SF-36 MCS	0.38	0.39	0.55
SF-36 PCS	0.60	0.60	0.65
SF-12 MCS	0.42	0.38	0.55
SF-12 PCS	0.60	0.62	0.68
EQ-5D	0.47	0.49	0.61
c. Health change question			
SF-36 MCS	-0.36	-0.32	-0.32
SF-36 PCS	-0.39	-0.37	-0.42
SF-12 MCS	-0.38	-0.36	-0.31
SF-12 PCS	-0.42	-0.42	-0.40
EQ-5D	-0.31	-0.32	-0.44

SF; Short Form, EQ-5D; EuroQol 5D, MCS; mental component summary, PCS; physical component summary

The calculated correlations for change in components and symptom frequency over time were very low for all components. The correlation of change in EQ-5D with change in symptom frequency was very similar to that of change in PCS-36. Correlations with change in VAS between 0 and 6 months were similar for change in SF-36 and SF-12 component scores, the change in EQ-5D correlation was similar to that of the change in PCS-36. Correlations with change in the health change question were very similar for change in SF-36 and SF-12 components, the change in EQ-5D had a correlation similar to that of the change in MCS-36.

Median follow-up time was 51 (38-63) months. Table 4a shows the association between the different questionnaire components with start of dialysis in a Cox regression analysis. An increase in MCS, PCS, and in EQ-5D score were all associated with a lower risk to start dialysis. This decrease was significant for the MCS in both the SF-36 and the SF-12. The SF-36 and SF-12 showed the same associations with start of dialysis. The EQ-5D showed the same association as the PCS-36. All 95% CIs overlapped. All five questionnaire components were associated with a lower mortality when increasing (table 4b). The MCS-12 and EQ-5D scores had the same association as compared with the MCS-36. The PCS-12 had similar associations as compared with the PCS-36.

Repeating all analyses without multiple imputation did not essentially change the results.

Table 4a. Crude and adjusted hazard ratio (95%CI) for start of dialysis per 1 SD increase in SF–36 component scores, SF–12 component scores, and EQ–5D scores

	HR (95% CI) Crude	HR (95% CI) Adjusted*
SF-36 MCS	0.83 (0.73-0.93)	0.83 (0.73-0.94)
SF-36 PCS	0.94 (0.84-1.05)	0.92 (0.82-1.04)
SF-12 MCS	0.83 (0.74-0.94)	0.83 (0.74-0.94)
SF-12 PCS	0.94 (0.84-1.06)	0.92 (0.82-1.04)
EQ-5D	0.93 (0.83-1.04)	0.92 (0.82-1.04)

SF; Short Form, EQ-5D; EuroQol 5D, SD; standard deviation, MCS; mental component summary, PCS; physical component summary, HR; hazard ratio, CI; confidence interval
*Adjusted for: Age, sex, race, primary kidney disease

Table 4b. Crude and adjusted hazard ratio (95% CI) for mortality per 1 SD increase in SF–36 component scores, SF–12 component scores, and EQ–5D scores

	HR (95% CI) Crude	HR (95% CI) Adjusted*
SF-36 MCS	0.88 (0.63-1.21)	0.82 (0.60-1.12)
SF-36 PCS	0.61 (0.45-0.81)	0.68 (0.51-0.92)
SF-12 MCS	0.83 (0.61-1.14)	0.81 (0.60-1.10)
SF-12 PCS	0.60 (0.44-0.81)	0.67 (0.49-0.91)
EQ-5D	0.84 (0.64-1.11)	0.84 (0.63-1.13)

SF; Short Form, EQ-5D; EuroQol 5D, SD; standard deviation, MCS; mental component summary, PCS; physical component summary, HR; hazard ratio, CI; confidence interval *Adjusted for: Age, sex, race, primary kidney disease

Discussion

In this study we found a better agreement between the SF-12 and the SF-36 as compared with the agreement between the EQ-5D and the SF-36, both for measurements at a single point in time and for changes over time. This was most pronounced in direct comparisons with the SF-36. In external validations the EQ-5D largely corresponded with the PCS-36.

In more detail, the Pearson correlation coefficient was high between the PCS-12 and PCS-36, as well as between the MCS-12 and MCS-36. The correlation was much lower between the EQ-5D and MCS-36. Between the EQ-5D and the PCS-36 the correlation was moderate. Over time correlations remained the same, except for the correlation between the EQ-5D and MCS-36 which improved slightly. The correlations for change over time were lower for most comparisons between the scores. In contrast, the correlations between the SF-12 and SF-36 remained very high. The Bland-Altman plots showed the largest 95% limits of agreement when comparing the EQ-5D with the SF-36 components and thus the lowest agreement.

