Clinical Kidney Journal, 2025, vol. 18, no. 5, sfaf112

https://doi.org/10.1093/ckj/sfaf112
Advance Access Publication Date: 24 April 2025
Original Article

ORIGINAL ARTICLE

Development, Rasch analysis and validation of the kidney symptom burden questionnaire (KSB-Q)

Derek Kyte [1,2,*], Benjamin R. Fletcher [2,*], Mike Horton [3], Sarah Damery [4], Olalekan Lee Aiyegbusi [2,4,5,6], Nicola Anderson [2,4,5], Andrew Bissell [7], Melanie Calvert [2,4,5,6], Paul Cockwell [2,8,9], James Ferguson [2,5], Muirne C.S. Paap [10], Chris Sidey-Gibbons [11], Neil Turner [12], Rav Verdi [7] and Anita Slade [2,5]

[1]Department of Allied Health, School of Health & Wellbeing, University of Worcester, Worcester, UK, [2]Centre for Patient-Reported Outcomes Research, Institute for Applied Health Research, University of Birmingham, Birmingham, UK, [3]Leeds Psychometric Laboratory for Health Sciences, University of Leeds, Leeds, UK, [4]National Institute for Health and Care Research Applied Research Collaboration West Midlands, University of Birmingham, Birmingham, UK, [5]National Institute for Health and Care Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, UK, [6]National Institute for Health and Care Research Blood and Transplant Research Unit in Precision and Cellular Therapeutics, [7]Kidney Patient Advisory Group, Centre for Patient-Reported Outcomes Research, Institute for Applied Health Research, University of Birmingham, Birmingham, UK, [8]Department of Renal Medicine, Queen Elizabeth Hospital Birmingham, University Hospitals Birmingham, Birmingham, UK, [9]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK, [10]Department of Child and Family Welfare, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands, [11]MD Anderson Center for INSPiRED Cancer Care, University of Texas, TX, USA and [12]Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK

*Joint first authors
Correspondence to: Derek Kyte; E-mail: d.kyte@worc.ac.uk

ABSTRACT

Background. Increasingly, patient-reported outcome measures (PROMs) are used to monitor chronic kidney disease (CKD) symptoms in routine clinical practice. However, such symptom measurement currently requires completion of multiple, often lengthy, PROMs, which may lead to questionnaire fatigue, lower levels of completion, and missing data. Moreover, many CKD-specific PROMs lack evidence of important measurement properties and few were developed using contemporary psychometric methods. The study objective was to develop and validate a short-form kidney symptom burden questionnaire (KSB-Q).

Methods. A cross-sectional item pool survey was distributed to adults (≥18 years) with CKD stages 3–5 [including individuals not receiving kidney replacement therapy (KRT), those receiving dialysis and those with a functioning kidney transplant] in England (Birmingham, London, Sheffield, and Nottingham) from March to September 2022. Rasch measurement was used to assess the psychometric properties of the item pool. Cognitive debriefing interviews were conducted to evaluate content validity.

Received: 30.9.2024; Editorial decision: 3.4.2025

Results. In total, 419/1464 participants responded (29% response rate), with 28% receiving dialysis and 30% in receipt of a kidney transplant. Rasch analysis indicated that nine items, one for each of nine symptom domains (fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs, and shortness of breath), formed a PROM with strong psychometric properties (including statistically acceptable reliability, test–retest reliability, and validity). Cognitive debriefing and survey responses confirmed content validity encompassing relevance, comprehensiveness, and clarity.
Conclusions. The KSB-Q is a nine-item PROM measuring somatic symptoms. The KSB-Q demonstrates strong psychometric properties in patients with CKD stages 3–5, including those not receiving KRT, individuals receiving dialysis, and those with a functioning kidney transplant.

Keywords: chronic kidney disease, measurement properties, patient-reported outcomes, Rasch analysis

KEY LEARNING POINTS

What was known:

- Measurement of chronic kidney disease (CKD) symptoms currently requires completion of multiple, often lengthy, patient-reported outcome (PRO) questionnaires.
- The study objective was to develop and validate a single short kidney symptom burden PRO questionnaire: the Kidney Symptom Burden Questionnaire (KSB-Q).

This study adds:

- We surveyed 419 people in the UK with kidney disease.
- Data analysis helped us to develop a validated nine-item KSB-Q covering: fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs, and shortness of breath.

