# Psychometric properties of the Chinese version of PROMIS-29 in patients receiving kidney replacement therapy

ZHAO Qin[1,2], LUO Haibo[3], WANG Liang[4], WANG Yu[4], ZHAO Hongyu[2], DONG Lei[4], LIU Shan[5], ZHU Xiao[2,6], LIU Jia[1,2]

*(1. Health Management Medical Center, Third Xiangya Hospital, Central South University, Changsha 410013, China; 2. Nursing Department, Third Xiangya Hospital, Central South University, Changsha 410013, China; 3. Logistics Support Department, Third Xiangya Hospital, Central South University, Changsha 410013, China; 4. Xiangya School of Nursing, Central South University, Changsha 410013, China; 5. College of Nursing and Public Health, Adelphi University, NY 11530-0701, USA; 6. Research Center of National Health Commission Key Laboratory of Translational Research on Transplantation Medicine on Transplantation Medicine Engineering and Technology, Changsha 410013, China)*

**ABSTRACT**

**Objective:** The Patient-Reported Outcomes Measurement Information System-29 (PROMIS-29) is widely used to measure health-related quality of life and common symptoms, but it has not yet been validated in patients receiving kidney replacement therapy (KRT). This study aims to evaluate the psychometric properties of the Chinese version of PROMIS-29 in patients receiving KRT.

**Methods:** A cross-sectional survey was conducted from November 2023 to January 2024 among adult patients receiving KRT. Participants completed a clinical information questionnaire, PROMIS-29, and conventional measurement questionnaires through an online survey platform. Evaluation indicators included internal consistency, ceiling and floor effects, structural validity, convergent and discriminant validity, and criterion validity, and measurement invariance. Item characteristic curves and item information curves were also plotted.

**Results:** A total of 500 participants were included, with a age of (48.76±12.59) years. Cronbach's α coefficients for the 7 domains of PROMIS-29 ranged from 0.75 to 0.96. Confirmatory factor analysis demonstrated good model fit: Comparative fit index (CFI)=0.975, Tucker-Lewis index (TLI)=0.971, root mean square error of approximation (RMSEA)=0.046, and standardized root mean square residual (SRMR)=0.056. All dimensions met the assumption of unidimensionality. No differential item functioning was detected for the majority of items across subgroups, indicating good measurement invariance. Most items showed ideal item characteristic curves and acceptable item information curves.

**Date of reception:** 2025-06-24

**First author:** ZHAO Qin, Email: zhaoqin274448@163.com, ORCID: 0000-0002-1787-7365; LUO Haibo, Email: 835191041@qq.com, ORCID: 0009-0004-7070-3622

**Corresponding author:** LIU Jia, chief nurse, Email: chucklejl@163.com, ORCID: 0000-0003-2973-2507

**Foundation item:** This work was partly supported by the Hunan Provincial Health High-Level Talent Scientific Research Project (R2023155) and the Natural Science Foundation of Hunan Province (2023JJ40885), China.

**Open access:** This is an open access article under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) (https://creativecommons.org/licenses/by-nc-nd/4.0/).

©Journal of Central South University (Medical Science). All rights reserved.

**Conclusion:** The Chinese version of PROMIS-29 has good reliability, validity, and measurement invariance among Chinese patients receiving KRT and can be used to assess health-related quality of life and symptoms in this population.

**KEY WORDS** PROMIS-29; patient-reported outcome measurement; kidney replacement therapy; psychometric properties; reliability and validity testing; differential item functioning

# 中文版PROMIS-29在肾脏替代治疗患者中的心理测量学特性

赵琴[1,2]，罗海波[3]，汪靓[4]，王昱[4]，赵洪圉[2]，董蕾[4]，刘珊[5]，朱肖[2,6]，刘佳[1,2]

(1. 中南大学湘雅三医院健康管理中心，长沙 410013；2. 中南大学湘雅三医院护理部，长沙 410013；3. 中南大学湘雅三医院后勤保障部，长沙 410013；4. 中南大学湘雅护理学院，长沙 410013；5. 艾德菲大学护理与公共卫生学院，美国 纽约 11530-0101；6. 国家卫生健康委员会移植医学转化研究重点实验室，长沙 410013)

**[摘 要] 目的：**患者报告结局测量信息系统-29(Patient-Reported Outcomes Measurement Information System-29，PROMIS-29)广泛应用于健康相关生活质量和常见症状的测量，但尚未在接受肾脏替代治疗的患者中进行验证。本研究旨在评估中文版PROMIS-29在肾脏替代治疗患者中的心理测量学特性。**方法：**采用横断面调查研究，于2023年11月至2024年1月招募成人肾脏替代治疗患者。患者通过在线调查平台填写临床资料问卷、PROMIS-29和传统测量问卷。评价指标包括内部一致性、天花板和地板效应、结构效度、聚合和区分效度、效标效度和测量等价性。同时绘制项目特征曲线和测试信息曲线。**结果：**最终纳入500例研究对象，年龄为(48.76±12.59)岁。PROMIS-29中7个领域的克龙巴赫α系数为0.75~0.96。验证性因子分析结果显示模型拟合良好：比较拟合指数(comparative fit index，CFI)=0.975，塔克-刘易斯指数(Tucker-Lewis index，TLI)=0.971，近似误差均方根(root mean square error of approximation，RMSEA)=0.046，标准化均方根残差(standardized root mean square residual，SRMR)=0.056，且各维度均符合单维性假设。大多数条目未检测到项目功能差异，具有良好的测量等价性。大部分条目拟合出理想的项目特征曲线和可接受的测试信息曲线。**结论：**中文版PROMIS-29在中国肾脏替代治疗患者中具有良好的信效度和测量等价性，可用于肾脏替代治疗患者健康相关生活质量和症状的评估。

**[关键词]** 患者报告结局测量信息系统-29；患者报告结局测量；肾脏替代治疗；心理测量学特性；信效度检验；项目功能差异

End-stage renal disease (ESRD), also known as kidney failure, refers to the final stage of chronic kidney disease (CKD), where the kidneys lose 85% − 90% function and can no longer sustain life in the long term[1-2]. Kidney replacement therapy (KRT) is the primary treatment for ESRD aimed at alleviating symptoms and sustaining life, including hemodialysis, peritoneal dialysis, and kidney transplantation[3]. However, due to continuous dialysis, taking immunosuppressants, and costly medical expenses, patients receiving KRT still experience substantial symptoms and live with poor health-related quality of life (HRQoL)[4-5]. Fatigue is the most common symptom among dialysis patients (70%), and over half of kidney transplant recipients report bone or joint pain (55%), constipation (53%), and muscle weakness (56%)[6]. These symptoms are associated with increased complications and mortality rates, yet they are not routinely assessed in clinical care[7-8].

Patient-reported outcome measures (PROMs) collect experiences and feelings directly reported from patients, without any interpretation from others, which are garnering increasing attention from researchers and healthcare professionals[9-11]. Derived from item response theory (IRT), the National Institutes of Health developed the Patient-Reported Outcomes Measurement Information System (PROMIS) to assess health-related outcomes in both healthy and diseased populations[12].

©Journal of Central South University (Medical Science). All rights reserved.

The scores of measures in PROMIS can be converted into standardized T-scores[13], allowing for comparability of health status across languages, conditions, and diseases. Despite the advantage of comparability, instruments in PROMIS need to be assured of good psychometric properties before being applied to health systems.

PROMIS profiles consist of 7 PROMIS short forms and one item measuring pain intensity, and they include 3 profile versions: PROMIS-29 (4 items for each short form), PROMIS-43 (6 items for each short form), and PROMIS-57 (8 items for each short form). The 7 short forms measure 7 health-related core domains, including physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities, and pain interference[14]. The shorter profiles are nested within the longer profiles. The CKD working group of the International Consortium of Health Outcomes Measurement (ICHOM) has recommended PROMIS-29 to evaluate HRQoL and symptoms (e. g., pain, fatigue, depression) of CKD[15]. As a precise, brief, and comparable tool, PROMIS-29 has been translated into various languages and validated among different populations[16-18]. Through rigorous translation and cross-cultural adaptation process, Cai, et al[19] have formed the Chinese version of PROMIS-29 (nested within PROMIS-57) and validated it in patients with breast cancer. However, the Chinese version has not yet been validated in the field of nephrology. Given the cultural and linguistic disparities, it is imperative to evaluate the psychometric properties of the Chinese version of PROMIS-29 before its implementation in nephrological clinical practice.

