Evaluation of PROMIS Preference Scoring System (PROPr) in Patients Undergoing Hemodialysis or Kidney Transplant

Jing Zhang,1 Barry Dewitt,2 Evan Tang,1 Daniel Breitner,1 Mohammed Saqib,1 Dan Li,1 Rabail Siddiqui,1 Nathaniel Edwards,1 John Devin Peipert 3,4 Ron D. Hays 5,5 Janel Hamner,6 and Istvan Mucsi 1,1

Abstract

Background and Objectives A preference-based health utility score (PROPr) can be calculated using Patient-Reported Outcomes Measurement Information System domain scores. We assessed the construct validity of PROPr among patients treated with KRT (hemodialysis or kidney transplant).

Design, setting, participants, & measurements We performed a secondary analysis of data collected in multicenter, cross-sectional studies of adults treated with KRT, recruited between April 2016 to March 2020 in Toronto, Canada. All participants provided informed consent. The outcome was the PROPr score. Coadministered outcome variables included the Short-Form Six-Domain (SF-6D) and EuroQol Five-Domain Five-Level (EQ-5D-5L) scores. Socioeconomic and clinical variables included age, sex, diabetes, eGFR, serum albumin, hemoglobin, KRT, and Charlson Comorbidity Index. Construct validity was assessed through correlations between PROPr and SF-6D or EQ-5D-5L, and associations between PROPr and other exposure variables. Health-condition impact estimates (coefficients for health conditions compared with a referent category, e.g., dialysis versus kidney transplant) were calculated using multivariable linear regression.

Results The mean (SD) age of the 524 participants was 57 (17) years, 58% were male, and 45% were White. Median (interquartile range) score was 0.39 (0.24–0.58) for PROPr, 0.69 (0.58–0.86) for SF-6D, and 0.85 (0.70–0.91) for EQ-5D-5L. Large correlations were observed between PROPr versus SF-6D (0.79; 95% confidence interval [95% CI], 0.76 to 0.82) and EQ-5D-5L (0.71; 95% CI, 0.66 to 0.75). Both PROPr and the other utility indices demonstrated health-condition impact in the expected direction. For example, the estimate for PROPr was −0.17 (95% CI, −0.13 to −0.21) for dialysis (versus kidney transplant), −0.05 (95% CI, −0.11 to 0.01; P = 0.08) for kidney transplant recipients with an eGFR of <45 versus ≥45 ml/min per 1.73 m², and −0.28 (95% CI, −0.32 to −0.33) for moderate/severe versus no/ mild depressive symptoms.

Conclusions Our results support the validity of PROPr among patients treated with KRT.

CJASN 16: 1328–1336, 2021. doi: https://doi.org/10.2215/CJN.01880212

Introduction

Patients with kidney failure require KRT in the form of kidney transplantation or dialysis to survive. Kidney transplant is the best treatment for many patients with kidney failure (1,2); however, kidney transplant recipients may still experience symptoms (3–5) that are associated with a lower health-related quality of life (HRQOL) (6). Patient-reported outcome measures (PROMs) such as the Kidney Disease Quality of Life 36-Item Short Form Survey (KDQOL-36) and the Patient-Reported Outcomes Measurement Information System (PROMIS) item banks can evaluate symptoms and the HRQOL of patients treated with KRT (7,8).

The PROMIS item banks were developed using item response theory (9,10), which allows for more reliable and efficient measurement of patient-valued and clinically actionable symptoms and functioning (e.g., physical function, fatigue) over a wide range of severity. They also serve as important outcome measures in trials or cohort studies.

PROMIS item banks can be administered as fixed-length short forms or using computer adaptive testing. The item response theory-based calibration of PROMIS item banks ensures that scores obtained with different versions (short forms or computer adaptive testing) are near identical (9).

Preference-based health utility measures (such as the Euroqol Five-Domain Five-Level [EQ-5D-5L] or Short Form Six-Domain [SF-6D] measures) are PROMs that yield summary scores that are used to estimate quality-adjusted life years. They account for population preference attributed to combinations of various health states, using various methods (e.g., time trade-off, standard gamble) (11,12). They range from less than zero...
(worse than dead) to one (full health). Anchoring dead at zero allows for the combination of morbidity and mortality to compare the cost-effectiveness of treatment options.

