# Mapping of the FKSI specific kidney disease measure onto three generic preference-based measures to generate utility values

**Manuel Monroy[1], Miguel A Ruiz[1]\*, Julio Bogueat[2], Isabel García[3] and Juan C Julián[2]**

[1]School of Psychology, Universidad Autónoma de Madrid, Spain

[2]Spanish Federation of Associations fighting Against Kidney Diseases – ALCER, Madrid, Spain

[3]Nephrology Department, University Hospital Josep Trueta, Girona, Spain

## Abstract

**Aim:** To obtain the mapping algorithms for translating the specific severity measures obtained by FKSI upon three generic utility instruments (SF-6D, EQ5D-3L and HUI-3) for chronic kidney disease (CKD) patients.

**Methods:** A cross-sectional prospective observational study design, including CKD patients recruited at random on demand for attention, was used. The final sample was composed by 161 patients suffering CKD. Women were 42%, mean age 54.6 years (SD=15.5) and average disease seniority of 2.82 years (SD=1.61). Three questionnaires measuring HRQoL (EQ5D-3L, SF-6D and HUI-3) were administered along with a CKD symptom severity specific (FKSI-9). Different regression models (linear, quadratic, cubic) were fitted for each separate generic instrument estimating preference-based utility values from FKSI symptom severity, and compared using several goodness-of-fit statistics. Empirical grouping of patients based on utilities and severity was explored using Latent Profile Analysis.

**Results:** Observed FKSI scores distributed between 0 and 29 points with an average M=7.5 (SD=6.39). Utility mean values for the three instruments were noticeably different: $M_{EQ}$=0.676 ($SD_{EQ}$=0.247), $M_{SF}$=0.514 ($SD_{SF}$=0.286) and $M_{HU}$=0.673 ($SD_{HU}$=0.232). In the three cases, the best fitting model was the cubic one, with the best fit attained by SF-6D ($R^2$=0.619) while EQ-5D ($R^2$=0.548) and HUI-3 ($R^2$=0.565) were lower. Latent profile analysis distinguished four clusters, with $R^2$=0.87 and 7% classification-error-rate.

**Conclusions:** Obtained results allow to transfer CKD deterioration values into social utility values for those health states, as measured by the three most widely used HRQoL instruments.

## Background

Chronic kidney disease (CKD) is a progressive and irreversible loss of renal function, due to different causes (diabetic nephropathy, hypertension, glomerulonephritis, hereditary renal failure, pharmacological poisoning, etc.) that lead to the kidneys losing their ability to eliminate waste, concentrate urine and preserve electrolytes in the blood, progressing towards the total loss of kidney function. At advanced stages, usual treatments are kidney transplantation, hemodialysis and peritoneal dialysis, all of which have a notable impact on daily life and the quality of life of patients.

The Spanish society of nephrology (SEN) [1] reported a CKD prevalence of 1,234 patients per million population (pmp) in 2016, varying between 1,752 pmp in the age group between 45 and 64 years and 2,888 pmp in people over 75 years. Etiology is known to vary with age, being vascular causes more frequent in patients older than 75 years, those caused by diabetes in the 65-74 age group, polycystic disease in the 45-65 age group and hereditary origin in ages less than 45 years. The prevalence of renal replacement therapy is 521 pmp in hemodialysis, 67 pmp in peritoneal dialysis and 647 pmp in renal transplant. The percentage of mortality in 2016 was 8%, being the most frequent causes cardiovascular problems and associated infections. Median survival is 6.3 years, with a 5 years' survival percentage of 57% [2].

Given the great impact of CKD on patient's wellbeing, it is important to bring forward indicators capable of quantifying the patient's vital state, towards an adequate therapeutic follow-up and, in particular, those reported by the patient. One of the patient reported outcome measures (PROM) most widely used as an indicator of the patient health status is Health Related Quality of Life (HRQoL), a measure that reflects the patient's subjective perception (without intervention of the clinical staff) in a repertoire of dimensions such as: emotional state, level of pain, physical functioning, social functioning and general perception of one's own health [3].

HRQoL is a particularly important output due to its diagnostic capabilities, since it has been shown to be directly related to mortality, hospitalization and consumption of clinical resources [4]. HRQoL has also shown relationship with other specific disease indicators, adding

complementary information for assessing clinical deterioration. Nowadays, HRQOL is accepted as a clinical goal by itself in patients with limited life expectancy or in therapies seeking to cope with the disease or to accommodate to symptoms, typical aspects of chronic diseases. PROMs have proven to be very sensitive when we study variations in health status for a particular pathology, being this sensitivity one of the reasons for usual inclusion in clinical studies [5-7].

Researchers commonly prefer to use pathology-specific questionnaires in patient follow-up, due to their greater sensitivity to health changes and better targeting to the pathology under study. But when the aim is to compare results with those of other pathologies or to perform economic evaluations, it is preferred to use generic (non disease-specific) HRQOL instruments. This is not without limitations, since generic instruments may capture information on patient characteristics (such as age, comorbidities or unwanted treatment effects) which may not be relevant or might be insensitive to mild health conditions.

The most popular generic instruments (such as SF6D, EQ-5D and HUI3) offer the possibility to calculate the utility value associated with each health condition (according to the profile given by the attributes measured by the instrument), which reflects the population preference towards each state of health, in a choice situation of uncertainty. This feature allows the use of utilities in the calculation of quality of life adjusted for years of life (QALY) and in any health-economy study in general.

In real life research, it is usually the case that we prefer to use a disease specific PROM instrument instead of a generic one, and not to include one of the later, so as not to overload the patient with self-reported measures. In such cases, the usual strategy is to perform a metric translation (mapping) from the specific measurement over the generic instrument. The mapping is also of interest when we want to compare our results with those obtained using a different generic instrument or even when there was no generic instrument available (as in the studies of retrospective databases or in meta-analyses).

## Objective

The objective of this study is to obtain the mapping algorithms necessary to translate the specific HRQoL measurement obtained using the FKSI specific CKD health index into three of the most popular preference-based generic instruments (SF6D, EQ-5D-3L and HUI-3). We will compare two procedures, on based on regression methods and another obtaining profiles by means of cluster analysis.

As a secondary benefit, we will be able to assess which one of the generic instruments is more adequate for capturing deterioration in HRQoL due to CKD condition.

## Methods

### Study design

The present study was designed as a cross-sectional prospective observational study. The sample was designed with the aim of reaching a large enough size to carry out the proposed multivariate analyses. Patients were included randomly on demand for treatment in the participating centers. A minimum target sample size of 150 patients with complete responses was determined. Patients were recruited by the collaborating therapists from the ALCER association, without limitations regarding the geographical origin, and including them as they gave their informed consent. The study protocol was approved by the Universidad Autómoma de Madrid (Spain) Research Ethic Committee. The Helsinki Declaration guidelines were met.