Patients undergoing KRT require regular assessment of symptoms and HRQoL to evaluate and improve treatment outcomes, and PROMs offer a valuable means to capture patient perspectives. Therefore, our study aims to evaluate the reliability and validity of the PROMIS-29 in adult patients on KRT in China. This validation would facilitate the integration of PROMIS-29 into nephrology practice and enhance HRQoL and symptom assessment in dialysis and transplant patients.

# 1 Materials and methods

## 1.1 Ehtics statements

Ethical approval was obtained from the Ethics Committee of the Third Xiangya Hospital, Central South University (No. 23892). Participants were informed about the purpose and content of the study, and provided informed consent before filling out the questionnaires. All data were accessible only to the researchers and kept strictly confidential.

## 1.2 Study design and sample

This was a cross-sectional study evaluating the Chinese version of PROMIS-29 in a sample of adult patients receiving KRT. Patients aged 18 years or older, who had undergone dialysis or kidney transplantation for more than 3 months, were recruited from November 2023 to January 2024 at a comprehensive hospital in Hunan, China. Patients in the acute phase of illness, with cognitive impairments preventing them from completing the questions on their own, or lacking basic reading ability were excluded. The sample size was determined with reference to the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline for study design[20]. Structural validity requires at least 5 times, the number of items and IRT-based analyses require at least 150 observations per group. Based on the projected analyses in this study, a minimum sample of 450 cases was considered adequate.

## 1.3 Data collection

Questionnaires including socio-demographic and clinical data, the PROMIS-29, and legacy measurements were completed on an online survey system (Wen Juan Xing), accessible conveniently by scanning a quick response (QR) code. Trained researchers and nurses administered the QR code in the WeChat groups for patients, transplant wards, nephrology wards, and hemodialysis centers. To ensure the quality of the questionnaire, all questions must be completed before submission, resulting in no missing data, and each WeChat account can only submit the questionnaire once. Additionally, questionnaires with a response time of less than 350 s or those that failed the attention-check item were excluded from the analysis. Two authors examined

©Journal of Central South University (Medical Science). All rights reserved.

the data and determined the final data to be included in the analysis based on inclusion and exclusion criteria.

## 1.4 Measures

### 1.4.1 Sociodemographic and clinical characteristics

Sociodemographic characteristics included self-reported age, gender, body mass index (BMI), marital status, years of education, working status, and household monthly income. Clinical characteristics included self-reported primary kidney disease, time on dialysis or time since transplant, and comorbidities. Comorbidities were assessed using a multiple-choice item, allowing participants to select all conditions that applied.

### 1.4.2 PROMIS-29

PROMIS-29 profile v2.1 comprises 29 items assessing 7 health-related domains (physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities, and pain interference) and a item pain intensity[21]. Except for the single pain intensity item utilizing a numeric rating scale from 0 to 10, each of the other 7 domains contains 4 items rated on a 5-point Likert scale (score 1−5)[22]. Raw scores for each domain are converted into T-scores using HealthMeasures Scoring Service, where a higher score represents a greater degree of the trait under assessment. T-scores are standardized scores with a mean of 50 and a standard deviation of 10, reflecting the respondent's standing relative to the general population, and therefore do not have a fixed absolute range[13].

### 1.4.3 Legacy measures

We chose validated and brief instruments to compare with PROMIS-29. The 12-Item Short Form Health Survey (SF-12) is a widely used questionnaire that consists of 8 dimensions: Physical functioning (PF), role physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH)[23]. According to the reference formula, the responses of the first 4 dimensions can be aggregated to calculate the physical component summary (PCS), while the scores from the last 4 dimensions can be calculated to determine the mental component summary (MCS)[24]. The score of each dimension and summary ranges from 0 to 100, with a higher score indicating better HRQoL. The Chinese version of SF-12 demonstrated good convergent and discriminant validity, along with good internal consistency (Cronbach's α=0.910)[25].

The Patient Health Questionnaire (PHQ-9) was developed based on the 9 symptoms outlined for diagnosing depression in the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (DSM-IV) and contains 9 items. This scale inquiries about the frequency of being bothered over the past 2 weeks, with each item scored on a scale from 0 to 3 (0 for "not at all" and 3 for "nearly every day")[26]. The total score ranges from 0 to 27, with higher scores indicating higher levels of depression. The Chinese version of PHQ-9 was validated with a Cronbach's α of 0.86 and a test-retest reliability of 0.86[27].

The Generalized Anxiety Disorder 7-Item Scale (GAD-7) was derived from the diagnostic criteria for generalized anxiety disorder (GAD) in the DSM-IV, and it is extensively employed for the assessment and screening of GAD[28]. It consists of 7 questions asking about the frequency of anxiety symptoms over the past 2 weeks. As with the PHQ-9, each entry is rated on a scale of 0−3, with a total score of 0−21. The Chinese version showed good internal consistency, with a Cronbach's α of 0.928[29].

## 1.5 Statistical analysis

### 1.5.1 Descriptive analysis

Descriptive analysis was performed to illustrate sample characteristics and PROMIS-29 scores. Mean± standard deviations and median ($P_{25}$, $P_{75}$) were analyzed for continuous variables, while frequencies and percentages were computed for categorical variables. We used the $\chi^2$ test, $F$ test and the Kruskal-Wallis $H$ test to compare the differences among patients undergoing hemodialysis, peritoneal dialysis, and kidney transplant recipients.

### 1.5.2 Reliability and floor/ceiling effects

Internal consistency was measured with Cronbach's α. It was regarded as excellent, good, and acceptable when coefficient α is ⩾0.90, 0.80−0.89, and 0.70−0.79, respectively[30]. We assessed floor and ceiling effects for each domain and the single pain intensity item by calculating the proportions of participants scoring at the lowest (floor) and highest (ceiling) levels. A threshold of exceeding 30% was considered to indicate significant floor or ceiling effects[31].

©Journal of Central South University (Medical Science). All rights reserved.

#### 1.5.3 Construct and criterion validity

Confirmatory factor analysis (CFA) using the maximum likelihood estimation method was employed to assess the one-factor structure of each domain of the PROMIS-29, confirming its unidimensionality. Based on the results of the CFA of each single domain, the 7-factor structure was examined. The model fit indices were comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The model fit is considered adequate when both CFI and TLI exceed 0.95, RMSEA is below 0.06, and SRMR is less than 0.08[32].

Convergent and discriminant validity was measured by calculating average variance extracted (AVE) and composite reliability (CR) of each construct. Good convergent validity was indicated if the AVE value for each construct was >0.50 and the CR value was >0.70[33]. The square root of AVE for a particular dimension being greater than the correlation coefficients between this dimension and others indicated good discriminant validity[34]. Criterion validity was examined by analyzing Spearman's correlations between PROMIS-29 domains and similar constructs in legacy measures. The values above 0.50 were considered acceptable criterion validity[18].

#### 1.5.4 Measurement invariance

Measurement invariance is an essential assumption for construct validity, as it ensures that the measurement assesses the same construct consistently across different groups or conditions[35]. In our study, measurement invariance was determined by analyzing whether differential item functioning (DIF) existed using the graded response model (GRM), which assumes that item responses are conditionally independent and residual correlations are not allowed. Using the "lordif" package in R[36], DIF was tested by ordinal Logistic regression method with respect to age (younger group ≤49 years, or older group >49 years), gender (male or female), years of education (≤12 years, or >12 years), and treatment (hemodialysis, peritoneal dialysis, or kidney transplant). The age group was divided according to the median. Monte Carlo simulations (S=1 000 replications) were employed to construct the empirical null distribution of McFadden's pseudo $R^2$ change under the null hypothesis of no DIF, and the 95th percentile of this distribution was used as the item-specific critical value to address multiple comparisons[36]. As suggested by PROMIS Health Organization, McFadden's pseudo $R^2$ with a change of ≥0.02[37] was employed as the threshold to signify significant DIF.

#### 1.5.5 IRT analysis

To visually investigate the difficulty, discrimination, and information value of the measurement, item characteristic curves (ICCs) and test information curves (TICs) were plotted based on the GRM model fit. ICCs provide the probability of a particular response to an item given different levels of latent trait (θ) [38]. TICs depict the distribution of test information across various ability levels, enabling the assessment of the test's measurement precision and accuracy[39]. In our study, we employed the "mirt" package in R to perform analyses of ICCs and TICs[40].

Except for the CFA conducted using AMOS version 26.0, and the DIF, ICCs, and TICs measured by R version 4.3.3, all other statistical analyses were performed in SPSS version 27.0. The two-tailed $P$ value <0.05 was considered statistically significant.