Validating the questionnaires externally with symptom frequency, the VAS, and the health change question resulted in similar correlations for the SF-12 and SF-36, both for

a single point in time and for change over time. The EQ-5D had correlations comparable to the SF-36 PCS score. The correlation between the EQ-5D and VAS was in between the correlations for VAS with PCS-36 and VAS with MCS-36, and the correlation of EQ-5D with the health change question was somewhat lower compared to the health change question with PCS-36 and MCS-36 correlations. Calculating associations with start of dialysis resulted in similar HRs for the SF-12 and the SF-36, the EQ-5D associations were comparable to the SF-36 PCS association. The associations with death were similar for the SF-12 and the SF-36. The EQ-5D association with death was comparable to that of the MCS-36.

The type of questions in the EQ-5D could explain why most comparisons show a similarity to the PCS-36 score, and to a lesser extent to the MCS-36 score. The five questions that result in the tariff score mainly reflect physical limitations, with just one question on anxiety and depression. Most likely, as a consequence this results in more similarities with the physical part of the SF-36 as compared with the mental part. Given this explanation, the similarity with MCS-36 for the association with death seems counterintuitive. However, the associations with death are simply lower for both the EQ-5D and the MCS-36 as compared with the PCS-36 score. This can be explained by the type of questions in the MCS-36, since it is likely that physical suffering has a higher association with death as compared with mental problems.

As far as we know this is the first study validating the SF-12 and EQ-5D in pre-dialysis patients. In dialysis patients the SF-12 has been validated multiple times.[18, 19, 36] Both cross-sectional validation and validation over time showed a good agreement between the SF-12 and the SF-36 in these studies which is in concordance with our study. The EQ-5D has not been completely validated in dialysis patients, but change over time in EQ-5D scores has been compared with change over time in SF-36 and SF-12 scores by Loosman ***et al***.[19] The correlations between the change in SF-scores and change in EQ-5D scores were moderate and equal for both the SF-36 and SF-12, which is similar to the results in our study. In kidney transplant patients the EQ-5D has been validated.[13] When testing concurrent validity, Cleemput ***et al***. found a larger agreement between the EQ-5D and the PCS as compared with the agreement between the EQ-5D and the MCS, this is similar in our study.[13]

Our study had several limitations. First, we decided to use the SF-36 as our reference. Although this is not formally established as being the gold standard to measure QoL, QoL cannot be measured directly and the use of a questionnaire is necessary. When searching PubMed for QoL questionnaires the SF-36 is by far the most often used questionnaire. Choosing the SF-36 as a reference results in two component scores as an outcome, namely the MCS and PCS. When comparing the SF-12 with the SF-36, both questionnaires have the same component scores which makes comparisons easy. In contrast, the EQ-5D generates one single score which makes the comparison with the SF-36 more cumbersome. The total score of the EQ-5D consists of 4 questions regarding physical problems and 1 question regarding mental problems, making the higher correlations with the PCS understandable.

Finally, some of the baseline characteristics between patients with and patients without a QoL questionnaire differed. It is not likely this influenced the correlations between the questionnaires, but this affects generalizability.

Main strength is that this is the first study validating the SF-12 and EQ-5D in pre-dialysis patients. This an important step in achieving large scale monitoring in this patient group with the use of validated questionnaires with limited time consumption for the patient. Another strength is the extensive validation of the questionnaires, not only comparing to our golden standard the SF-36, but also with different external constructs, such as symptom frequency, the VAS from the EuroQol, and start of dialysis. Finally, we had a large sample of pre-dialysis patients who were followed over time, providing us with the opportunity to expand the validation with measurements over time.

In conclusion, the SF-12 had good agreement with the SF-36 and can be used as a substitute for the SF-36 when this questionnaire is considered too long. The EQ-5D in general had lower agreement with the SF-36 as compared with the SF-12 and should not be first choice to substitute the SF-36 when measuring quality of life in pre-dialysis patients.