### Participants

The following inclusion criteria were applied: both genders, age above 18 years old, being in treatment of a chronic kidney disease, no cognitive impairment, being able to answer the questionnaires on their own, and having given their informed consent.

The final sample consisted of 161 patients, 41.6% of the participants being women. Mean age was 54.6 years (SD=15.5) and with an average time from diagnosis of 2.82 years (SD=1.61). A total of 18.6% were obese, 37.3% suffered from congenital pathology (14.9% Polycystic, 16.1% Glomerulonephritis, 2.5% Pyelonephritis) and 44.7% from acquired pathology (28.1% due to diabetes, hypertension or cardiovascular accident, 2.5% due to food or drug intoxication and 1.2% due to trauma). The most common treatments were: conservative treatment (13.7%), dialysis (62.7%) and transplantation (23.1%). Form them, 34.8% were above the cut-off score for clinical anxiety, while 23.6% could be classified with a clinical level of depression. Average anger expression score was 32.6 (SD=9.45). It is worth mentioning that 88.2% were also in treatment, at least, for another comorbidity (Table 1).

### Instruments

An ad-hoc data collection form was designed including the four questionnaires to be administered: three to measure generic HRQoL (EQ5D-3L, SF6D and HUI3) and a specific instrument measuring severity of CKD symptomatology (FKSI-9). In addition, Hamilton's Hospital Anxiety-Depression Scale (HADS) and the State-Trait Anger Expression Inventory (STAXI) were also administered. Three data collection forms were created, so that each one of the HRQoL questionnaires was presented first in turn, with the aim to control for any possible carryover effect among the quality of life measurements.

### Questionnaire EQ5D-3L

EuroQol-5D-3L (EQ5D-3L) [9,10] is a generic instrument of HRQOL based on population preferences. It assesses the level of deterioration in 5 attributes: mobility, self-care, daily activities, pain/discomfort and anxiety/depression; using items with 3 response levels (1=none, 2=some problems, 3=many problems). Each combination of levels creates a health profile, with a total of 243 possible health states, although not all are equally likely. The profile [11111] corresponds to perfect health and the profile [33333] represents the worst possible state of health (pits). Based on the sorting of health profiles according to social preferences, each health state is translated into a social utility value, which may be computed from the 5 attribute levels using multi-attribute utility function (MAUF). Different MAUFs are used in different countries, mainly using estimates based on standard gamble, time trade-off and visual analog scale (VAS) procedures. The basic MAUF equation is additive:

$$u_i = 1 - \left[ q + \sum_{j=1}^{j=5} \sum_{k=1}^{k=3} b_{jk} D_{ijk} + b_{N3} N3_i \right]$$

where the utility/preference value for health status $i$ is obtained by subtracting from 1 the disutility of the health status. Disutility is obtained by weighting by $b_{jk}$ the level of deterioration $k$ reached in dimension $j$ (dummy variable $D_{ijk}$) plus an interaction term ($N3$), which adds a constant when any of the dimensions reaches its maximum level of deterioration, plus a constant ($q$). It should be noted that the first

**Table 1.** Sociodemographic and clinical descriptors

| Variable | | Freq. | % | Variable | Level | Freq. | % |
|---|---|---|---|---|---|---|---|
| Place of treatment | Dialysis Unit | 30 | 18.6 | Etiology | Hereditary | 60 | 37.3 |
| | Hospital | 39 | 24.2 | | Acquired | 72 | 44.7 |
| | Home | 21 | 13.0 | | Unknown | 29 | 18.0 |
| | Patient Association | 65 | 40.4 | Number of concomitant diseases | 0 | 19 | 11.8 |
| | Unknown | 6 | 3.7 | | 1 | 45 | 28.0 |
| Age (years) | 18-29 | 8 | 5.0 | | 2 | 29 | 18.0 |
| | 30-39 | 25 | 15.5 | | 3 | 20 | 12.4 |
| | 40-49 | 27 | 16.8 | | 4 | 22 | 13.7 |
| | 50-59 | 38 | 23.6 | | 5 | 9 | 5.6 |
| | 60-69 | 33 | 20.5 | | 6 | 7 | 4.3 |
| | 70-79 | 24 | 14.9 | | 7 | 1 | 0.6 |
| | ≥ 80 | 4 | 2.5 | | 8 | 3 | 1.9 |
| | Unknown | 2 | 1.2 | | 9 | 4 | 2.5 |
| Gender | Female | 67 | 41.6 | | 10 | 2 | 1.2 |
| | Male | 93 | 57.8 | Treatment | Conservatory | 22 | 13.7 |
| Body Mass Index | Underweight | 5 | 3.1 | | Dialysis | 101 | 62.7 |
| | Normal | 56 | 34.8 | | Transplantation | 37 | 23.0 |
| | Overweight | 65 | 40.4 | | Not known | 3 | 0.6 |
| | Obesity | 30 | 18.6 | Anxiety | Subclinical | 85 | 52.8 |
| | Unknown | 5 | 3.1 | | Uncertain | 20 | 12.4 |
| Education | No studies | 12 | 7.5 | | Present | 56 | 34.8 |
| | Primary | 44 | 27.3 | Depression | Subclinical | 104 | 64.6 |
| | Secondary | 33 | 20.5 | | Uncertain | 19 | 11.8 |
| | Vocational Training | 37 | 23.0 | | Present | 38 | 23.6 |
| | Postgraduate | 29 | 18.0 | Disease seniority (years) | ≤5 | 49 | 30.4 |
| Civil State | Single | 34 | 21.1 | | 6-10 | 27 | 16.8 |
| | Married | 87 | 54.0 | | 11-15 | 20 | 12.4 |
| | Divorced | 22 | 13.7 | | 16-20 | 17 | 10.6 |
| | Widower | 13 | 8.1 | | > 20 | 40 | 24.8 |
| | Other | 2 | 1.9 | | Unknown | 8 | 5.0 |

level on any dimension ($k = 1$), represents that there is no deterioration in that dimension, $D_{ijk} = 0$, and the perfect health profile is anchored at a utility value of 1.

## SF6D questionnaire

The medical results survey (MOS), in its 6 dimension utility form (SF6D) [11], is a generic HRQoL instrument based on preferences derived from the MOS SF-36 (36 items). It summarizes the level of deterioration in 6 dimensions: physical functioning, role limitations, social functioning, pain, mental health and vitality; using a coding in 4 to 6 levels of 11 items. It is possible to obtain a total of 18,000 health profiles, with profile [111111] corresponding to perfect health, and [645655] representing the worst possible state of health. Different MAUFs have been estimated to derive utilities in different countries, with the particularity that no constant of severity (interaction) is used. A value of 0 is assigned to the first level for each dimension/attribute [12,13].