## 2 Results

### 2.1 Participant characteristics

A total of 559 adult patients with KRT were invited and assessed for eligibility. Fifty-nine participants were excluded based on the inclusion and exclusion criteria, resulting in a final sample of 500 patients for analysis. Among the cohort, the age was (48.76±12.59) years, 56.0% of them were male, and the majority (77.0%) were married. Most participants had undergone dialysis or received transplants for a period spanning 1 to 5 years (52.3% for patients on hemodialysis, 55.6% for patients on peritoneal dialysis, and 45.9% for kidney transplant recipients). Excluding unclear cases of primary kidney disease, the top 3 causes were glomerulonephritis/sclerosis (23.2%), hypertension/renal vascular disease (20.0%), and diabetes mellitus (10.4%). Statistically significant differences were observed in age, gender, years of education, working status, household monthly income, primary kidney disease, and partial comorbidities among patients receiving different treatments. In health-related

©Journal of Central South University (Medical Science). All rights reserved.

PROMIS-29, kidney transplant recipients got higher scores in positive domains and lower scores in negative domains than patients on dialysis. The results of sample characteristics are presented in Table 1.

**Table 1 Characteristics of participants**

| Characteristics | Total (*n*=500) | Patients on HD (*n*=151) | Patients on PD (*n*=153) | KTR (*n*=196) | $\chi^2/F/H$ | *P* |
|---|---|---|---|---|---|---|
| Age/Years | 48.76±12.59 | 54.07±13.44 | 46.86±12.83 | 46.15±10.32 | 23.709 | <0.001 |
| Gender/[No.(%)] | | | | | 8.289 | 0.016 |
| Male | 280(56.0) | 92(60.9) | 71(46.4) | 117(59.7) | | |
| Female | 220(44.0) | 59(39.1) | 82(53.6) | 79(40.3) | | |
| BMI/[No.(%)] | | | | | 8.497 | 0.075 |
| <18.5 kg/m$^2$ | 55(11.0) | 24(15.9) | 14(9.2) | 17(8.7) | | |
| 18.5−<24.0 kg/m$^2$ | 295(59.0) | 82(54.3) | 86(56.2) | 127(64.8) | | |
| ≥24.0 kg/m$^2$ | 150(30.0) | 45(29.8) | 53(34.6) | 52(26.5) | | |
| Marital status/[No.(%)] | | | | | 3.089 | 0.543 |
| Single | 60(12.0) | 14(9.3) | 18(11.8) | 28(14.3) | | |
| Married | 385(77.0) | 117(77.5) | 118(77.1) | 150(76.5) | | |
| Divorced or widowed | 55(11.0) | 20(13.2) | 17(11.1) | 18(9.2) | | |
| Years of education/[No.(%)] | | | | | 15.350 | <0.001 |
| ≤12 | 308(61.6) | 102(67.5) | 106(69.3) | 100(51.0) | | |
| >12 | 192(38.4) | 49(32.5) | 47(30.7) | 96(49.0) | | |
| Working status/[No.(%)] | | | | | 28.388 | <0.001 |
| Yes | 159(31.8) | 31(20.5) | 39(25.5) | 89(45.4) | | |
| No | 341(68.2) | 120(79.5) | 114(74.5) | 107(54.6) | | |
| Household monthly income/[No.(%)] | | | | | 20.071 | 0.003 |
| 0−5 000 yuan | 263(52.6) | 88(58.3) | 95(62.1) | 80(40.8) | | |
| 5 001−10 000 yuan | 152(30.4) | 43(28.5) | 33(21.6) | 76(38.8) | | |
| 10 001−20 000 yuan | 66(13.2) | 16(10.6) | 19(12.4) | 31(15.8) | | |
| >20 000 yuan | 19(3.8) | 4(2.6) | 6(3.9) | 9(4.6) | | |
| Primary kidney disease/[No.(%)] | | | | | 63.398 | <0.001 |
| Glomerulonephritis/sclerosis | 116(23.2) | 28(18.5) | 39(25.5) | 49(25.0) | | |
| Pyelonephritis | 26(5.2) | 5(3.3) | 11(7.2) | 10(5.1) | | |
| Polycystic kidney disease | 16(3.2) | 4(2.6) | 7(4.6) | 5(2.6) | | |
| Other congenital/hereditary kidney diseases | 13(2.6) | 0(0) | 6(3.9) | 7(3.6) | | |
| Hypertension/renal vascular disease | 100(20.0) | 32(21.2) | 36(23.5) | 32(16.3) | | |
| Diabetes mellitus | 52(10.4) | 37(24.5) | 9(5.9) | 6(3.1) | | |
| Miscellaneous | 49(9.8) | 11(7.3) | 14(9.2) | 24(12.2) | | |
| Unknown | 128(25.6) | 34(22.5) | 31(20.3) | 63(32.1) | | |
| Time on dialysis or time since transplant/[No.(%)] | | | | | | |
| 3−12 months | | 37(24.5) | 36(23.5) | 22(11.2) | | |
| 13−60 months | | 79(52.3) | 85(55.6) | 90(45.9) | | |
| 61−120 months | | 25(16.6) | 25(16.3) | 52(26.6) | | |
| >120 months | | 10(6.6) | 7(4.6) | 32(16.3) | | |

©Journal of Central South University (Medical Science). All rights reserved.

**Table 1 Characteristics of participants**

| Characteristics | Total (*n*=500) | Patients on HD (*n*=151) | Patients on PD (*n*=153) | KTR (*n*=196) | $\chi^2/F/H$ | *P* |
|---|---|---|---|---|---|---|
| Comorbidities/[No.(%)] | | | | | | |
| Hypertension | 330(66.0) | 115(76.2) | 121(79.1) | 94(48.0) | 47.046 | <0.001 |
| Diabetes mellitus | 87(17.4) | 43(28.5) | 15(9.8) | 29(14.8) | 19.958 | <0.001 |
| Chronic lung disease | 25(5.0) | 9(6.0) | 6(3.9) | 10(5.1) | 0.672 | 0.715 |
| Gastrointestinal disorders | 65(13.0) | 17(11.3) | 17(11.1) | 31(15.8) | 2.262 | 0.323 |
| Cardiovascular disease | 84(16.8) | 35(23.2) | 27(17.6) | 22(11.2) | 8.833 | 0.012 |
| Musculoskeletal disorders | 57(11.4) | 19(12.6) | 17(11.1) | 21(10.7) | 0.313 | 0.855 |
| Malignancy | 5(1.0) | 2(1.3) | 0(0) | 3(1.5) | 2.263 | 0.336 |
| Others | 37(7.4) | 14(9.3) | 10(6.5) | 13(6.6) | 1.107 | 0.575 |
| PROMIS-29 Physical function | 48.6(42.9, 57.0) | 44.2(38.7, 48.7) | 47.3(42.4, 47.0) | 57.0(47.8, 57.0) | 85.581 | <0.001 |
| PROMIS-29 Anxiety | 56.0(40.3, 63.5) | 56.0(48.6, 63.5) | 56.2(48.1, 63.5) | 54.1(40.3, 59.6) | 12.905 | 0.002 |
| PROMIS-29 Depression | 55.9(41.0, 62.2) | 57.3(41.0, 62.2) | 56.1(52.0, 62.2) | 54.1(41.0, 57.5) | 24.658 | <0.001 |
| PROMIS-29 Fatigue | 51.0(48.6, 57.0) | 53.2(48.6, 58.9) | 53.2(48.6, 57.1) | 48.6(44.6, 53.2) | 63.052 | <0.001 |
| PROMIS-29 Sleep disturbance | 52.8(48.2, 55.5) | 52.8(49.6, 57.1) | 52.8(49.6, 56.2) | 51.1(45.0, 54.5) | 23.628 | <0.001 |
| PROMIS-29 Ability to participate in social roles and activities | 51.8(46.0, 55.5) | 49.9(44.2, 51.8) | 49.9(44.2, 53.6) | 51.8(50.0, 62.8) | 43.839 | <0.001 |
| PROMIS-29 Pain interference | 55.7(41.6, 58.5) | 55.7(41.6, 61.3) | 55.7(41.6, 58.7) | 50.2(41.6, 55.7) | 20.278 | <0.001 |
| PROMIS-29 Pain intensity | 2(1, 4) | 3(2, 5) | 2(1, 4) | | 16.321 | <0.001 |
| SF-12 PCS | 41.2(35.9, 46.9) | 37.7(32.2, 42.27) | 39.2(33.7, 45.1) | 46.2(41.3，50.6) | 115.692 | <0.001 |
| SF-12 MCS | 46.2(38.8, 51.7) | 42.9(36.7, 49.8) | 43.46(38.1, 50.6) | 49.7(43.4, 54.5) | 38.773 | <0.001 |
| PHQ-9 | 4.0(1.0, 7.0) | 8(3, 12) | 6(3, 10) | 4(1, 8) | 31.705 | <0.001 |
| GAD-7 | 6.0(2.0, 9.0) | 6(1, 8) | 4(1, 7) | 4(0, 7) | 7.012 | 0.030 |