The PROMIS-Preference (PROPr) summary score was developed using standard-gamble methodology (11,13). PROPr is calculated from seven PROMIS domains: cognitive function abilities, depression, fatigue, pain interference, physical functioning, sleep disturbance, and ability to participate in social roles. These domains were selected using a modified Delphi procedure with input from experts and structural independence testing in general community members (12–14).

In this paper, we introduce the PROPr summary score to the nephrology community and assess its construct validity in patients with kidney failure treated with KRT. This assessment is necessary before its use in nephrology clinical research because it confirms comparable properties of PROPr among patient populations with different treatments, illnesses, and symptom experiences.

Materials and Methods

Study Design

This is a secondary analysis of data collected in multicenter, cross-sectional studies evaluating the PROMIS short-form profiles and the PROMIS computer adaptive testing in adult (≥18 years of age) patients on hemodialysis for ≥90 days or who received a kidney transplant ≥30 days before enrollment. Patients were recruited between April 2016 and March 2020 at four hospitals in Toronto, Ontario, Canada. Patients with severe acute illness or those with a diagnosis of dementia and inability to communicate in English, as determined by the clinical team, were excluded (Supplemental Figure 1). We excluded participants who had sociodemographic or clinical data missing from this secondary analysis. Furthermore, we excluded those who had any PROMIS domain scores, other than cognitive function, missing because their PROPr score could not be computed: some participants did not complete all domains on the short forms and others missed domains due to a technical problem with the electronic data capture. Furthermore, participants in the PROMIS computer adaptive testing validation studies initially received only three to four PROMIS domains with corresponding legacy instruments, to reduce question burden. Subsequently, we changed our approach and administered all PROMIS domains to all participants, but administered only half of the legacy questionnaires in each set.

Participants completed study questionnaires on iPads, using an electronic data capture system (DADOS; Techna Institute, University Health Network, Toronto, Canada), during scheduled clinic visits or before/during dialysis treatment. Sociodemographic data (age, sex, marital status, education level, employment status, and ethnic background) were self-reported. Clinical characteristics were collected from the electronic medical record using a standardized data extraction form. We calculated eGFR (ml/min per 1.73 m²) from serum creatinine using the Chronic Kidney Disease Epidemiology Collaboration equation (15). All participants signed informed consent before enrollment. Approval for the studies was obtained from the research ethics boards (REBs) of all participating sites (REB numbers 15-9645, 2016-003, 17-0061, and 377-2017). All study procedures were conducted in accordance with the standards of the University Health Network REB and with the 1964 Declaration of Helsinki and its later amendments. The data were stored in the Comprehensive Psychosocial Research Data System for Patients with CKD (REB number 17-5916-AE).

Health Utility Measures

PROMIS T-scores are based on the item response theory to represent the measured construct, with a mean (SD) of 50 (10) for a relevant reference population (e.g., US general population) (13). We calculated PROPr using scores generated from seven domains that may be assessed by PROMIS short forms and computer adaptive tests. These seven domains, as well as the total number of items included in the respective item banks, are as follows: depression (28), fatigue (95), physical function (165), sleep disturbance (27), ability to participate in social roles and activities (35), pain interference (40), and cognitive function abilities (31). The PROMIS short-form profiles PROMIS-57, PROMIS-43, and PROMIS-29 assess each domain using eight, six, or four items, respectively. When computer adaptive testing is used, an algorithm administers items selected from the item bank on the basis of responses to previous items, until the stopping rule is satisfied. Generally, no more items are administered if reliability reaches 90% or the patient answered 12 items. The vast majority of patients answered four to eight items.

We used the PROPr scoring algorithm that was developed using the standard-gamble valuation in a US sample (11,13) because there are currently no Canadian preference weights. The cognitive function domain was not included in the PROMIS profile instruments used in this study and was estimated using published methods (11). We used PROMIS-29 scores, and not PROMIS-57, due to its brevity and good psychometric properties (16). The theoretic range of PROPr scores is −0.022 to one, with higher score indicating better health status.