## HUI3 questionnaire

The Health Utilities Index Mark 3 (HUI-3) [14] survey covers several aspects of health, intentionally restricted to skills (physical and emotional), and excludes role performance and social interaction. It covers eight attributes: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain, using five to six response levels. Each combination of levels indicates a unique health status. The MAUF of this instrument is multiplicative and different functions have been

estimated for different countries; the state of perfect health [11111111] has a utility of 1, while the utility for the lowest level in the eight attributes [66566565] is -0.36, which is considered a health situation equivalent to worse than being dead. The MAUF for deriving the utility from a profile in the Spanish population, is given by:

$$u_i = (1{,}0078 \times b1 \times b2 \times b3 \times b4 \times b5 \times b6 \times b7 \times b8) - 0.0078$$

Where the values $b1$-$b8$, in the pits-Full Health metric, are the coefficients calculated in the Spanish population [15], which correspond to the response level attained in each one of the dimensions.

## FKSI-DRS questionnaire

The Functional Assessment of Cancer Therapy - Kidney Symptom Index - Disease Related Symptoms (FKSI-DRS-9) [16-18] is a self-reported questionnaire composed by the first 9 items of the instrument FKSI-15, which are answered on a 0 to 4 points Likert scale assessing the level of limitation due to the symptoms of kidney diseases. This instrument was used as the disease-specific measure of deterioration. Two dimensions may be distinguished: physical and psychological. Items used to assess perceive individual change were discarded from the FKSI-15. The scoring ranges from 0 to 36 points. A higher score reflects greater deterioration [19].

## Statistical analyses

The criterion score on the specific kidney disease health status was obtained computing the factor score for FKSI items, assuming

one overall dimension (Principal Components extraction, factor score regression method), which produces a summary score with 0 mean and standard deviation proportional to the eigenvalue of the dimension.

To interpretation easier, it was re-scaled to a 0-1 metric, since the attainable minimum and maximum scores are known. The score obtained was considered the specific kidney symptomatology indicator of reported severity.

Once the specific severity indicator was obtained, a metric translation of the indicator values was performed on each of the three generic measures of HRQoL used in this study, each one separately. In this way, the predicted utility values were obtained for each generic instrument given a level of kidney symptoms severity. Several regression models, linear and non-linear, were tested and compared using various goodness-of-fit statistics.

In all regression models, the values of disutility ($d_i = 1 - u_i$) were used, instead of the values of utility, for the following reasons. First, the data mass is usually concentrated around the most favorable health states with least disutilities, so that the points of greatest mass are close to the origin of the coordinate axes, the independent and dependent variables (severity and disutility) are measured in the same direction and the slope of the model is always positive. Secondly, it is always possible to estimate a model without the intersection term, anchoring the 0 value of disutility (perfect health) at the origin, and making it match with the minimum severity value of the FKSI (which will also be 0). Subsequently, it suffices to subtract from 1 the predicted disutility to obtain the model utility predicted value.

The following regression models were estimated: linear, quadratic and cubic, using the density function values; and Tobit, using cumulative values of the distribution function. To anchor the best possible health states in both instruments, symptom severity scores were scaled within the 0-1 range.

Before estimation of the different prediction models, those patients with evident outlier values in two or more of the generic instruments were discarded, since their score could be reflecting peculiarities that were not typical of the pathology under study. Outlier values were identified as those clearly falling outside the 95% individual confidence interval for the linear model (departing in more than 3 standardized residuals, Figure 1).

Along with the statistical significance for the regression coefficient estimates, goodness of fit (GOF) of each model was assessed using $R^2$ statistic, average absolute error (MAE) and percent average absolute error (MAPE). MAE and MAPE were computed overall and by quintile

groups according to the severity scores, in order to assess the local GOF at the different levels of severity. MAE and MAPE indices should be studied with caution since very small utility values can inflate the mismatch values substantially, when dividing by quantities close to 0.

Covariates were not included in the regression models (age, disease seniority, number of treatments, comorbidities, depression level, etc.) with the aim to consider only the direct effect of the disease. In addition, the inclusion of covariates would limit the use of the models in retrospective studies in which the possible covariates could have not been gathered.

As an additional procedure, a latent profile analysis (LCP) was carried out exploring how health states summarized by the three generic instruments rank patients. It could be the case that patients are sorted differently by each generic instrument or that utility measures might show different sensitivity at different levels of severity. The disutilities of the generic instruments (HUI-III, SF-6D and EQ-5D) as well as the severity of the specific instrument (FKSI) were included as active variables in the LCP. Sociodemographic variables and disease descriptors were also included as inactive covariables in order to describe the profiles obtained.

All analyzes were carried out using IBM SPSS v23 software and LatentGold V.5.0.

## Results

Observed direct scores on the FKSI renal symptoms severity scale were distributed between 0 and 29 points, with a mean M=7.5 (SD=6.39), while transformed factor scores varied between 0 and 1, with an average value of 0.261 (SD=0.219), and with positive skewness $g_1$=0.954 (SE=0.191). Only 6.2% of patients scored at the minimum scale value.

Average utility scores obtained with the generic instruments were significantly different: $M_{EQ}$=0.676 ($SD_{EQ}$=0.247), $M_{SF}$=0.514 ($SD_{SF}$=0.286) and $M_{HU}$=0.663 ($SD_{HU}$=0.232), being SF6D mean significantly lower than the other two (p<0.001). Correlations between them were all significant (p<0.001), $r$ (EQ,SF)=0.797, $r$ (EQ,HU)=0.764 and $r$ (SF,HU)=0.763. Scores presented a clear negative bias in all cases, with accumulation of cases at the top of the scale: $g_{EQ}$=0.682 (SE=0.191), $g_{SF}$=0.904 (SE=0.195), $g_{HU}$=0.944 (SE=0.194).