Potential impact:

- The KSB-Q is a short/easy to complete questionnaire that can be used to capture information about patient-important symptoms in individuals with CKD who are not receiving kidney replacement therapy, those who are receiving dialysis, and those who have a functioning kidney transplant.

INTRODUCTION

Chronic kidney disease (CKD) is a global health concern [1] with a high prevalence [2] and substantial healthcare and societal costs [3, 4, 5]. Patients with CKD commonly experience considerable somatic symptom burden [6] and associated poorer long-term health-related quality of life [7] and mortality [8,9]. The exact combination of symptoms differs across recognized clinical groups, including patients not on kidney replacement therapy (KRT), those receiving dialysis, or kidney transplant recipients [6]. Thus, granular measurement of individual symptoms is an important tool to support value-based person-centred care [10], helping to target the symptoms that matter most to individuals and optimizing patient benefit [11].

Increasingly, routine clinical measurement of symptoms in CKD populations is undertaken using patient-reported outcome measures (PROMs). These questionnaires, traditionally used in research, gather valuable data directly from the patient about their lived experience of a health condition [12]. This data can be collected in clinical practice to augment clinical data, build a longitudinal picture of disease burden and improve care [11]. Routine clinical PROM-based symptom monitoring has demonstrated efficacy in patients with cancer [13, 14, 15] and feasibility has been established in patients with advanced CKD [16]. Currently, studies are ongoing to explore effectiveness in patients receiving dialysis [17,18].

However, there remains heterogeneity in the PROMs used to measure symptoms in CKD. A recent systematic review and meta-analysis identified that 54 different PROMs were used to collect data on symptoms across the included studies, with no single tool measuring >45% of symptoms reported in the population [6]. In addition, few CKD-specific PROMs appear to have demonstrated adequate item targeting alignment in a broad sample population including those not receiving KRT, individuals receiving dialysis and those with a functioning kidney transplant. Thus, measurement of the symptoms that matter most to patients with CKD currently requires the use of multiple, often lengthy, PROMs This may lead to questionnaire burden/fatigue amongst patients and staff, a widely recognized threat to PRO data collection that may lead to missing data [19,20]. Moreover, many commonly used PROMs that measure CKD symptoms lack evidence of important measurement properties, and few were developed using current guidelines or contemporary psychometric methods such as item response theory [21,22].

In this report, we present the results of a study undertaken to develop a psychometrically robust short-form PRO tool measuring somatic symptoms in patients with CKD: the Kidney Symptom Burden Questionnaire (KSB-Q).

MATERIALS AND METHODS

Setting and study design

This two-phase PROM development study (Fig. 1) was undertaken between October 2019 and September 2022 (UK National Health Research Authority ethical approval, 20/WM/0182). Recruitment was conducted at four research sites (all UK NHS trust secondary care providers): Birmingham, Nottingham, London, and Sheffield. In this paper, we focus on the analysis and development of the phase 1 project output: the KSB-Q. The phase

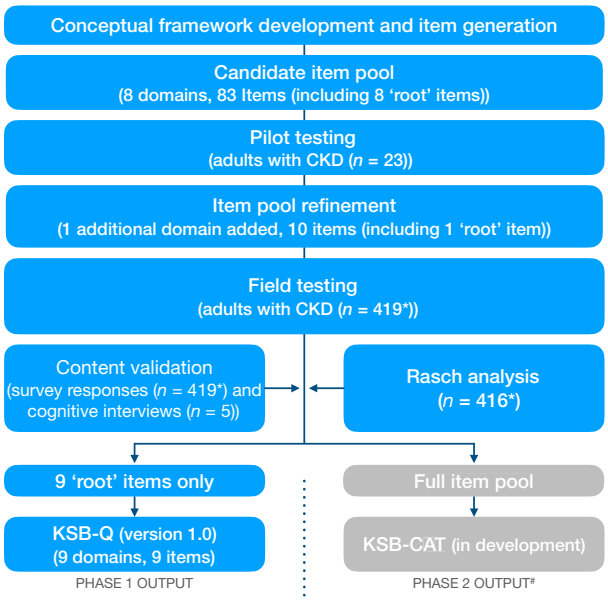

Figure 1: Development of the renal symptom burden questionnaire (KSB-Q). Adults with CKD (stages 3–5) included patients: (i) not receiving KRT, (ii) receiving dialysis, and (iii) with a functioning kidney transplant. *Includes internal pilot data. #KSB-CAT development to be reported in a future paper.