References

1. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. ***N Engl J Med*** 2004; 351: 1296-1305
2. Roth M RP, Mindell J. Kidney disease and renal function. Health Center Survey for England. London, UK: NHS Information Centre; 2010.
3. Coresh J, Selvin E, Stevens LA***, et al.*** Prevalence of chronic kidney disease in the United States. ***JAMA*** 2007; 298: 2038- 2047
4. Thong MS, van Dijk S, Noordzij M***, et al.*** Symptom clusters in incident dialysis patients: associations with clinical variables and quality of life. ***Nephrol Dial Transplant*** 2009; 24: 225-230
5. Pagels AA, Soderkvist BK, Medin C, Hylander B, Heiwe S. Health-related quality of life in different stages of chronic kidney disease and at initiation of dialysis treatment. ***Health Qual Life Outcomes*** 2012; 10: 71
6. Tsai YC, Hung CC, Hwang SJ***, et al.*** Quality of life predicts risks of end-stage renal disease and mortality in patients with chronic kidney disease. ***Nephrol Dial Transplant*** 2010; 25: 1621-1626
7. de Goeij MC, Ocak G, Rotmans JI, Eijgenraam JW, Dekker FW, Halbesma N. Course of symptoms and health-related quality of life during specialized pre-dialysis care. ***PLoS One*** 2014; 9: e93069
8. Lowrie EG, Curtin RB, LePain N, Schatell D. Medical outcomes study short form-36: a consistent and powerful predictor of morbidity and mortality in dialysis patients. ***Am J Kidney Dis*** 2003; 41: 1286-1292
9. Merkus MP, Jager KJ, Dekker FW, Boeschoten EW, Stevens P, Krediet RT. Quality of life in patients on chronic dialysis: self-assessment 3 months after the start of treatment. The Necosad Study Group. ***Am J Kidney Dis*** 1997; 29: 584-592
10. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. ***Med Care*** 1992; 30: 473-483
11. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. ***Med Care*** 1996; 34: 220-233
12. Liem YS, Bosch JL, Hunink MG. Preference-based quality of life of patients on renal replacement therapy: a systematic review and meta-analysis. ***Value Health*** 2008; 11: 733-741
13. Cleemput I, Kesteloot K, Moons P***, et al.*** The construct and concurrent validity of the EQ-5D in a renal transplant population. ***Value Health*** 2004; 7: 499-509
14. Dale PL, Hutton J, Elgazzar H. Utility of health states in chronic kidney disease: a structured review of the literature. ***Curr Med Res Opin*** 2008; 24: 193-206
15. Herdman M, Gudex C, Lloyd A***, et al.*** Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). ***Qual Life Res*** 2011; 20: 1727-1736
16. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. ***Rheumatology (Oxford)*** 1999; 38: 807-813
17. Breckenridge K, Bekker HL, Gibbons E***, et al.*** How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. ***Nephrol Dial Transplant*** 2015; 30: 1605-1614

18. Lacson E, Jr., Xu J, Lin SF, Dean SG, Lazarus JM, Hakim RM. A comparison of SF-36 and SF-12 composite scores and subsequent hospitalization and mortality risks in long-term dialysis patients. *Clin J Am Soc Nephrol* 2010; 5: 252-260
19. Loosman WL, Hoekstra T, van Dijk S*, et al.* Short-Form 12 or Short-Form 36 to measure quality-of-life changes in dialysis patients? *Nephrol Dial Transplant* 2015; 30: 1170-1176
20. Nacak H, van Diepen M, de Goeij MC, Rotmans JI, Dekker FW, group P-s. Uric acid: association with rate of renal function decline and time until start of dialysis in incident pre-dialysis patients. *BMC Nephrol* 2014; 15: 91
21. Tattersall J, Martin-Malo A, Pedrini L*, et al.* EBPG guideline on dialysis strategies. *Nephrol Dial Transplant* 2007; 22: 5-21
22. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis* 2002; 39: S1-266
23. European Best Practice Guidelines Expert Group on Hemodialysis European Renal Association. Section I. Measurement of renal function, when to refer and when to start dialysis. *Nephrol Dial Transplant* 2002; 17 Suppl 7: 7-15
24. Dutch Federation of Nephrology. Multidisciplinary guidelines pre-dialysis. 2011
25. Ware JE SK, Kosinski M, Gandek B. SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: New England Medical Center, The Health Institute; 1993.
26. Ware JE SK, Kosinski M, Gandek B. SF-36 Physical and Mental Summary Scales: A User's Manual. Boston MA: New England Medical Center, The Health Institute; 1994.
27. Aaronson NK, Muller M, Cohen PD*, et al.* Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998; 51: 1055-1068
28. Ware JE, Jr., Gandek B, Kosinski M*, et al.* The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; 51: 1167-1170
29. Ware JE KM, Keller SD. SF-12: How to score the SF-12 Physical and Mental Health Summary Scales. Boston, MA: The Health Institute, New England Medical Center, 1995.
30. Gandek B, Ware JE, Aaronson NK*, et al.* Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; 51: 1171-1178
31. Brooks R. EuroQol: the current state of play. *Health Policy* 1996; 37: 53-72
32. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; 35: 1095-1108
33. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; 59: 1087-1091
34. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. *PLoS One* 2013; 8: e73990
35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310

36. Osthus TB, Preljevic VT, Sandvik L*, et al.* Mortality and health-related quality of life in prevalent dialysis patients: Comparison between 12-items and 36-items short-form health survey. ***Health Qual Life Outcomes*** 2012; 10: 46