Figure 2 shows the cumulative distribution of re-scaled factor scores. The equation needed to compute the corrected factor scores out from the scores on the individual items of the FKSI is given by
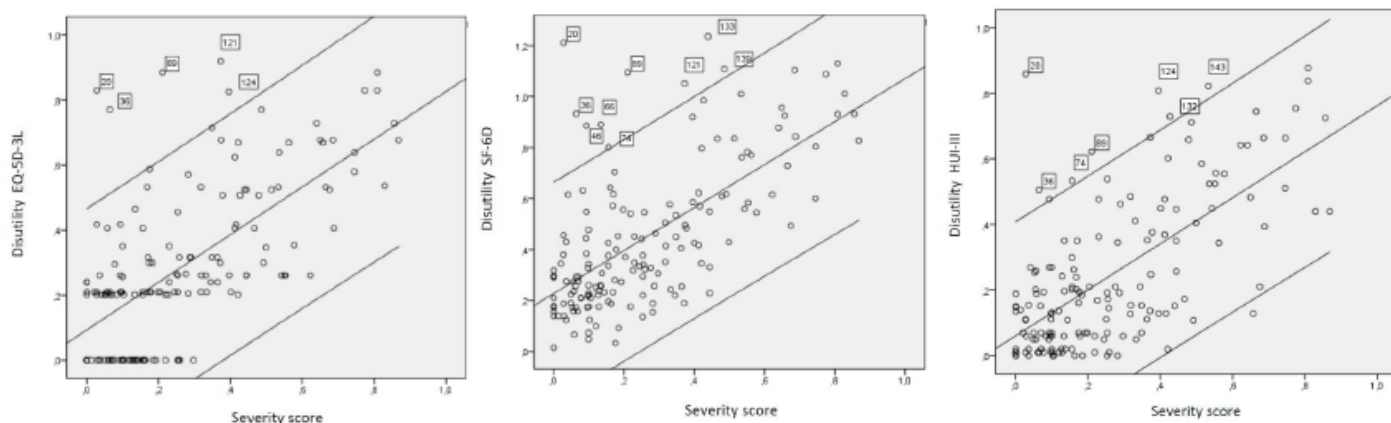


**Figure 1.** Outlier identification. EQ-5D, SF-6D and HUI-3 vs FKSI scatter-plots showing the fitted linear model with 95% individual confidence interval

$$\hat{f}_i = 0.226 * \begin{bmatrix} (0.155*(X_1-1.593))+(0.121*(X_2-1.185))+(0.029*(X_3-0.673))+ \\ (0.128*(X_4-1.216))+(0.282*(X_5-1.525))+(0.139*(X_6-0.772))+ \\ (0.090*(X_7-0.512))+(0.089*(X_8-0.148))+(0.071*(X_9-0.062))+1.17 \end{bmatrix}$$

Where $X_1$-$X_9$ are the scores on the FKSI items, and the values 0.266 and 1.17 are scale constants needed to translate the values into the 0-1 range.

Regarding the degree of sensitivity shown by the instruments, it was observed that the EQ-5D was the least sensitive, obtaining only 36 profiles of the possible ones and accumulating 55.9% of the patients in 4 of them (11111, 11112, 11121, 11122), while the 89 profiles were obtained using the HUI-3 and 146 using the SF-6D (Table 2).

Table 3 shows the percentage of patient accumulated at the different response levels of the attributes and for each one of the instruments. It can be observed that the patients tend to be located at the less severe health levels, although patients can be found in the higher levels of severity of most attributes.

The cubic model was the best fitting one for the mapping functions three, although it should be noted that the differences in fit were minimal
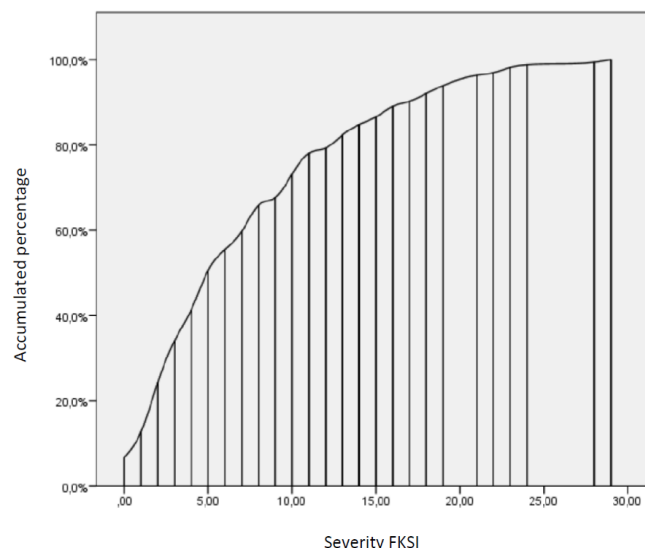