2 project output, a multi-symptom renal computer adaptive test (CAT) [23], is in development and will be reported in a future paper.

Study population

The field test population comprised adults (≥18 years) with stage 3–5 CKD from one of three clinical groups: (i) not receiving KRT, (ii) receiving dialysis, and (iii) with a functioning kidney transplant. Individuals were excluded if they were unable to speak, read, or write English, or if they had experienced an episode of acute kidney injury in the previous 3 months. The minimum sample size was 250 respondents, which provides 99% confidence that Rasch item calibrations/person measures are stable within ±0.50 logits [24]. Recruitment to the cognitive debriefing interviews, using the same eligibility criteria, was conducted with patients at University Hospitals Birmingham NHS Foundation Trust.

Conceptual framework development and item generation

The conceptual framework for the candidate item pool was developed following a global systematic review and meta-analysis of symptom burden and health-related quality of life in CKD [6]. The review analysed 449 symptom prevalence/severity studies, including 199 147 participants from 62 countries. Somatic symptoms identified in the review were mapped to the World Health Organization International Classification of Functioning, Disability and Health (ICF) [25] and cross-mapped against published qualitative literature and established CKD core outcome sets including SONG (Standardized Outcomes in Nephrology) [26, 27, 28] and ICHOM (International Consortium for Health Outcomes Measurement) [29].

The results of the mapping exercise were reviewed by the study management group, which included patients ($n = 2$), clinicians ($n = 4$), and psychometricians/outcome methodologists ($n = 4$). The group selected somatic symptoms for inclusion in the candidate item pool that: (i) showed evidence of prevalence/severity in our global systematic review [6], (ii) were supported by qualitative evidence, and (iii) were included in established CKD core outcome sets. Symptom domains meeting these three requirements were included in the candidate item pool, and those not meeting one or more elements were excluded. Full details of the symptom domains that were reviewed, along with the summary data used in the decision-making process, are included in the Supplementary Appendix. The final selected item pool domains are reported in Table 1.

To develop the additional domain items, we reviewed and adapted relevant items extracted from the questionnaires identified in our global systematic review and meta-analysis [6]. This process involved a thorough examination of existing PROMs to ensure that the items selected were both comprehensive and relevant to the CKD population. By integrating these adapted items, we aimed to enhance the robustness and applicability of the KSB-Q, ensuring it efficiently captured a range of symptom experiences reported by patients.

Table 1: Candidate item pool symptom domains and definitions.

Candidate item pool symptom domain	Definition[b]
Fatigue	A feeling of tiredness, or lack of energy, often linked with low motivation.
Pain	Painful muscle cramps, bone or joint pain, or general pain.
Memory/concentration	Problems with memory or concentration.
Poor sleep	Difficulty falling asleep and/or difficulties staying asleep.
Skin problems	Dryness and itchiness of the skin, or changes in skin appearance.
Gastrointestinal problems	Problems with the stomach and digestion, e.g. nausea, vomiting, or constipation.
Dizziness	Feeling faint, woozy, weak, or unsteady.
Restless legs	An overwhelming and uncomfortable urge to move one's legs. Or an unpleasant crawling or creeping sensation in the feet, calves, and thighs.
Shortness of breath[a]	A sensation of being unable to breathe normally or feeling of breathlessness.

[a]Shortness of breath domain added following pilot phase.
[b]Definitions presented to participants during the item pool survey.

Study procedures

During field testing, each research site screened clinic lists for potentially eligible patients, before distributing a paper version of the item pool survey (Supplementary Appendix). Each survey included a unique ID number. Respondents either returned the paper survey to the research team using a pre-paid envelope, or submitted the survey online using an optional QR-code/web address. An internal pilot was conducted in March 2022.