The SF-6D was calculated from seven of the SF-12 version 1 items (physical functioning, pain, role limitation, mental health, social functioning, and vitality) (17). Health-state valuations were determined using the standard-gamble technique (17,18). The theoretic range of the SF-6D is ~0.23 to one, with a higher score indicating better health status. Because there is no available Canadian valuation set, we used the UK population preference weights. The EQ-5D-5L covers five domains—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression—using a single item for each domain, measured on five response levels (19). The theoretic range of EQ-5D-5L is ~0.65 to one, with a higher score indicating better health status. In this analysis, the Canadian value set was used (20–22).

Other PROMs

Recent studies considered the revised Edmonton Symptom Assessment System (ESAsr) to routinely assess symptoms experienced by patients on dialysis (23,24). The ESAsr assesses nine symptoms: pain, tiredness, drowsiness, nausea, lack of appetite, shortness of breath, depression, anxiety, and general well-being (5,25). Each symptom is scored between zero and ten, with a lower score indicating a lower symptom
severity. The total ESASr score ranges from zero to 90, with ≥30 suggested as the definition of moderate/severe overall symptom burden (26).

The Patient Health Questionnaire-9 assesses depressive symptoms using nine items (27,28), with total score ranging between zero and 27. A cutoff of ten or more defines moderate/severe depressive symptoms (28).

**Construct Validity**

Convergent validity was assessed by examining correlations between the PROPr and legacy preference–based measures (SF-6D and EQ-5D-5L). We also assessed correlations of the PROMIS domain T-scores with the SF-12 physical and mental component scores and the EQ-5D-5L item scores.

Additionally, in a condition impact analysis (11–13) (see Statistical Analysis below), we examined whether preference-based scores differed between groups, as defined by clinical- and patient-reported variables that are expected to have different HRQOLs (on the basis of clinical experience and the published literature). These groups included the presence or absence of diabetes or significantly impaired kidney graft function (eGFR of <45 ml/min per 1.73 m²), or dialysis versus kidney transplant. Other groups were formed by cutoff points on the basis of clinical relevance and distribution of the variables in the sample. Low albumin was defined as <3.7 g/dl, and low hemoglobin was defined as <10.8 g/dl. The cutoff for low albumin and low hemoglobin represents the lowest tertile that we believe to be clinically relevant, and yield groups large enough for meaningful analysis. We also compared the PROPr between patients with versus without moderate/severe symptom burden (ESASr score of ≥30) and with versus without moderate/severe depressive symptoms (Patient Health Questionnaire-9 score of ten or more).

**Statistical Analyses**

Variables are described using the mean (SD), median (interquartile range [IQR]), or frequency (%), as appropriate. Convergent validity between PROPr and SF-6D was assessed using Pearson correlation, whereas the correlation between PROPr and EQ-5D-5L was assessed using Spearman correlation given the skewness of EQ-5D-5L. Correlations between PROPr and SF-6D/EQ-5D-5L were assessed using Spearman correlation, whereas the correlation between PROPr and SF-6D/EQ-5D-5L item scores was assessed using Pearson correlation. Convergent validity was assessed by examining correlations of the PROMIS domain T-scores with the SF-12 physical and mental component scores and the EQ-5D-5L item scores. Known-groups validity was assessed by examining differences between groups, as defined by clinical- and patient-reported variables that are expected to have different HRQOLs (on the basis of clinical experience and the published literature). These groups included the presence or absence of diabetes or significantly impaired kidney graft function (eGFR of <45 ml/min per 1.73 m²), or dialysis versus kidney transplant. Other groups were formed by cutoff points on the basis of clinical relevance and distribution of the variables in the sample. Low albumin was defined as <3.7 g/dl, and low hemoglobin was defined as <10.8 g/dl. The cutoff for low albumin and low hemoglobin represents the lowest tertile that we believe to be clinically relevant, and yield groups large enough for meaningful analysis. We also compared the PROPr between patients with versus without moderate/severe symptom burden (ESASr score of ≥30) and with versus without moderate/severe depressive symptoms (Patient Health Questionnaire-9 score of ten or more).

**Results**

A total of 677 participants (for PROMIS-29, n=291; for PROMIS computer adaptive testing, n=386) were enrolled in the original PROMIS validation studies. Participants with missing sociodemographic or clinical variables (n=23) were excluded. We also excluded 11 participants who did not complete one or more PROMIS domains and 119 who had not been offered all of the PROMIS item banks due to protocol-related changes or problems with the electronic data capture (Supplemental Figure 1). Therefore, 524 participants were included in the final cohort. The mean (SD; range) age of the 524 participants was 57 (17; 18–92) years. The majority were male (58%), 45% were White, and 56% reported more than high school level education. The median (IQR) dialysis vintage was 3.8 (1.6–6.9) years, whereas median (IQR) time after kidney transplant was 7.4 (2.5–13.2) years (Table 1).