**Figure 2.** Cumulative distribution for FKSI observed values

**Table 2.** Generic HRQoL prevalent profiles, frequency, percentage and cumulative percentage. Partial listing

| | | | | | Instrument | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **EQ5D** | | | | **SF6D** | | | | **HUI** | | | |
| Profile | n | % | Cum % | Profile | n | % | Cum % | Profile | n | % | Cum % |
| 11111 | 44 | 27.3 | 27.3 | 211224 | 3 | 1.9 | 1.9 | 21111212 | 10 | 6.2 | 6.2 |
| 11121 | 20 | 12.4 | 39.8 | 211123 | 2 | 1.2 | 3.1 | 11111111 | 9 | 5.6 | 11.8 |
| 11112 | 14 | 8.7 | 48.4 | 212123 | 2 | 1.2 | 4.3 | 21111111 | 9 | 5.6 | 17.4 |
| 11122 | 12 | 7.5 | 55.9 | 212125 | 2 | 1.2 | 5.6 | 21111112 | 8 | 5.0 | 22.4 |
| 11222 | 6 | 3.7 | 59.6 | 212224 | 2 | 1.2 | 6.8 | 21111211 | 8 | 5.0 | 27.3 |
| 21111 | 5 | 3.1 | 62.7 | 212225 | 2 | 1.2 | 8.1 | 11111211 | 6 | 3.7 | 31.1 |
| 21222 | 5 | 3.1 | 65.8 | 212324 | 2 | 1.2 | 9.3 | 11111112 | 5 | 3.1 | 34.2 |
| 11123 | 4 | 2.5 | 68.3 | 213323 | 2 | 1.2 | 10.6 | 21111232 | 5 | 3.1 | 37.3 |
| 11132 | 4 | 2.5 | 70.8 | 223114 | 2 | 1.2 | 11.8 | 21111213 | 4 | 2.5 | 39.8 |
| 21121 | 4 | 2.5 | 73.3 | 233224 | 2 | 1.2 | 13.0 | 11111113 | 3 | 1.9 | 41.6 |
| 21232 | 4 | 2.5 | 75.8 | 312325 | 2 | 1.2 | 14.3 | 21111311 | 3 | 1.9 | 43.5 |
| 12222 | 3 | 1.9 | 77.6 | 343445 | 2 | 1.2 | 15.5 | 21111313 | 3 | 1.9 | 45.3 |
| 21223 | 3 | 1.9 | 79.5 | 344556 | 2 | 1.2 | 16.8 | 11111122 | 2 | 1.2 | 46.6 |
| 22222 | 3 | 1.9 | 81.4 | 443434 | 2 | 1.2 | 18.0 | 21111132 | 2 | 1.2 | 47.8 |
| 22233 | 3 | 1.9 | 83.2 | 111121 | 1 | 0.6 | 18.6 | 21111221 | 2 | 1.2 | 49.1 |
| 11221 | 2 | 1.2 | 84.5 | 111122 | 1 | 0.6 | 19.3 | 21111222 | 2 | 1.2 | 50.3 |
| 11233 | 2 | 1.2 | 85.7 | 111123 | 1 | 0.6 | 19.9 | 21111333 | 2 | 1.2 | 51.6 |
| 21122 | 2 | 1.2 | 87.0 | 111124 | 1 | 0.6 | 20.5 | 21111334 | 2 | 1.2 | 52.8 |
| 21233 | 2 | 1.2 | 88.2 | 111133 | 1 | 0.6 | 21.1 | 21111412 | 2 | 1.2 | 54.0 |
| 22223 | 2 | 1.2 | 89.4 | 112223 | 1 | 0.6 | 21.7 | 21111434 | 2 | 1.2 | 55.3 |
| 22333 | 2 | 1.2 | 90.7 | 112323 | 1 | 0.6 | 22.4 | 21112323 | 2 | 1.2 | 56.5 |
| 11113 | 1 | 0.6 | 91.3 | 112324 | 1 | 0.6 | 23.0 | 23111212 | 2 | 1.2 | 57.8 |
| 11211 | 1 | 0.6 | 91.9 | 112326 | 1 | 0.6 | 23.6 | 41111111 | 2 | 1.2 | 59.0 |
| 11212 | 1 | 0.6 | 92.5 | 121222 | 1 | 0.6 | 24.2 | 11111132 | 1 | 0.6 | 59.6 |
| 11223 | 1 | 0.6 | 93.2 | 131133 | 1 | 0.6 | 24.8 | 11111222 | 1 | 0.6 | 60.2 |
| 11231 | 1 | 0.6 | 93.8 | 133334 | 1 | 0.6 | 25.5 | 11111232 | 1 | 0.6 | 60.9 |
| 11232 | 1 | 0.6 | 94.4 | 211111 | 1 | 0.6 | 26.1 | 11111311 | 1 | 0.6 | 61.5 |
| 12312 | 1 | 0.6 | 95.0 | 211113 | 1 | 0.6 | 26.7 | 11111312 | 1 | 0.6 | 62.1 |
| 21131 | 1 | 0.6 | 95.7 | 211122 | 1 | 0.6 | 27.3 | 11111313 | 1 | 0.6 | 62.7 |
| 21211 | 1 | 0.6 | 96.3 | 211211 | 1 | 0.6 | 28.0 | 11111324 | 1 | 0.6 | 63.4 |
| 21212 | 1 | 0.6 | 96.9 | 211214 | 1 | 0.6 | 28.6 | 11111413 | 1 | 0.6 | 64.0 |
| 21221 | 1 | 0.6 | 97.5 | 211221 | 1 | 0.6 | 29.2 | 11112222 | 1 | 0.6 | 64.6 |
| 22221 | 1 | 0.6 | 98.1 | 211222 | 1 | 0.6 | 29.8 | 11121112 | 1 | 0.6 | 65.2 |
| 22312 | 1 | 0.6 | 98.8 | 211223 | 1 | 0.6 | 30.4 | 11121213 | 1 | 0.6 | 65.8 |
| 22323 | 1 | 0.6 | 99.4 | 211322 | 1 | 0.6 | 31.1 | 11121333 | 1 | 0.6 | 66.5 |
| 32233 | 1 | 0.6 | **100.0** | 211323 | 1 | 0.6 | **31.7** | 11131122 | 1 | 0.6 | **67.1** |

between the models of different shape (linear, quadratic and cubic). The cubic pattern was chosen due to better represent the expected evolution of the utilities, starting at a floor value corresponding to the perfect health state (disutility = 0) and growing towards an asymptotic value at the ceiling of the scale (disutility=1) (Figure 3). Table 4 shows the coefficients needed to estimate the disutilities for the three instruments. Predicted utilities are obtained by subtracting from 1 the value of predicted disutility.

Moderate fit was attained by all models, with the SF-6D reaching the best fit ($R^2$=0.619), while EQ-5D ($R^2$=0.548) and HUI-3 ($R^2$=0.565) were lower. However, the relative error obtained with the SF-6D model was much higher (MAPE=56.9%) than the 20% obtained by the two other models. As expected, the size of residuals stratified by quintiles turned out to be especially bad at the quintile corresponding to high utility values, that is, in less serious health conditions.

Although determining the number of clusters for this validation test is not crucial, the LCP analysis identified 4 clusters with centroids shown in Table 5. The solution reached good fit $R^2$=0.87 with an error classification rate of 7%. Cluster profiles (Figure 4) show that averages are arranged in parallel (without crossings) implying that clusters are collecting groups of patients with levels of progressive deterioration in the disease (FKSI) and also in the three generic instruments of HRQOL. In the absence of crosses, we can infer that there are no other aspects of health not being considered, which might be influencing substantially the measurement of HRQoL, but those corresponding to the CKD itself. It is also true that if we would increase enough the number of conglomerates, profiles would end up showing crossings between clusters. Inspection of profiles also shows that the SF-6D tends to assign slightly higher disutility values, and the EQ-5D usually assigns lower disutility values. Progression of disutility when moving between

**Table 3A.** Generic HRQoL percentage of responses by attribute/dimension response level

| | EQ5D-3L Levels | | |
|---|---|---|---|
| Dimension | 1 | 2 | 3 |
| Mobility | 73.3 | 26.1 | 0.6 |
| Personal Care | 88.8 | 11.2 | 0 |
| Daily Activities | 68.9 | 28.0 | 3.1 |
| Pain | 43.5 | 43.5 | 13.0 |
| Anxiety/Depression | 50.3 | 36.0 | 13.7 |

**Table 3B.** Generic HRQoL percentage of responses by attribute/dimension response level

| | SF-6D Levels | | | | | |
|---|---|---|---|---|---|---|
| Dimension | 1 | 2 | 3 | 4 | 5 | 6 |
| Physical Function | 7.5 | 44.7 | 29.8 | 6.8 | 5.6 | 5.6 |
| Role Limitations | 42.9 | 15.5 | 8.7 | 32.9 | * | * |
| Social Function | 23.6 | 25.5 | 31.7 | 11.8 | 7.5 | * |
| Pain | 21.1 | 18.6 | 24.2 | 13.0 | 14.9 | 8.1 |
| Mental Health | 6.8 | 52.8 | 16.8 | 12.4 | 11.2 | |
| Vitality | 5.0 | 9.9 | 16.1 | 24.8 | 21.7 | 22.4 |