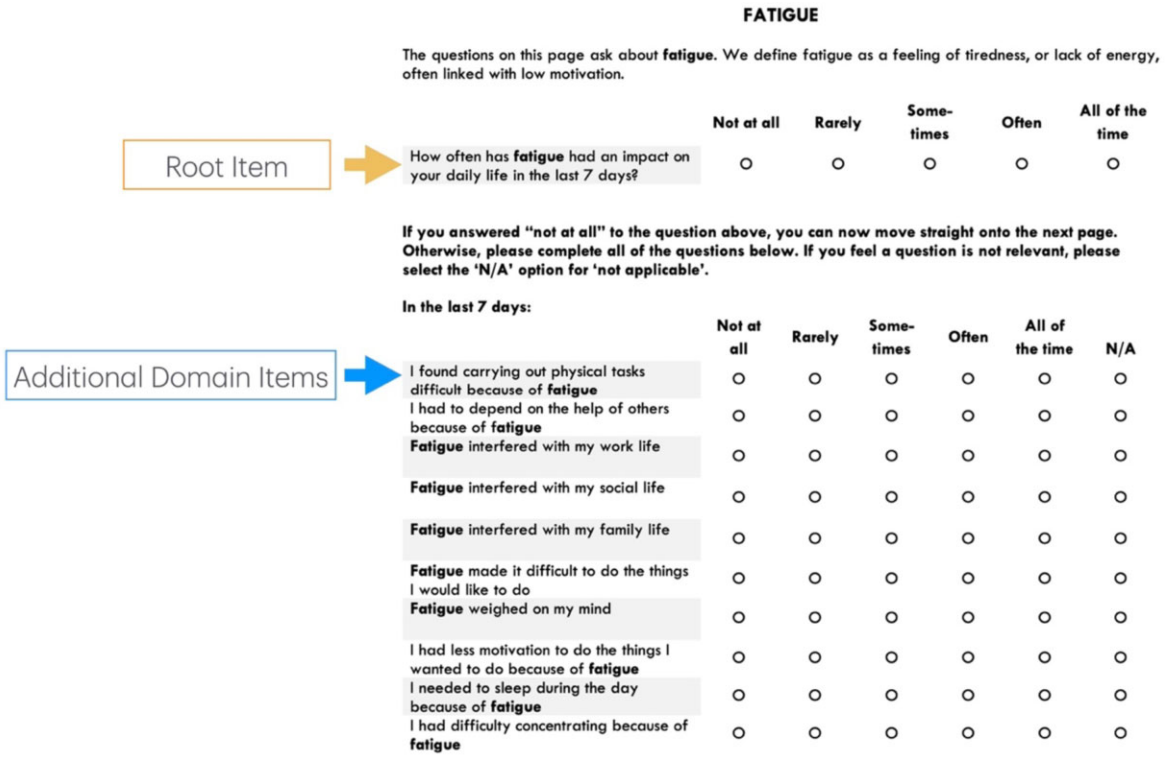

Figure 2: Example item pool survey page.

Response options for each of the candidate items were: 'not at all' (least burden), 'rarely', 'sometimes', 'often', and 'all of the time' (most burden). We chose a 7-day recall period, reasoning that this would: (i) reduce the likelihood of recall bias, (ii) improve the consistency of patient responses [30], and (iii) capture recent changes in the patient's condition or treatment effects, making the data potentially more relevant for clinical decision-making and monitoring. Each symptom domain in the survey included an initial 'root' item, for example, 'How often has fatigue had an impact on your daily life in the last 7 days?' (see example in Fig. 2). Those respondents selecting the 'not at all' response option to this root item were directed to move onto the next domain. Otherwise, respondents were asked to answer all the other items within the current section. The nine root items were used to develop the KSB-Q reported in this paper. Additional items in each domain will be used in the development of the KSB-CAT (phase 2 of the project, to be reported in a future paper).

The survey also included demographic questions, additional items regarding the content validity of the item pool and, er to explore concurrent validity, a copy of the Integrated Palliative Outcome Score (IPOS) renal survey [31]. The IPOS-Renal is a commonly used PROM that measures symptoms and other concerns (e.g. information needs, practical issues, family anxiety etc.) [31].

The ID numbers of returned surveys were provided to each research site to link participants' estimated glomerular filtration rate (eGFR) data (closest to survey receipt). To establish test–retest reliability, those respondents consenting to receive a 'time 2' survey were posted a second version at least 2 weeks following the first. This version included a global item to establish the participants' current CKD status and capture any perceived changes: 'Thinking about your kidney disease, how do you feel now?': 'A lot worse', 'A little worse', 'About the same', A little better', and 'A lot better'. Test–retest reliability was calculated for those patients who selected the option 'About the same', using an intraclass correlation (ICC) two-way mixed-effect analysis of variance model, with interaction for the absolute agreement between single scores [32]. A test–retest correlation value $>.70$ is recommended as a minimum standard for reliability [33].