Values of variables are expressed as means±standard deviation for normally distributed continuous variable, medians (25th percentile, 75th percentile) for non-normally distributed continuous variable, and frequencies (percentages) for categorical variable. Between-group comparisons were performed using Kruskal-Wallis test or analysis of variance for continuous variables, depending on their distribution, and $\chi^2$ test for categorical variables. HD: Hemodialysis; PD: Peritoneal dialysis; KTR: Kidney transplant recipients; PROMIS-29: Patient-Reported Outcomes Measurement Information System-29; SF-12: 12-Item Short Form Health Survey; PHQ-9: Patient Health Questionnaire; GAD-7: Generalized Anxiety Disorder 7-item scale; PCS: Physical component; MCS: Mental component summary; BMI: Body mass index. The 7 domains of PROMIS-29 were described using T-scores.

## 2.2 Reliability, and floor/ceiling effects

The PROMIS-29 T-scores and pain intensity raw score are displayed in Table 1. Physical function domain obtained the lowest median T-score at 48.6 (42.9, 57.0) and anxiety domain got the highest at 56.0 (40.3, 63.5). Internal consistency was sufficient for all domains, with the Cronbach's α ranging from 0.75 to 0.96 (Table 2). According to the criteria of ⩾30%, a notable ceiling effect was observed in physical function domain (42.2%), whereas pain interference domain exhibited a substantial floor effect (37.4%; Table 2).

## 2.3 Structural validity

CFA was individually conducted across the 7 domains of PROMIS-29, and the results demonstrated satisfactory model fit for all subscales. As shown in Table 3, the CFI and TLI values for all domains were > 0.95, RMSEA values ranged between 0.000 to 0.069, and all SRMR values were <0.08. Therefore, the unidimensional structure of the subscales was confirmed. Moreover, the CFA of the whole PROMIS-29 also revealed a favorable model fit (CFI=0.975, TLI=0.971, RMSEA=0.046, SRMR=0.056), confirming the 7-

©Journal of Central South University (Medical Science). All rights reserved.

factor structure. Except for item Sleep116 within the sleep disturbance domain (factor loading=0.17), factor loadings for all other items were above 0.5, ranging from 0.70 to 0.97.

**Table 2 Reliability, floor, and ceiling effects of PROMIS-29 and legacy measurements**

| Seven domains and 1 item | Cronbach's α of PROMIS-29 | PROMIS-29 | | Legacy measurements | | |
|---|---|---|---|---|---|---|
| | | Floor/% | Ceiling/% | Classification | Floor/% | Ceiling/% |
| Physical function | 0.91 | 1.8 | 42.2 | SF-12 PCS | 0.2 | 0.2 |
| Anxiety | 0.94 | 25.6 | 1.6 | GAD-7 | 22.0 | 0.6 |
| Depression | 0.94 | 26.6 | 1.0 | PHQ-9 | 14.0 | 0.2 |
| Fatigue | 0.93 | 7.8 | 2.0 | SF-12 VT | 6.0 | 5.0 |
| | | | | PHQ-9 item 4 | 23.6 | 8.2 |
| Sleep disturbance | 0.75 | 4.0 | 0.2 | PHQ-9 item 3 | 33.2 | 6.0 |
| Ability to participate in social roles and activities | 0.91 | 2.0 | 17.0 | SF-12 SF | 7.2 | 17.8 |
| Pain interference | 0.96 | 37.4 | 1.8 | SF-12 BP | 4.0 | 25.2 |
| Pain intensity | NA | 0.6 | 2.8 | SF-12 BP | 4.0 | 25.2 |

VT: Vitality; SF: Social functioning; BP: Bodily pain; NA: Not applicable.

**Table 3 Model fit indices of confirmatory factor analysis for PROMIS-29 domain**

| Domain | $\chi^2/df$ | CFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|
| Physical function | 0.286 | 1.000 | 1.003 | 0 | 0.002 |
| Anxiety | 1.807 | 1.000 | 0.997 | 0.040 | 0.003 |
| Depression | 0.895 | 1.000 | 1.000 | 0 | 0.004 |
| Fatigue | 0.916 | 1.000 | 1.000 | 0 | 0.002 |
| Sleep disturbance | 1.778 | 0.999 | 0.994 | 0.039 | 0.010 |
| Ability to participate in social roles and activities | 0.589 | 1.000 | 1.002 | 0 | 0.004 |
| Pain interference | 3.346 | 0.999 | 0.994 | 0.069 | 0.004 |
| PROMIS-29 | 2.043 | 0.975 | 0.971 | 0.046 | 0.056 |

CFI: Comparative fit index; TLI: Tucker-Lewis index; RMSEA: Root mean square error of approximation; SRMR: Standardized root mean square residual. Domains of physical function, anxiety, fatigue, sleep disturbance, and pain interference were modified according to the modification index.

## 2.4 Convergent validity and discriminant validity

In our study, we analyzed AVE, CR, and correlation coefficients through the AMOS plugin developed by Gaskin, et al[41]. Good convergent validity was evidenced, having AVE values that fall within the range of 0.506 to 0.863 and CR values of 0.778 to 0.962 (Table 4). Additionally, the square root of AVE value for each domain displayed in the diagonal were all greater than the correlation coefficients with other dimensions in its row and column, signaling robust discriminant validity (Table 4).

## 2.5 Criterion validity

Correlations between PROMIS-29 subscales and the corresponding structure of legacy measurements were presented in Table 5, showing acceptable criterion validity ($r$>0.50). Strong correlations were found between the anxiety domain and GAD-7 ($r$=0.71, 95% *CI* 0.66 to 0.75), as well as between the depression domain and PHQ-9 ($r$=0.73, 95% *CI* 0.68 to 0.77).

©Journal of Central South University (Medical Science). All rights reserved.

**Table 4 Convergent validity and discriminant validity**

| Domain | CR | AVE | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|
| Physical function | 0.909 | 0.716 | 0.846 | | | | | | |
| Anxiety | 0.933 | 0.778 | −0.401*** | 0.882 | | | | | |
| Depression | 0.943 | 0.805 | −0.370*** | 0.818*** | 0.897 | | | | |
| Fatigue | 0.927 | 0.760 | −0.514*** | 0.617*** | 0.654*** | 0.872 | | | |
| Sleep disturbance | 0.778 | 0.506 | −0.370*** | 0.457*** | 0.482*** | 0.611*** | 0.711 | | |
| Ability to participate in social roles and activities | 0.915 | 0.731 | 0.508*** | −0.464** | −0.511*** | −0.556*** | −0.435*** | 0.855 | |
| Pain interference | 0.962 | 0.863 | −0.482*** | 0.458*** | 0.517*** | 0.609*** | 0.432*** | −0.530*** | 0.929 |

CR: Composite reliability; AVE: Average variance extracted. Diagonal is the square root of AVE. ** $P$<0.01, *** $P$<0.001.

**Table 5 Correlations between PROMIS-29 domains and legacy measurements**

| Seven domains and 1 item | Legacy measurement | Correlation coefficient (95% *CI*) |
|---|---|---|
| Physical function | SF-12 PCS | 0.65(0.59 to 0.70) |
| Anxiety | GAD-7 | 0.71(0.66 to 0.75) |
| Depression | PHQ-9 | 0.73(0.68 to 0.77) |
| Fatigue | SF-12 VT | −0.60(−0.65 to −0.54) |
| | PHQ-9 item 4 | 0.66(0.61 to 0.71) |
| Sleep disturbance | PHQ-9 item 3 | 0.66(0.61 to 0.71) |
| Ability to participate in social roles and activities | SF-12 SF | 0.51(0.44 to 0.57) |
| Pain interference | SF-12 BP | −0.57(−0.63 to −0.51) |
| Pain Intensity | SF-12 BP | −0.51(−0.58 to −0.44) |

All correlation coefficients were significant with $P$<0.001.