**Table 3C.** Generic HRQoL percentage of responses by attribute/dimension response level

| | HUI-3 Level | | | | | |
|---|---|---|---|---|---|---|
| Dimension | 1 | 2 | 3 | 4 | 5 | 6 |
| Vision | 26.7 | 63.4 | 5.0 | 3.1 | 1.9 | 0 |
| Hearing | 92.5 | 1.9 | 3.7 | 1.2 | 0.6 | 0 |
| Speech | 95.0 | 3.7 | 0.6 | 0 | 0.6 | * |
| Mobility | 82.6 | 8.7 | 4.3 | 3.1 | 0.6 | 0.6 |
| Dexterity | 36.6 | 33.5 | 17.4 | 9.9 | 2.5 | 0 |
| Cognition | 63.4 | 12.4 | 20.5 | 3.1 | 0.6 | * |
| Pain | 31.7 | 37.9 | 19.3 | 9.9 | 1.9 | * |

**Table 3D.** Generic HRQoL percentage of responses by attribute/dimension response level

| | FKSI Level | | | | |
|---|---|---|---|---|---|
| Dimension | 0 | 1 | 2 | 3 | 4 |
| Lack of Energy | 25.5 | 28.6 | 19.9 | 14.9 | 11.2 |
| General Pain | 42.2 | 21.1 | 21.1 | 8.7 | 6.8 |
| Weight Loss | 64.6 | 13.0 | 14.9 | 7.5 | 0 |
| Bone Pain | 43.5 | 23.0 | 13.7 | 9.9 | 9.9 |
| Exhaustion | 29.8 | 23.0 | 24.2 | 12.4 | 10.6 |
| Breathing | 64.0 | 11.8 | 13.7 | 6.2 | 1.2 |
| Cough | 72.0 | 15.5 | 5.0 | 6.2 | 1.2 |
| Fever | 91.9 | 5.6 | 0.6 | 1.9 | 0 |
| Hematuria | 98.1 | 1.2 | 0 | 0 | 0.6 |

*Dimension level not used

**Table 4.** Estimated model coefficients, and goodness of fit statistics, overall (top) and by quintiles (bottom)

| Coefficients | Instrument | | |
|---|---|---|---|
| | EQ-5D | SF-6D | HUI-3 |
| $b_0$ | 0.103 | 0.248 | 0.092 |
| $b_1$ | 0.234 | -0.061 | -0.082 |
| $b_2$ | 1.462 | 2.460 | 2.107 |
| $b_3$ | -1.053 | -1.648 | -1.455 |
| Fit | | | |
| $R^2$ | 0.548 | 0.619 | 0.565 |
| $F_{3,157}$ | 61.384* | 80.228* | 64.441* |
| MAE | 0.128 | 0.121 | 0.108 |
| MAPE | 22.4% | 56.9% | 17.6% |

* $p<0.001$; MAE=Mean Absolute Error; MAPE: Mean absolute percentage Error

| | MAE (MAPE) | | | | | |
|---|---|---|---|---|---|---|
| | Overall | Q1 | Q2 | Q3 | Q4 | Q5 |
| EQ-5D-3L | 0.128 (22.4) | 0.118 (14.6) | 0.139 (16.6) | 0.122 (17.1) | 0.133 (27.2) | 0.132 (36.9) |
| SF-6D | 0.121 (56.9) | 0.093 (24.8) | 0.101 (40.0) | 0.134 (72.6) | 0.137 (59.9) | 0.153 (82.2) |
| HUI-3 | 0.108 (17.6) | 0.073 (8.6) | 0.074 (8.2) | 0.114 (15.2) | 0.142 (19.7) | 0.152 (38.0) |

MAE=Mean Absolute Error; MAPE:Mean absolute percentage Error; Q1-Q5=quintile groups

**Table 5.** Cluster centroids (means disutilities) and standard deviations by class latent profiles

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| FKSI Severity | 0.1413 (0.098) | 0.2754 (0.171) | 0.4419 (0.234) | 0.7611 (0.203) |
| EQ5D Disutility | 0.0563 (0.095) | 0.2541 (0.054) | 0.4227 (0.166) | 0.6809 (0.133) |
| SF6D Disutility | 0.2283 (0.094) | 0.3626 (0.182) | 0.5507 (0.221) | 0.9493 (0.136) |
| HUI3 Disutility | 0.0808 (0.084) | 0.1504 (0.101) | 0.3835 (0.192) | 0.5990 (0.209) |
| N (%) | 46 (32%) | 40 (28%) | 36 (25%) | 21 (15%) |

disability strata within each instrument is rather similar for all three instruments.