PROPr had the lowest median (IQR) score at 0.39 (0.24–0.58), followed by SF-6D (0.69; 0.58–0.86) and EQ-5D-5L (0.85; 0.70–0.91). PROPr had the lowest proportion of patients who reached the ceiling (the highest possible score in our sample, 0.2%), followed by SF-6D (4%), and both exhibited a near-symmetrical distribution. EQ-5D-5L had the highest ceiling effect (22%) and the most skewed distribution (Table 2, Supplemental Figure 2). The distributions for PROPr29 (n=279) and PROPr computer adaptive testing (n=245) were similar (Supplemental Table 1).

PROPr was strongly correlated with the legacy preference–based measures (for PROPr versus SF-6D, Rho=0.79; 95% CI, 0.76 to 0.82 [P<0.001]; for PROPr versus EQ-5D-5L, Rho=0.71; 95% CI, 0.66 to 0.75 [P<0.001]) (Figure 1). Similar correlations were found between PROPr29 and PROPr computer adaptive testing and SF-6D/EQ-5D-5L (Supplemental Figure 3). Furthermore, we observed the expected correlations of the PROMIS domain scores (which were used to compute PROPr) with the SF-12 summary scores and the EQ-5D-5L domain scores. (Supplemental Table 2).

Associations between eGFR and the preference-based scores followed an overall similar pattern for all scores, with a sharp decline of the scores below the eGFR of 20–35 ml/min per 1.73 m² (Supplemental Figure 4).

In univariable analysis, the health-condition impact estimates for PROPr scores were as follows: for patients with diabetes mellitus, −0.11 (95% CI, −0.07 to −0.15; P<0.001); for patients on hemodialysis (versus kidney transplant), −0.17 (95% CI, −0.13 to −0.21; P<0.001); for patients reporting moderate/severe general symptom burden, −0.25 (95% CI, −0.20 to −0.29; P<0.001); and for patients reporting moderate/severe depressive symptoms, −0.28 (95% CI, −0.22 to −0.33; P<0.001). Furthermore, the health-condition impact estimate for PROPr was −0.05
(95% CI, −0.11 to 0.01; P=0.08; observed power 45%) for eGFR <45 (n=79) versus ≥45 ml/min per 1.73 m² (n=141) among kidney transplant recipients (Figure 2). The magnitude of the health-condition impact estimate for PROP was similar to the coefficients obtained for SF-6D and EQ-5D-5L, except for dialysis versus kidney transplant, general symptom burden, and depressive symptoms. After adjusting for potential confounding variables in a multivariable model, similar impact estimates were observed for all preference-based measures (Figure 3). Health-condition impact estimates for PROP scores calculated from PROMIS-29 or PROMIS computer adaptive testing scores showed similar results, except with self-reported depressive symptoms (Supplemental Figure 5).

**Discussion**

In this study, we assessed the construct validity of PROP, a preference-based health utility score derived from PROMIS domain scores, among patients treated with KRT. PROP had no detectable floor or ceiling effects. PROP was strongly correlated with legacy preference–based measures. Additionally, PROP was associated with other variables in the expected directions.

Our findings provide support for the construct validity of PROP. We propose that PROMIS tools, including PROP, offer significant advantages for clinical research and, potentially, in clinical practice, compared with traditional PROMs. Compared with some of the domains of the frequently used EQ-5D-5L (e.g., combined anxiety/depression), PROMIS...
domains are more specific, better defined, and clinically actionable. Further, PROMIS domains demonstrated greater reliability over a broader range of symptom severity or functional impairment compared with SF-12 or SF-36 (31,32). The precise, versatile PROMIS item banks (administered either as four- or eight-item short forms, or computer adaptive test) have also already been used as outcome measures studies (33,34) and in clinical practice for symptom monitoring (35–37). In addition, PROPr, and the overall mental and physical health summary scores (38), provide reliable measures of health utility and overall HRQOL. PROMIS offers this “three-in-one” combination (39,40) with unparalleled precision and reliability and acceptable question burden.