## 2.6 Measurement invariance

DIF was detected only for item PFA21 (“Are you able to go up and down stairs at a normal pace”) across education and treatment groups. No DIF was detected for remaining items across different age, gender, years of education, and treatment groups. For all items except PFA21, the changes in McFadden’s pseudo $R^2$ values were less than 0.02, indicating negligible DIF. Overall, these findings support the measurement invariance of the instrument across demographic and clinical subgroups. Detailed data is presented in Supplementary Table 1 (https://doi.org/10.57760/sciencedb.xbyxb.00187).

## 2.7 Item characteristic and test information

As illustrated in Figure 1, the ICCs of most items exhibited monotonicity in their first and last curves, while the remaining 3 curves followed a normal distribution, which was close to the ideal form. However, the curves of item Sleep116 showed a flat pattern and crossover, indicating poor discrimination of individuals with different abilities.

The results of TIC for each domain are depicted in Figure 2. The sleep disturbance provided the lowest information, with maximum information reaching about 10. This was followed by the fatigue and sociability domains, with a maximum of about 20. The pain interference domain demonstrated a maximum information content exceeding 60. However, its curve displayed relatively apparent spikes, indicating the test’s high precision in measuring at specific levels of ability.

©Journal of Central South University (Medical Science). All rights reserved.

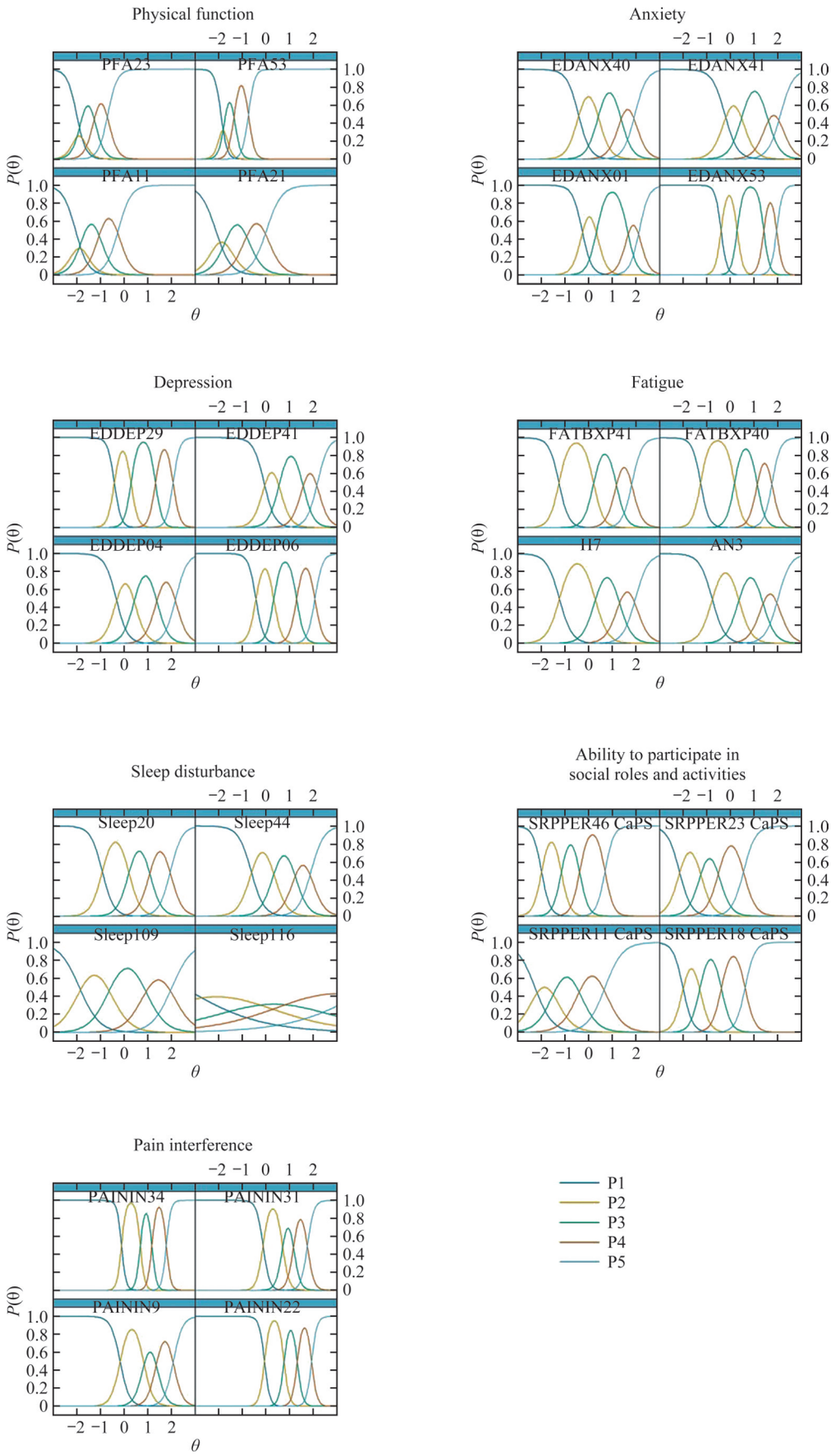



**Figure 1 Item characteristic curves for each domain of PROMIS-29**

PROMIS-29: Patient-Reported Outcomes Measurement Information System Profile-29.

©Journal of Central South University (Medical Science). All rights reserved.