## Discussion

Disease specific HRQoL are the preferred choice for measuring health given their high sensitivity to changes in the patient health state (treatment effectiveness, disease progression, coping with symptoms, etc.). Therefore, using generic instruments instead implies loosing sensitivity and also involves other measurement problems since it is difficult to make the patient isolate the health aspects related only to the pathology that is being assessed. Naturally, patients have an overall view on their health state and it is difficult to filter out the effect of possible comorbidities, adverse events or the affective state. However, even if it is unadvisable to use generic instruments for an accurate assessment of the health state and, therefore, for patients follow-up, there are research situations where obtaining generic measures is
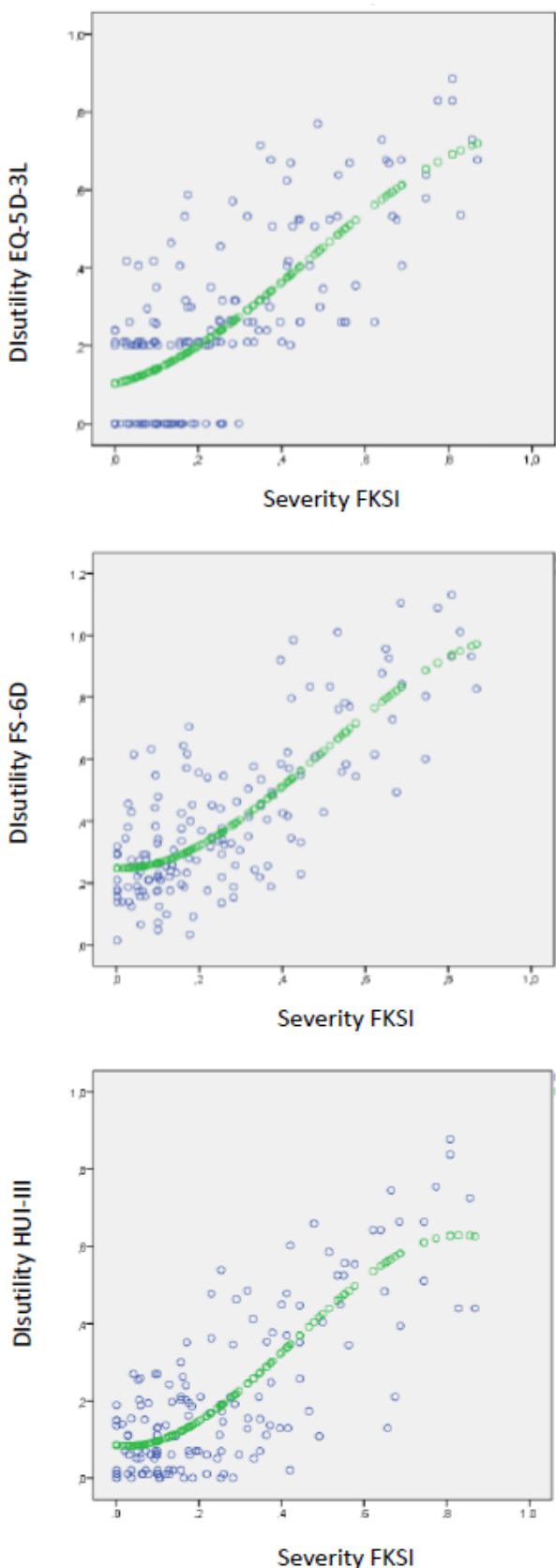
**Figure 3.** EQ-5D, SF-6D and HUI-3 observed (blue) and predicted (green) disutilities vs FKSI observed values
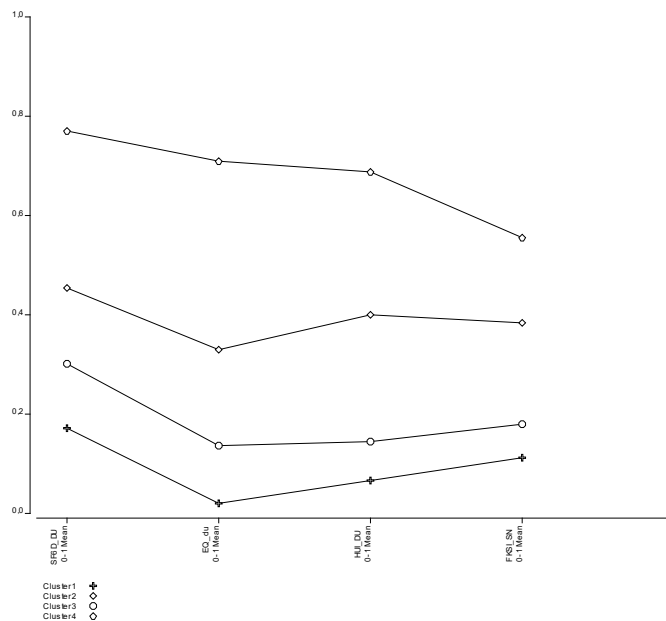


**Figure 4.** Cluster profiles showing average generic disutilities (EQ-5D, SF-6D, HUI-3) and symptom severity (FKSI)

crucial. We must remember that the generic measures reflect the social value of the patient health state (compared to other possible health states) and not really their vital situation. Which is the reason why they are the measures of choice in pharmaco-economic valuations.

A possible strategy to avoid these problems would be to design preference elicitation choice experiments using vignettes based on the health conditions derived from the specific instrument, but this would not prevent from the inflation of marginal utilities due to other serious comorbidities being present. Another possibility would be to determine the generic health profiles that are really prevalent and meaningful in the particular disease, and only to mapping those conditions. This approach could be used when observed distributions are found such as that obtained for EQ-5D-3L, where a small number of health states gather together the majority of patients. However, if we intend to obtain representative results, very large samples should be used, and it could be cumbersome when the number of possible health states is very large, as has happened empirically with the SF-6D (with 127 states) or the HUI- 3 (with 64 states, Table 3).

For the time being, the direct mapping of specific health states into generic utility values seems to be the most accepted option [20]. Nevertheless, another possible way to determine the mapping between generic instruments anchored by a specific instrument could be to identify empirical profiles of health states shared by groups of patients, using cluster generation procedures such as the LPA. This procedure would allow to determine as many clusters as considered appropriate, and to obtain the table of correspondences between utility values of different instruments based on the average utility value on each instrument, represented by the centroid of each cluster. We have seen that for a small number of clusters this option is possible (Table 5), but the behavior with a high number of strata (clusters) might not be as uniform as in our case, and inversions between the instruments (profile crossings) may appear, which could be difficult to understand. In fact, in our findings, the LPA solution may have been particularly insensitive to comorbidities due to the removal of extreme cases.

In our study, CKD has shown to be a quite disabling pathology, with low average utility values: $M_{EQ}$=0.676, $M_{SF}$=0.514 and $M_{HU}$=0.673. However, we have observed a large number of patients whose scores are at the lowest level (without deterioration or with mild deterioration) in most of the attributes of the generic instruments (Table 4). It is also true that our sample, even being representative of patients with CKD in touch with patient organizations, is not a sample with a high level of deterioration since 50% of the subjects obtain scores between 0 and 5 points (from a possible maximum of 36 points).

Utility scores obtained using SF-6D and HUI-3 instruments showed to be more sensitive to CKD severity than those obtained using the EQ-5D-3L. This behavior is known and currently a new version of the EQ-5D is being developed with five levels per attribute [21,22]. In addition, the distribution of scores of the first two instruments was more disperse and they did not show a gap between perfect health and the following health profile. The observed cumulative distribution functions for SF-6D and HUI-3 disutility scores were more uniform, while the EQ-5D-3L showed a steeper function, especially at the mild health states.

In the regression models, the strategy of using factor scores to summarize CKD severity is technically preferable to the use of the score obtained directly from the algebraic sum of FKSI item scores, since each item is weighted according to its individual reliability for optimally sorting patients according to CKD severity. Furthermore, it avoids having to decide on how to sum-up the scores when building the criterion variable (disutilities) based on the response levels in each item, and minimizes a possible impact of the covariates over particular levels of response.

Although we have not considered any covariates in the prediction of disutilities, we did check for the influence of other variables in the mapping functions. Variables able to contribute in explaining additional variability present in utility scores where "number of concomitant diseases", "anxiety" and "frequency of anger situations", and also "years since diagnosis" in the case of predicting SF-6D disutilities, results departing from the inclusion of obesity, age and hypertension in the EPIRCE study [2].