The correlations between PROPr and the legacy utility scores were large similar to a previous study (12) in samples drawn from the US general population. Our results extend those findings to patients with kidney failure treated with KRT, lending further support to the construct validity of PROPr.

Lower average scores in PROPr, compared with legacy preference–based measures, could be attributed to the fact that the description of the best possible health state (corresponding to the score of one) in PROPr is better compared with the best health states described in legacy measures. As a result, ceiling effects are less common, even in the general population without significant health concerns or in patients
Figure 2. Univariable point estimates of the health-condition impacts on PROPr, SF-6D, and EQ-5D-5L score (represented by different markers) for health conditions in the whole sample. Error bars represent SEM.

Figure 3. Multivariable-adjusted (age, sex, income level, race, education level, and income level) point estimates of the health-condition impacts on PROPr, SF-6D, and EQ-5D-5L score (represented by different markers) for health conditions in the whole sample. Error bars represent SEM.
with chronic conditions and good functional status (such as many kidney transplant recipients).

Clinical and PROM variables were associated with PROPr scores consistently in the expected directions. The strength of associations was similar to the legacy preference-based scores. However, eGFR was not associated with EQ-5D-5L scores among kidney transplant recipients, whereas there was a nonlinear association for SF-6D and PROPr. This association was similar in the low eGFR range (<30 ml/min per 1.73 m²), however, the association was more pronounced for PROPr in the range of 30–120 ml/min per 1.73 m². The condition impact coefficient for significantly impaired graft function for the PROPr score was numerically greater than the other preference-based measures, but it was not statistically significant. The lack of statistical significance was most likely due to the low statistical power of that analysis, because it was restricted to kidney transplant recipients only. This health-condition impact estimate is greater than the suggested minimally important difference (0.04) for PROPr (41). These results are consistent with the hypothesis that PROPr may better differentiate between various health states, and be more responsive in detecting changes over time. However, this will need to be tested in prospective studies.

The relevance of PROPr is potentially supported by our health-condition impact analysis. Condition impacts for all health utility scores, including PROPr, were significantly different from zero for all health conditions, similarly to Hamner et al. (12). Interestingly, PROPr yielded a larger condition impact for dialysis versus kidney transplant and lower versus higher eGFR among kidney transplant recipients. Although PROMIS assesses generic physical, emotional, and social domains of health, many of these domains (e.g., fatigue, sleep disturbance, depression, pain interference) are particularly relevant to patients with kidney failure (3,27,42,43). Furthermore, these domains improve substantially after kidney transplantation. This provides plausibility for a relatively larger difference in utility for kidney transplant compared with dialysis.

We analyzed an all-source PROPr score that was generated by merging the data obtained from PROMIS short forms or PROMIS computer adaptive testing. The use of PROMIS-57 may be somewhat more precise and reliable than PROMIS-29, but measurement characteristics of PROMIS-29 are still excellent, and its brevity is advantageous in reducing responder burden and improving feasibility (8,38). Our results demonstrate the potential versatility of the PROMIS tools, both for research and clinical purposes and to calculate health utility. Scores obtained with various administration methods (short form, computer adaptive testing) can be combined, allowing collection and comparison of data in large health systems, where technical requirements or patient or provider preference makes it necessary to administer PROMIS tools using different methodologies in different settings.

Our sample is sociodemographically and ethnically diverse and included both patients on dialysis and kidney transplant recipients, supporting generalizability.

However, limitations of this work also need to be considered. The PROMIS profile instruments used in our validation studies did not include the cognitive functioning domain, which is needed to calculate PROPr. We generated PROPr scores using predicted cognitive function data, which may affect the accuracy of the scores. However, our results and those of Hamner et al. (12) showed that the PROPr scores generated by this methodology demonstrated strong and consistent convergent and construct validity in diverse samples. A recently created PROMIS-29 +2 Profile version 2.1 short form addresses this concern because it includes two items to measure cognitive functioning (44). Additionally, we performed analysis on a convenience sample, and excluded patients who had significant cognitive impairment and those who were non-English speaking, which could limit generalizability. Validity of PROPr obtained using the translated version of PROMIS tools will need to be established in future studies.