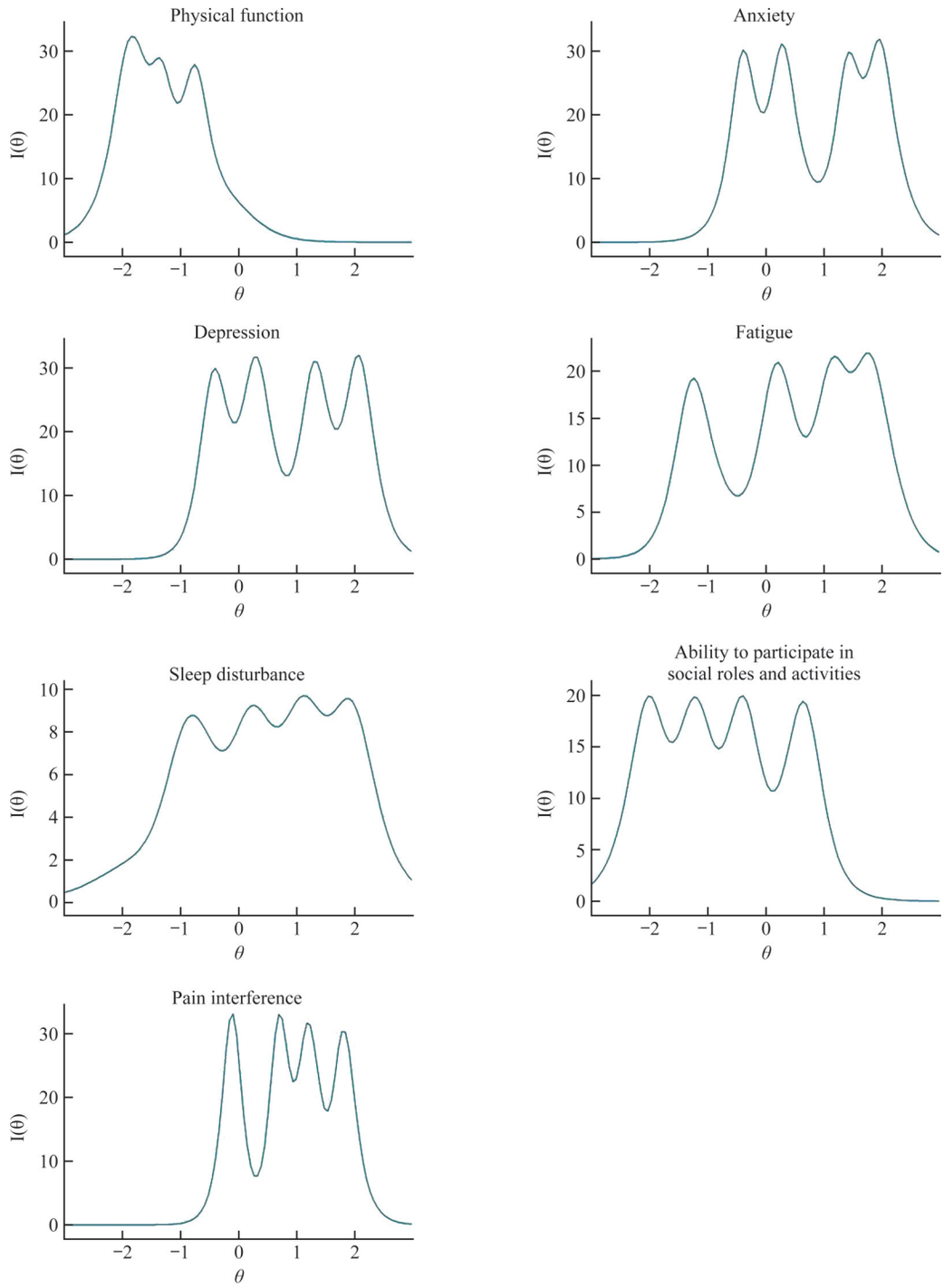



**Figure 2 Test information curves for each domain of PROMIS-29**

## 3 Discussion

This was the first study to evaluate the psychometric performance of the Chinese version of PROMIS-29 in patients undergoing KRT. Our findings indicated sufficient internal consistency, well-performed structural validity, excellent convergent and discriminant validity, acceptable criterion validity, and reliable measurement equivalence of each subscale. Notably, a significant ceiling effect was observed in the physical function domain, while the pain interference domain exhibited a prominent floor effect. Additionally, the analysis of ICCs and TICs provided further valuable insights into the measurement properties.

The mean T-scores of PROMIS-29 revealed that the anxiety and depression were higher compared with

©Journal of Central South University (Medical Science). All rights reserved.

the US general population, highlighting the importance of MH care. Among different types of KRT, patients undergoing hemodialysis exhibited the most severe symptoms, demonstrating the lowest physical function score, the highest levels of fatigue, anxiety, depression, and pain intensity and interference. According to a survey conducted in three major cities in China, there were still 20.2% of patients receiving hemodialysis twice a week, indicating inadequate dialysis frequency which could aggravate related symptoms[42-43]. Patients with longer interdialytic interval manifested higher symptoms frequency and severity[44]. Additionally, the prolonged and frequent hospital visits, rapid fluctuations in osmolality and body fluid levels, as well as the necessity of punctures and tube placement associated with hemodialysis treatment would also contribute to the heightened symptom levels of patients on hemodialysis[4, 45]. In contrast, kidney transplant recipients had the best health-related scores, which were similar to the US general population. Kidney transplantation provides a functional kidney and offers patients the opportunity for a more normal lifestyle, leading to reduced symptom burden.

Overall, all subscales of PROMIS-29 demonstrated acceptable internal consistency, with Cronbach's α ranging from 0.75 to 0.96. Among the 7 domains, the sleep disturbance domain was the only one with a Cronbach's α lower than 0.90, obtaining a value of 0.75, which was consistent with other language versions[17, 46]. This may be attributed to the difficulty in understanding the item Sleep116 (my sleep was refreshing), which responders reflected was hard to understand and answer. This item seeks to understand whether participants feel comfortable and energized after sleeping over the past 7 days, emphasizing the effects of sleep. However, it might be misinterpreted as assessing the refreshing aspect of the sleep itself rather than focusing on one's personal feelings or experiences. The same question also occurred in the studies conducted by Kang, et al[47] on the Korean version and van Kooten, et al[48] on the Dutch version, indicating the term "refreshing" is likely a suboptimal descriptor of sleep quality in this cultural context. Additionally, the factor loading of this item falling below 0.50 and the significant deviation of its ICC from the ideal curve both corroborate the earlier observation. To improve comprehension, cognitive interview is demanded to clarify participants' understanding and adjust the description of this item. To enhance participants' comprehension and ensure accurate responses, a cognitive interview is recommended to explore how participants interpret this item and refine the wording accordingly, thereby improving the reliability of this item. In its current form, this item should be interpreted in conjunction with the overall scale score and related items, rather than used in isolation for clinical decision-making.

The proportion of best score for physical function domain accounts for 42.2% of the whole cohort, manifesting a huge ceiling effect, which may limit the instrument's ability to distinguish individuals with higher levels of PF. Adversely, the pain interference domain was noted significant floor effect of 37.4% of the worst score, making it difficult to differentiate among individuals with minimal pain interference. These findings were similar to the results of a validation study among kidney transplant recipients[19]. Since each subscale of the PROMIS-29 contains only 4 items, they reduce the response burden, but at the same time also limit the depth and breadth of information that can be obtained[49-50]. Another contributing factor could be that all participants were in stable condition and had received dialysis or transplant for at least 3 months, thereby resulting in better performance in the physical function and pain interference domains. In clinical practice, changes in patients with extreme scores may be underestimated. Therefore, this limitation should be considered when using these two domains to monitor treatment response or disease progression in KRT populations.

It is noteworthy that several PROMIS-29 domains showed only moderate correlations with their SF-12 counterparts. Although the SF-12 and PROMIS-29 assess similar health domains, their operational definitions are not fully aligned. For example, the SF-12 VT domain reflects positive states such as "feeling energetic", whereas the PROMIS fatigue domain focuses on negative experiences such as "feeling tired or run-down". In addition, the SF-12 uses only one to 2 items per domain, whereas the PROMIS-29 was developed using IRT with more comprehensive, construct-focused item banks. These differences may

©Journal of Central South University (Medical Science). All rights reserved.

partly explain the attenuated correlations between the 2 instruments. Therefore, instrument selection should be guided by specific research purposes.

Measurement invariance and DIF are crucial in evaluating measures to mitigate bias and ensure comparability across different groups[51-52]. In the current study, DIF was detected only for item PFA21 across education and treatment groups, and no DIF was detected for the remaining items. Similarly, McMullen, et al[53] did not find any DIF in PROMIS-29 items among adult burn survivors across age, gender, education levels and burn size. Nevertheless, owing to the cross-sectional nature of our study design, we were unable to examine DIF over time, which is also crucial for measurement equivalence analysis[54]. Future longitudinal studies are needed to investigate temporal measurement consistency.

As item banks in PROMIS were developed based on IRT, we plotted the ICCs and TICs to provide a more detailed and comprehensive understanding of psychometric properties. Both types of curves were analyzed with GRM, a model proposed by Samejima , et al[55-56] that is suitable for ordered polytomous categories. Most items showed near-ideal shapes in their ICC plots, mirroring findings observed in a similar study involving a Chinese population with breast cancer[19]. Test information was aggregated from the item information of all items within their respective domains. In this study, the test information of PROMIS-29 most subscales ranged from 10 to 30. Rimehaug, et al[57] compared the test information between PROMIS-29 and PROMIS-57, finding that the curves of the shorter profile were under those of the longer one. In previous studies, the test information of 8-item short forms consistently surpassed the threshold of 30[19, 58]. When selecting measurement tools for clinical or research purposes, considerations regarding brevity, validity, and precision of the tool should be thoroughly addressed. Though PROMIS-29 may offer less precision in providing information than PROMIS-57, its satisfactory psychometric properties and lower response burden render it favored in outcomes measurement.

Our study showcased the satisfactory reliability and validity of PROMIS-29 among adult patients with KRT in China. With its coverage of 7 domains closely linked to HRQoL encompassing physical, mental, and social aspects, this tool proves invaluable in assessing HRQoL and related symptoms like fatigue, anxiety, and depression[59]. PROMIS-29 is a generic PROM with fewer items and less response burden, making it preferred for regular surveys in clinical settings. However, it is worth noting that in many instances, a combination of generic and disease-specific tools is essential to furnish more comprehensive information for clinical diagnosis, treatment, and care.

There are several limitations of our study that should be considered. First, we did not assess psychometric properties over time, such as test-retest reliability, time measurement invariance, and responsiveness. Future research employing longitudinal study designs are needed to address these aspects. Second, the standardized T-scores utilized in our study were derived from US norms, as no norms specific to the Chinese general population were available. Third, this study was a single-center study conducted in Hunan Province, which may limit the generalizability of our findings. A multi-center survey should be considered in the future. Fourth, our study employed convenience sampling to recruit participants, which may introduce bias and further limit. However, the inclusion of patients undergoing hemodialysis, peritoneal dialysis, and kidney transplantation extends the applicability of our findings across different treatment modalities.

In conclusion, this study provides evidence endorsing the reliability and validity of the Chinese version of PROMIS-29 in patients undergoing KRT. Combined with its advantages of comparability, simplicity, and precision, we recommend its consideration for assessing HRQoL and symptoms in these populations. Future studies could explore its acceptability, applicability, and utility in collecting clinical outcomes pertinent to CKD.