The model with best fit for predicting disutility values was the cubic model. All proposed models presented the same problem, the great dispersion of the utility scores observed at the non-severe health states of the FKSI (Figure 3). But this phenomenon should not be understood as an anomalous behavior, rather it reflects the limitation of specific instruments themselves to capture the effect of covariates (that may explain the overall level of deterioration), and not so much due to the limitation of generic instruments for measuring benign health states. In fact, a not irrelevant group of patients obtained very high disutility values (probably due to other aspects of their health deterioration) but with a very low specific CKD deterioration level. Studying these cases with large residuals and low FKSI scores, we found that they were subjects with notable high levels of anxiety and depression, among other possible confounding factors. Better fitting models could have been obtained including covariates not specific to CKD (such as age, psychological health, comorbidities, type of treatment, etc.), but this would lead to a limited applicability of models to other data sets and, subsequently, the mapping models would not be generalizable.

Our study on the behavior of utilities in subpopulations of cases produced the stratification of the sample by levels of severity. The clusters corresponded to strata of patients with progressive levels of deterioration, in which all instruments showed a similar progression, both generic and specific. Although the technique used is very sensitive to the presence of atypical cases, the solution obtained discriminated levels of deterioration but not the presence of this type of cases (perhaps due to the previous filtering of outliers).

## Conclusions

The mapping of disease-specific instruments into health related generic measures is a common methodological strategy which takes advantage of the high sensitivity of specific instruments and the broad generalizability of generic measures. It was shown that it is possible to map CKD specific FKSI scores into generic disutilities (SF-6D, HUI-3 and EQ-5D-3L), achieving adequate goodness of fit values and an acceptable amount explained variance (between 55 % and 62%).

The supremacy of the cubic model was not very evident, since the MAPE values of the different models were very similar. The similarity of the models is due to the lack of fit obtained by all of them at low values of disutility (best health states). This is an inherent problem for generic instruments, which have shown to capture health impairments not attributable to the specific deterioration measured by the FKSI.

Our results allow transferring the values of CKD impairment onto the utility attributed by society to those health states, as they are appraised by the three HRQoL instruments most frequently used in research.

## Limitations

The present study has been carried out in the Spanish population and it is possible that cultural biases might be present.

## Availability of data and materials

Raw data must be kept confidential due to the Ethics Committee requirements. Nevertheless, anonymized data may be shared on reasonable demand to the authors.

## Competing interests

The authors declare that they have no competing interests.

## Funding

## Author's contributions

MR, MM, JCJ & JB were responsible for the design of the study; MR, JB and JCJ decided on clinical variables and psychological measurements to be gathered from patients, MM & MR analyzed and interpreted patient data regarding the mapping estimation procedures. MM and JCJ performed the bibliographic search on epidemiological data, JB and IG were responsible for data collection; MM and MR were major contributors in writing the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

## References

1. Sociedad Española de Nefrología (2017). XLVII congress Spanish Nephrology Society abstracts [Resúmenes del XLVII congreso de la Sociedad Española de Nefrología]. Nefrología, 37(Supl. 1). http://www.registrorenal.es/download/documentacion/Informe_REER_2017.pdf

2. Otero González A, De Francisco A, Gayoso P, García F, EPIRCE Study Group (2010) Prevalence of chronic renal disease in Spain: results of the EPIRCE study. *Nefrología* 30: 78-86.

3. Ruiz MA, Pardo A (2005) Health-related quality of life: definition and use in medical practice. *Pharmaco Economics Spanish Research Articles* 2: 31-43.

4. Montilla CP, Duschek S, del Paso GR (2016) Health-related quality of life in chronic kidney disease: predictive relevance of mood and somatic symptomatology. *Nefrología* 36: 275-282.

5. Álvarez-Ude F (2001) Factors associated with the perceived state of health (health-related quality of life) of patients on chronic hemodialysis. *Magazine of the Spanish Society of Nephrological Nursing* 14: 64-68.

6. Souza D, Bernal M (2012) Incidence, prevalence and mortality of renal cancer in Spain: estimates and projections for the period 1998-2022. *Spanish Urological Records* 36: 521-526.

7. Rebollo-Rubio A, Morales-Asencio J, Pons-Raventos E, Mansilla-Francisco J (2015) Review of studies on health-related quality of life in advanced chronic kidney disease in Spain. *Nefrología* 35: 92-109.

8. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, et al. (2017) Mapping to estimate health-state utility from non–preference-based outcome measures: an ISPOR good practices for outcomes research task force report. *Value in Health* 20: 18-27.

9. EuroQoL Group (1990) EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 6: 199-208.

10. Badia X, Roset M, Montserrat S, Herdman M, Segura A (1999) The Spanish version of EuroQol: a description and its applications. European Quality of Life scale. *Medicina clinica* 112: 79-85.

11. Brazier J, Roberts J, Deverill M (2002) The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 21: 271-292.

12. Abellan Perpinan JM (2012) Utilities SF-6D for Spain. 2012/8 usage guide. Seville: Chair of Health Economics. Pablo de Olavide University. Ministry of Health of the Junta de Andalucía.

13. Vilagut G (2005) The Spanish SF36 health questionnaire: a decade of experience and new developments. Health Services Research Unit. Municipal Institute of Medical Research (IMIM-IMAS). Barcelona. Spain. *Gac Sanit* 19: 135-150.

14. Furlong W, Feeny D, Torrance GW, Goldsmith CH, DePauw S, et al. (1998) Multiplicative multi-attribute utilitiy function for the health utilites index mark 3 (HUI3) System: A technical report. *Centre for Health Economics and Policy Analysis* 98: 11.

15. Ruiz M, Rejas J, Soto J, Pardo A, Rebollo I (2003) Adaptation and validation of the Health Utilities Index Mark 3 to Spanish and correction scales in the Spanish population. *Med Clin* 120: 89-96.

16. Cella D, Yount S, Brucker PS, Du H, Bukowski R, et al. (2007) Development and validation of a scale to measure disease-related symptoms of kidney cancer. *Value Health* 10: 285-293.

17. Cella DC, Yount S, Du H, Dhanda R, Gondek K, et al. (2006) Development and validation of the functional assessment of cancer therapy-kidney symptom index (FKSI). *J Support Oncol* 4: 191-199.

18. Butt Z, Peipert J, Webster K, Chen C, Cella D (2013) General population norms for the functional assessment of cancer therapy–Kidney Symptom Index (FKSI). *Cancer* 119: 429-437.

19. Ortega F, Rebollo P, Bobes J, González M, Saiz P (2000) Interpretation of the results of health-related quality of life of patients in replacement therapy for end-stage renal failure. *Nefrología* 20: 431-439.

20. Mukuria C, Rowen D, Harnan S, Rawdin A, Wong R, et al. (2019) An updated systematic review of studies mapping (or cross-walking) measures of health-related quality of life to generic preference-based measures to generate utility values. *Applied health economics and health policy* 2019: 1-19.

21. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, et al. (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research* 20: 1727-1736.

22. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, et al. (2013) Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research* 22: 1717-1727.