Both SF-6D and PROPr were calculated using non-Canadian population preference weights, potentially resulting in inaccurate representation of population preferences. However, it is likely that, in well-resourced countries with similar cultural and economic backgrounds, the preference weights are sufficiently comparable (45,46). Indeed, Poder et al. (47) observed that SF-6D scores calculated using preference weights from the United Kingdom or the United States were comparable. Therefore, we propose that potential differences between UK, US, or Canadian population preference weights would not qualitatively affect our analyses. However, future studies should develop Canadian value sets for the various health utility scores. Additionally, future longitudinal studies should determine if changes in health conditions are reflected in changes in the PROPr score. Future research should also consider the overlap of measures of depressive symptoms with HRQOL (48).

In conclusion, our results provide strong support for the validity of PROPr among patients with kidney failure on dialysis and after kidney transplant. We propose that PROPr could be considered for use in clinical research and analyses in these patient populations. Future studies should determine the applicability of using PROPr in assessing cost-effectiveness of interventions.

Disclosures

I. Mucsi reports receiving research funding from, and serving as a scientific advisor or member of, Paladin Labs Inc. J.D. Peipert reports receiving research funding from Pfizer and Veloxis Pharmaceuticals. All remaining authors have nothing to disclose.

Funding

The study was supported, in part, by Kidney Foundation of Canada grant KFOCI90008 and the Canadian Institutes of Health Research grant PJT 165915.

Acknowledgments

All authors contributed to the conceptualization and interpretation of data and to writing the manuscript. J. Zhang and I. Mucsi proposed the analysis models and drafted the manuscript; all authors contributed to the finalization of analysis, interpretation of results, and in writing the manuscript; and administration of the project was coordinated between J. Zhang, J. Hamner, B. Dewitt, and I. Mucsi. All authors have approved the final version of the manuscript.
Supplemental Material

This article contains the following supplemental material online at http://cjASN.asnjournals.org/lookup/suppl/doi:10.2215/CJN.01880221/-/DCSupplemental.

Supplemental Table 1. Summary statistics for the PROMIS-29 and PROMIS computer adaptive test and their respective preference scores in the whole sample.

Supplemental Table 2. Correlation between PROMIS domain scores versus SF-12 physical and mental component score and item scores of EQ-5D-5L for the whole sample.

Supplemental Figure 1. Patient flow diagram for the whole sample.

Supplemental Figure 2. Histogram plot presenting the distribution of PROPr (top), SF-6D (middle), and EQ-5D-5L (bottom) with curves of normality to represent the underlying distribution of PROPr, SF-6D, and EQ-5D-5L for the whole sample.

Supplemental Figure 3. Scatter plot and linear fit plots (gray region corresponding to 95% confidence interval) between PROPr and PROPr computer adaptive testing and SF-6D (top) or EQ-5D-5L (bottom) for the whole sample.

Supplemental Figure 4. Nonlinear fitted curves (surrounding lines correspond to 95% confidence interval) of preference-based measures by eGFR for the whole sample.

Supplemental Figure 5. Unvariable model of point estimates of the impacts of health conditions using PROPr, PROPr29, and PROPr computer adaptive testing, adjusted for age, sex, education level, and ethnicity.

References

(PHQ-ADS): Initial validation of structural validity in dialysis patients. *Gen Hosp Psychiatry* 50: 15–19, 2018
37. Papuga MO, Dasilva C, McIntyre A, Mitten D, Kates S, Baumhauer JF: Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. *Health Syst (Basingstoke)* 7: 1–12, 2017

Received: February 6, 2021 Accepted: June 22, 2021

Published online ahead of print. Publication date available at www.cjasn.org.

Supplemental Material

Table of content

Supplemental Table 1. Summary statistics for the PROMIS-29 and PROMIS computer adaptive test and their respective preference scores in the whole sample

Supplemental Table 2. Correlation between PROMIS domain scores versus SF-12 physical and mental component score and item scores of EQ-5D-5L for the whole sample

Supplemental Figure 1. Patient flow diagram for the whole sample.