**Contributions:** ZHAO Qin Conceptualization, formal analysis, writing-original draft; WANG Liang and WANG Yu Investigation, data curation, methodology; ZHAO Hongyu, DONG Lei, and LIU Shan Methodology, validation; ZHU Xiao Methodology, validation, funding acquisition; LUO Haibo Methodology, visualization, writing-review & editing; LIU Jia Conceptualization, funding acquisition, writing- review & editing. The final version of the manuscript has been

©Journal of Central South University (Medical Science). All rights reserved.

read and approved by all authors.

**Conflict of interest:** The authors declare that they have no conflicts of interest to disclose.

## References


[1] Ho WY, Yen CL, Lee CC, et al. Use of fibrates is not associated with reduced risks of mortality or cardiovascular events among ESRD patients: a national cohort study[J]. Front Cardiovasc Med, 2022, 9: 907539. https://doi.org/10.3389/fcvm.2022.907539.

[2] National Kidney Foundation. Kidney failure (ESRD): Symptoms, stages, and treatment[EB/OL]. (2023-09-05)[2024-12-20]. https://www.kidney.org/atoz/content/KidneyFailure.

[3] Ma X, Tao M, Hu Y, et al. Clinical outcomes, quality of life, and costs evaluation of peritoneal dialysis management models in Shanghai Songjiang District: a multi-center and prospective cohort study[J]. Ren Fail, 2021, 43(1): 754-765. https://doi.org/10.1080/0886022X.2021.1918164.

[4] Bossola M, Hedayati SS, Brys ADH, et al. Fatigue in patients receiving maintenance hemodialysis: a review[J]. Am J Kidney Dis, 2023, 82(4): 464-480. https://doi.org/10.1053/j.ajkd.2023.02.008.

[5] Kalantar-Zadeh K, Lockwood MB, Rhee CM, et al. Patient-centred approaches for the management of unpleasant symptoms in kidney disease[J]. Nat Rev Nephrol, 2022, 18(3): 185-198. https://doi.org/10.1038/s41581-021-00518-z.

[6] Fletcher BR, Damery S, Aiyegbusi OL, et al. Symptom burden and health-related quality of life in chronic kidney disease: a global systematic review and meta-analysis[J/OL]. PLoS Med, 2022, 19(4): e1003954[2025-04-11]. https://doi.org/10.1371/journal.pmed.1003954.

[7] Cukor D, Donahue S, Tummalapalli SL, et al. Anxiety, comorbid depression, and dialysis symptom burden[J]. Clin J Am Soc Nephrol, 2022, 17(8): 1216-1217. https://doi.org/10.2215/cjn.01210122.

[8] Mehrotra R, Davison SN, Farrington K, et al. Managing the symptom burden associated with maintenance dialysis: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference[J]. Kidney Int, 2023, 104(3): 441-454. https://doi.org/10.1016/j.kint.2023.05.019.

[9] Duncanson E, Bennett PN, Viecelli A, et al. Feasibility and acceptability of e-PROMs data capture and feedback among patients receiving haemodialysis in the Symptom monitoring WIth Feedback Trial (SWIFT) pilot: protocol for a qualitative study in Australia[J]. BMJ Open, 2020, 10(11): e039014. https://doi.org/10.1136/bmjopen-2020-039014.

[10] Churruca K, Pomare C, Ellis LA, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues[J]. Health Expect, 2021, 24(4): 1015-1024. https://doi.org/10.1111/hex.13254.

[11] van der Willik EM, Hemmelder MH, Bart HAJ, et al. Routinely measuring symptom burden and health-related quality of life in dialysis patients: first results from the Dutch registry of patient-reported outcome measures[J]. Clin Kidney J, 2021, 14(6): 1535-1544. https://doi.org/10.1093/ckj/sfz192.

[12] Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS) [J]. Med Care, 2007, 45(5 Suppl 1): S22-S31. https://doi.org/10.1097/01.mlr.0000250483.85507.04.

[13] Rossi MJ, Sheean AJ, Cote MP, et al. The patient-reported outcomes measurement information system (PROMIS): can we finally compare apples to oranges? [J]. Arthroscopy, 2020, 36(5): 1215-1217. https://doi.org/10.1016/j.arthro.2020.03.001.

[14] Siembida EJ, Reeve BB, Zebrack BJ, et al. Measuring health-related quality of life in adolescent and young adult cancer survivors with the National Institutes of Health Patient-Reported Outcomes Measurement Information System®: Comparing adolescent, emerging adult, and young adult survivor perspectives[J]. Psychooncology, 2021, 30(3): 303-311. https://doi.org/10.1002/pon.5577.

[15] Verberne WR, Das-Gupta Z, Allegretti AS, et al. Development of an international standard set of value-based outcome measures for patients with chronic kidney disease: a report of the international consortium for health outcomes measurement (ICHOM) CKD working group[J]. Am J Kidney Dis, 2019, 73(3): 372-384. https://doi.org/10.1053/j.ajkd.2018.10.007.

[16] Rawang P, Janwantanakul P, Correia H, et al. Cross-cultural adaptation, reliability, and construct validity of the Thai version of the Patient-Reported Outcomes Measurement Information System-29 in individuals with chronic low back pain[J]. Qual Life Res, 2020, 29(3): 793-803. https://doi.org/10.1007/s11136-019-02363-x.

[17] Elsman EBM, Roorda LD, Smidt N, et al. Measurement properties of the Dutch PROMIS-29 v2.1 profile in people with and without chronic conditions[J]. Qual Life Res, 2022, 31(12): 3447-3458. https://doi.org/10.1007/s11136-022-03171-6.

[18] Tang E, Ekundayo O, Peipert JD, et al. Validation of the Patient-Reported Outcomes Measurement Information System (PROMIS)-57 and -29 item short forms among kidney transplant recipients[J]. Qual Life Res, 2019, 28(3): 815-827. https://doi.org/10.1007/s11136-018-2058-2.

[19] Cai T, Wu F, Huang Q, et al. Validity and reliability of the Chinese version of the Patient-Reported Outcomes Measurement Information System adult profile-57 (PROMIS-57)[J]. Health Qual Life Outcomes, 2022, 20(1): 95. https://doi.org/10.1186/s12955-022-01997-9.

[20] Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN Study Design checklist for patient-reported outcome measurement instruments[EB/OL]. (2019-07-01) [2024-12-20]. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_


©Journal of Central South University (Medical Science). All rights reserved.

final.pdf.

[21] Kaplan RM, Hays RD. Health-related quality of life measurement in public health[J]. Annu Rev Public Health, 2022, 43: 355-373. https://doi.org/10.1146/annurev-publhealth-052120-012811.

[22] Aghdaee M, Gu YY, Sinha K, et al. Mapping the patient-reported outcomes measurement information system (PROMIS-29) to EQ-5D-5L[J]. PharmacoEconomics, 2023, 41(2): 187-198. https://doi.org/10.1007/s40273-022-01157-3.

[23] Yadegari I, Bohm E, Ayilara OF, et al. Differential item functioning of the SF-12 in a population-based regional joint replacement registry[J]. Health Qual Life Outcomes, 2019, 17 (1): 114. https://doi.org/10.1186/s12955-019-1166-1.

[24] Lin Y, Yu Y, Zeng J, et al. Comparing the reliability and validity of the SF-36 and SF-12 in measuring quality of life among adolescents in China: a large sample cross-sectional study[J]. Heal Qual Life Outcomes, 2020, 18(1): 360. https://doi.org/10.1186/s12955-020-01605-8.

[25] Shou J, Ren LM, Wang HT, et al. Reliability and validity of 12-item Short-Form health survey (SF-12) for the health status of Chinese community elderly population in Xujiahui district of Shanghai[J]. Aging Clin Exp Res, 2016, 28(2): 339-346. https://doi.org/10.1007/s40520-015-0401-9.

[26] Stocker R, Tran T, Hammarberg K, et al. Patient Health Questionnaire 9 (PHQ-9) and General Anxiety Disorder 7 (GAD-7) data contributed by 13, 829 respondents to a national survey about COVID-19 restrictions in Australia[J]. Psychiatry Res, 2021, 298: 113792. https://doi.org/10.1016/j.psychres.2021.113792.

[27] Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population[J]. Gen Hosp Psychiatry, 2014, 36(5): 539-544. https://doi.org/10.1016/j.genhosppsych.2014.05.021.