Supplemental Figure 2. Histogram plot presenting the distribution of PROPr (top), SF-6D (middle) and EQ-5D-5L (bottom) with curves of normality to represent the underlying distribution of PROPr, SF-6D and EQ-5D-5L for the whole sample

Supplemental Figure 3. Scatter plot and linear fit plots (gray region corresponding to 95% confidence interval) between PROPr29 and PROPr computer adaptive testing and SF-6D (top) or EQ-5D-5L (bottom) for the whole sample

Supplemental Figure 4. Non-linear fitted curves (surrounding lines correspond to 95% confidence interval) of preference-based measures by eGFR for the whole sample

Supplemental Figure 5. Univariable model of point estimates of the impacts of health conditions using PROPr, PROPr29 and PROPr computer adaptive testing, adjusted for: age, sex education level and ethnicity. Point estimates are represented by different markers with upper and lower 95% confidence interval for the whole sample
Supplemental Table 1. Summary statistics for the PROMIS-29 and PROMIS computer adaptive test and their respective preference scores in the whole sample

<table>
<thead>
<tr>
<th>PROMIS Item Banks</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>25%</th>
<th>75%</th>
<th>Min</th>
<th>Max</th>
<th>Percent at min.</th>
<th>Percent at max.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical Function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMIS-29</td>
<td>279</td>
<td>42.7</td>
<td>10.6</td>
<td>41.8</td>
<td>34.4</td>
<td>56.9</td>
<td>22.9</td>
<td>56.9</td>
<td>3.2%</td>
<td>28.7%</td>
</tr>
<tr>
<td>PROMIS-CAT</td>
<td>245</td>
<td>41.3</td>
<td>11.1</td>
<td>40.1</td>
<td>34.3</td>
<td>50.1</td>
<td>14.7</td>
<td>75.6</td>
<td>1.6%</td>
<td>0.4%</td>
</tr>
<tr>
<td><strong>Social Roles and activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMIS-29</td>
<td>279</td>
<td>49.5</td>
<td>10.4</td>
<td>49.8</td>
<td>41.7</td>
<td>58.1</td>
<td>29.0</td>
<td>64.1</td>
<td>4.7%</td>
<td>23.7%</td>
</tr>
<tr>
<td>PROMIS-CAT</td>
<td>245</td>
<td>49.4</td>
<td>10.2</td>
<td>49.4</td>
<td>43.8</td>
<td>57.3</td>
<td>21.5</td>
<td>67.5</td>
<td>1.6%</td>
<td>8.6%</td>
</tr>
<tr>
<td><strong>Sleep disturbance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMIS-29</td>
<td>279</td>
<td>49.4</td>
<td>10.4</td>
<td>50.5</td>
<td>41.1</td>
<td>56.1</td>
<td>32.0</td>
<td>73.3</td>
<td>12.2%</td>
<td>2.2%</td>
</tr>
<tr>
<td>PROMIS-CAT</td>
<td>245</td>
<td>51.5</td>
<td>11.3</td>
<td>51.6</td>
<td>44.2</td>
<td>57.9</td>
<td>26.3</td>
<td>83.8</td>
<td>2.0%</td>
<td>2.0%</td>
</tr>
<tr>
<td><strong>Pain interference</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMIS-29</td>
<td>279</td>
<td>54.1</td>
<td>11.3</td>
<td>55.6</td>
<td>41.6</td>
<td>63.8</td>
<td>41.6</td>
<td>75.6</td>
<td>37.3%</td>
<td>7.5%</td>
</tr>
<tr>
<td>PROMIS-CAT</td>
<td>245</td>
<td>52.7</td>
<td>11.9</td>
<td>52.6</td>
<td>38.7</td>
<td>61.5</td>
<td>38.7</td>
<td>83.8</td>
<td>0.8%</td>
<td>0.8%</td>
</tr>
<tr>
<td></td>
<td>PROMIS-29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMIS-CAT</td>
<td>245</td>
<td>50.7</td>
<td>10.0</td>
<td>50.8</td>
<td>44.7</td>
<td>57.6</td>
<td>34.2</td>
<td>84.4</td>
<td>14.3%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Fatigue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROMIS-CAT</td>
<td>245</td>
<td>53.6</td>
<td>10.9</td>
<td>53.5</td>
<td>47.4</td>
<td>59.4</td>
<td>24.3</td>
<td>84.7</td>
<td>0.8%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Preference-based measure</td>
<td>n</td>
<td>Mean</td>
<td>SD</td>
<td>Median</td>
<td>25%</td>
<td>75%</td>
<td>Min</td>
<td>Max</td>
<td>Percent at min.</td>
<td>Percent at max.</td>
</tr>
<tr>
<td>PROPr29</td>
<td>279</td>
<td>0.43</td>
<td>0.24</td>
<td>0.40</td>
<td>0.25</td>
<td>0.64</td>
<td>0.01</td>
<td>0.91</td>
<td>0.4%</td>
<td>2.9%</td>
</tr>
<tr>
<td>PROPr-CAT</td>
<td>245</td>
<td>0.38</td>
<td>0.21</td>
<td>0.37</td>
<td>0.21</td>
<td>0.53</td>
<td>-0.02</td>
<td>0.94</td>
<td>0.4%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