[28] Saunders R, Moinian D, Stott J, et al. Measurement invariance of the PHQ-9 and GAD-7 across males and females seeking treatment for common mental health disorders[J]. BMC Psychiatry, 2023, 23(1): 298. https://doi.org/10.1186/s12888-023-04804-x.

[29] Shih YC, Chou CC, Lu YJ, et al. Reliability and validity of the traditional Chinese version of the GAD-7 in Taiwanese patients with epilepsy[J]. J Formos Med Assoc, 2022, 121(11): 2324-2330. https://doi.org/10.1016/j.jfma.2022.04.018.

[30] Tavakol M, Dennick R. Making sense of cronbach's alpha[J]. Int J Med Educ, 2011, 2: 53-55. https://doi.org/10.5116/ijme.4dfb.8dfd.

[31] Katz P, Pedro S, Michaud K. Performance of the patient-reported outcomes measurement information system 29-item profile in rheumatoid arthritis, osteoarthritis, fibromyalgia, and systemic lupus erythematosus[J]. Arthritis Care Res, 2017, 69 (9): 1312-1321. https://doi.org/10.1002/acr.23183.

[32] Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives [J]. Struct Equ Model A Multidiscip J, 1999, 6(1): 1-55. https://doi.org/10.1080/10705519909540118.

[33] Tavassolmand SS, Montazeri A, Madadizadeh F, et al. Translation and validation of the Persian version of diabetic foot ulcer scaleshort form (DFS-SF)[J]. Int Wound J, 2023, 20 (3): 822-830. https://doi.org/10.1111/iwj.13929.

[34] Xu S, Chen Y, Li Y, et al. Factor structure and psychometric properties of the affective lability scale-short form in Chinese adolescents[J]. Front Psychiatry, 2022, 13: 881541. https://doi.org/10.3389/fpsyt.2022.881541.

[35] Fan CW, Huang PC, Chen IH, et al. Differential item functioning for the Tendency of Avoiding Physical Activity and Sport Scale across two subculture samples: Taiwanese and mainland Chinese university students[J/OL]. Heliyon, 2023, 9 (12): e22583[20225-04-01]. https://doi. org/10.1016/j. heliyon.2023.e22583.

[36] Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations[J]. J Stat Softw, 2011, 39(8): 1-30. https://doi.org/10.18637/jss.v039.i08.

[37] PROMIS ® Minimum requirements for the release of PROMIS instruments after translation and recommendations for further psychometric evaluation, Version: March 3, 2014. PROMIS Statistical Center[EB/OL]. (2014-03-03)[2025-01-03]. http://www.healthmeasures. net/images/PROMIS/Standards_for_release_of_PROMIS_instruments_after_translation_v8.pdf.

[38] Wolf EJ, Ellickson-Larew S, Guetta RE, et al. Psychometric performance of the miller forensic assessment of symptoms test (M-FAST) in veteran PTSD assessment[J]. Psychol Inj Law, 2020, 13(3): 284-302. https://doi.org/10.1007/s12207-020-09373-y.

[39] Park K, Jaekal E, Yoon S, et al. Diagnostic utility and psychometric properties of the beck depression inventory-II among Korean adults[J]. Front Psychol, 2019, 10: 2934. https://doi.org/10.3389/fpsyg.2019.02934.

[40] Chalmers RP. mirt: a multidimensional item response theory package for the*R*Environment[J]. J Stat Soft, 2012, 48(6): 1-29. https://doi.org/10.18637/jss.v048.i06.

[41] Gaskin J, James M, Lim J. "Master Validity Tool", AMOS Plugin. Gaskination's StatWiki[EB/OL. (2019-09-19)[2025-04-01]. http://statwiki.gaskination.com/.

[42] Zhao X, Niu Q, Gan L, et al. Baseline data report of the China dialysis outcomes and practice patterns study (DOPPS)[J]. Sci Rep, 2021, 11(1): 873. https://doi.org/10.1038/s41598-020-79531-4.

[43] Htay H, Bello AK, Levin A, et al. Hemodialysis use and practice patterns: an international survey study[J]. Am J Kidney Dis, 2021, 77(3): 326-335. https://doi. org/10.1053/j.ajkd.2020.05.030.

[44] Chauhan K, Wen HH, Gupta N, et al. Higher symptom frequency and severity after the long interdialytic interval in patients on maintenance intermittent hemodialysis[J]. Kidney Int Rep, 2022, 7(12): 2630-2638. https://doi.org/10.1016/j.ekir.2022.09.032.

[45] Kassim MAK, Pantazi AC, Nori W, et al. Non-pharmacological

©Journal of Central South University (Medical Science). All rights reserved.

interventions for pain management in hemodialysis: a narrative review[J]. J Clin Med, 2023, 12(16): 5390. https://doi.org/10.3390/jcm12165390.

[46] Kang D, Lim J, Kim BG, et al. Psychometric validation of the Korean Patient-Reported Outcome Measurement Information System (PROMIS)-29 Profile V2.1 among patients with chronic pulmonary diseases[J]. J Thorac Dis, 2021, 13(10): 5752-5764. https://doi.org/10.21037/jtd-21-591.

[47] Kang D, Kim Y, Lim J, et al. Validation of the Korean version of the patient-reported outcomes measurement information system 29 profile V2.1 among cancer survivors[J]. Cancer Res Treat, 2022, 54(1): 10-19. https://doi.org/10.4143/crt.2020.1200.

[48] van Kooten JAMC, Terwee CB, Kaspers GJL, et al. Content validity of the Patient-Reported Outcomes Measurement Information System Sleep Disturbance and Sleep Related Impairment item banks in adolescents[J]. Health Qual Life Outcomes, 2016, 14(1): 92. https://doi.org/10.1186/s12955-016-0496-5.

[49] Kruyen PM, Emons WHM, Sijtsma K. Assessing individual change using short tests and questionnaires[J]. Appl Psychol Meas, 2014, 38(3): 201-216. https://doi.org/10.1177/0146621613510061.

[50] Pecorelli N, Guarneri G, Vallorani A, et al. Validation of the PROMIS-29 questionnaire as a measure of recovery after pancreatic surgery[J]. Ann Surg, 2023, 278(5): 732-739. https://doi.org/10.1097/SLA.0000000000006020.

[51] Belzak WCM, Bauer DJ. Improving the assessment of measurement invariance: Using regularization to select anchor items and identify differential item functioning[J]. Psychol Methods, 2020, 25(6): 673-690. https://doi.org/10.1037/met0000253.

[52] Bauer DJ. A more general model for testing measurement invariance and differential item functioning[J]. Psychol Methods, 2017, 22(3): 507-526. https://doi.org/10.1037/met0000077.

[53] McMullen K, Bamer A, Ryan CM, et al. Validation of PROMIS-29 domain scores among adult burn survivors: a National Institute on Disability, Independent Living, and Rehabilitation Research Burn Model System Study[J]. J Trauma Acute Care Surg, 2022, 92(1): 213-222. https://doi.org/10.1097/TA.0000000000003365.

[54] Arenas E, Teruel G, Gaitán-Rossi P. Time and gender measurement invariance in the modified Calderon depression scale[J]. Health Qual Life Outcomes, 2022, 20(1): 100. https://doi.org/10.1186/s12955-022-02007-8.

[55] Samejima F. Estimation of latent ability using a response pattern of graded scores[J]. Psychometrika, 1969, 34(1): 1-97. https://doi.org/10.1007/BF03372160.

[56] Samejima F. Graded response model[M]//Van Der Linden WJ, Hambleton RK. Handbook of Modern Item Response Theory. New York, NY: Springer New York, 1997: 85-100. https://doi.org/10. 1007/978-1-4757-2691-6_5.

[57] Rimehaug SA, Kaat AJ, Nordvik JE, et al. Psychometric properties of the PROMIS-57 questionnaire, Norwegian version [J]. Qual Life Res, 2022, 31(1): 269-280. https://doi.org/10.1007/s11136-021-02906-1.

[58] Zhou T, Wang Y, Chen J, et al. Psychometric properties of the Chinese version of the PROMIS-Cancer-Anxiety item bank assessed using a graded response model[J]. Asia Pac J Oncol Nurs, 2023, 10(12): 100312. https://doi.org/10.1016/j.apjon.2023.100312.

[59] Cella D, Choi SW, Condon DM, et al. PROMIS® adult health profiles: efficient short-form measures of seven health domains [J]. Value Health, 2019, 22(5): 537-544. https://doi.org/10.1016/j.jval.2019.02.004.

(Edited by PENG Minning)

©Journal of Central South University (Medical Science). All rights reserved.