*Footnote: Abbreviations: PROPr29 – PROPr estimated from PROMIS-29; PROPr-CAT – PROPr estimated from PROMIS computer-adaptive test; SD – Standard deviation*
Supplemental Table 2. Correlation between PROMIS domain scores versus SF-12 physical and mental component score and item scores of EQ-5D-5L for the whole sample.

<table>
<thead>
<tr>
<th>PROMIS-short form and computer adaptive test item banks</th>
<th>SF-12 v1 Domains</th>
<th>EQ-5D-5L items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physical component score</td>
<td>Mental component score</td>
</tr>
<tr>
<td>Anxiety</td>
<td>-0.26</td>
<td>-0.49</td>
</tr>
<tr>
<td>Depression</td>
<td>-0.30</td>
<td>-0.58</td>
</tr>
<tr>
<td>Fatigue</td>
<td>-0.50</td>
<td>-0.54</td>
</tr>
<tr>
<td>Pain interference</td>
<td>-0.67</td>
<td>-0.36</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>0.77</td>
<td>0.35</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>-0.36</td>
<td>-0.38</td>
</tr>
<tr>
<td>Social roles</td>
<td>0.60</td>
<td>0.51</td>
</tr>
<tr>
<td>Cognitive function - abilities</td>
<td>0.60</td>
<td>0.58</td>
</tr>
<tr>
<td>Pain intensity</td>
<td>-0.66</td>
<td>-0.29</td>
</tr>
</tbody>
</table>

Footnote: Abbreviations: PROMIS - Patient reported outcome measurement information system; SF-12 - Short form 12-item; EQ-5D-5L - EuroQol 5-domain 5-level
Supplemental Figure 1. Patient flow diagram for the whole sample.

Total number of patients assessed for eligibility (n=2,195)

Excluded (n=1,671)
- Screen failure (n=780)
- Declined participation (n=738)
Missingness (n=153)
- Missing key demographic information (n=23)
- Missing PROMIS domain T-scores (n=130)

Cohort included in the study (n=524)
Supplemental Figure 2. Histogram plot presenting the distribution of PROP (top), SF-6D (middle) and EQ-5D-5L (bottom) with curves of normality to represent the underlying distribution of PROP, SF-6D and EQ-5D-5L for the whole sample.
Footnote: Abbreviations: PROPr - Patient reported outcome measurement information system preference score; SF6D - Short form 6-domain; EQ-5D-5L - EuroQol 5-domain 5-level
Supplemental Figure 3. Scatter plot and linear fit plots (gray region corresponding to 95% confidence interval) between PROPr29 and PROPr-computer adaptive test and SF-6D (top) or EQ-5D-5L (bottom) for the whole sample

Footnote: Abbreviations: SF6D - Short form 6-domain; EQ-5D-5L - EuroQol 5-domain 5-level
Supplemental Figure 4. Non-linear fitted curves (surrounding lines correspond to 95% confidence interval) of preference-based measures by eGFR for the whole sample.
Footnote: Abbreviations: eGFR – estimated glomerular filtration rate; PROPr - Patient reported outcome measurement information system preference score; SF6D - Short form 6-domain; EQ-5D-5L - EuroQol 5-domain 5-level
Supplemental Figure 5. Univariable model of point estimates of the impacts of health conditions using PROPr, PROPr29 and PROPr-CAT, adjusted for: age, sex education level and ethnicity. Point estimates are represented by different markers with upper and lower 95% confidence interval for the whole sample.
Footnote: Abbreviations: PROPr – Patient reported outcomes measurement information system (PROMIS®) preference scores;

PROPr29 – PROPr estimated from PROMIS-29