Rasch analysis

Rasch analysis was conducted on the short-form KSB-Q, constructed from the nine root domain-level items only (Table 2). Rasch Measurement Theory [34] provides a unified, unidimensional confirmatory framework to assess multi-item latent scales for several aspects of internal construct validity, ensuring that it is valid to add the items together to form an overall total score and highlighting any measurement anomalies within a multi-item scale [35, 36]. Rasch analysis was completed using RUMM2030 software to test whether each item uniquely contributed to the underlying construct of kidney symptom burden [37].

The following elements were assessed:

- **Local dependence**: which refers to the situation where responses to certain items are more closely related to each other than is explained by the underlying trait being measured. Items were tested for local dependency, where the residual correlation Q3 criterion cut point to indicate dependency was taken as 0.2 above the average residual correlation [38].
- **Individual item fit**: each item was assessed for its fit to the Rasch model using chi-square statistics and standardized fit residuals (non-significant at Bonferroni-adjusted chi-square P value, standardized (z-score) fit residuals within ± 2.5) [36].

Table 2: Item pool 'Root' items subjected to Rasch analysis.

Domain	Root item	Response options
Fatigue	How often has **fatigue** had an impact on your daily life in the last 7 days?	'not at all' (least burden), 'rarely', 'sometimes', 'often', 'all of the time' (most burden)
Pain	How often has **pain** had an impact on your daily life in the last 7 days?	
Memory/concentration	How often have **problems with memory or concentration** had an impact on your daily life in the last 7 days?	
Poor sleep	How often has **poor sleep** had an impact on your daily life in the last 7 days?	
Skin problems	How often have **skin problems** had an impact on your daily life in the last 7 days?	
Gastrointestinal problems	How often have **gastrointestinal problems** had an impact on your daily life in the last 7 days?	
Dizziness	How often has **dizziness** had an impact on your daily life in the last 7 days?	
Restless legs	How often have **restless legs** had an impact on your daily life in the last 7 days?	
Shortness of breath	How often has **shortness of breath** had an impact on your daily life in the last 7 days?	

Items that do not fit well may indicate that they are not measuring the same underlying construct as other items.

- **Response category functioning**: was assessed to ensure that item response categories were operating as intended, where thresholds (the crossover points between adjacent response categories) should progressively increase across the underlying trait, to represent distinctly separate response categories.
- **Differential item functioning (DIF)**: was used to check for item bias across different subgroups (age group (18–30; 31–40; 41–50; 51–60; 61–70; 71+), sex (male; female; unspecified), and ethnicity (White; Asian; Black; Other)). Items should perform consistently across these groups (DIF non-significant at Bonferroni-adjusted ANOVA *P* value) to ensure the scale's fairness and stability.
- **Visual scale targeting**: was used to compare the relative distribution of item locations with the distribution of person locations to ensure that the scale was well-targeted to the sample. Good targeting means that the items are well-aligned with the respondents in terms of the symptom burden expressed. We also reviewed floor and ceiling effects to ensure alignment coverage.
- **Person estimates**: the symptom burden levels of individuals (theta) were estimated using the Rasch model. These estimates, expressed in logits, represent the latent trait being measured. Higher theta values indicate higher levels of symptom burden. These estimates were used to assess the distribution of person abilities and to ensure that the scale was well-targeted to the sample.
- **Reliability indices**: reliability was assessed using the Person Separation Index (PSI) and Cronbach's alpha [36]. These indices indicate the internal consistency of the scale and its capacity to reliably order the people that it is measuring.
- **Unidimensionality**: the assumption of unidimensionality [39] was assessed using a series of *t*-tests (confirmed where the person estimates did not differ in more than 5% of cases, with a lower bound 95% confidence interval applied). Unidimensionality confirms that the items measure a single underlying construct, which is essential for the validity of the total score. When the assumptions of the Rasch model are satisfied, the sufficiency of the raw score allows for a linear, interval-level transformation of scores [36].

For all individuals, KSB-Q scale scores corresponded with an interval-level logit value that was extracted from the Rasch analysis software. To aid interpretability, the linear logit values were subsequently converted into 0–100 scale values, where a higher score represents higher symptom burden. These 0–100 transformed metric scores were also used to calculate the standard error of measurement (SEM) and minimal detectable change (MDC) of the KSB-Q. The SEM is calculated with the formula: $= SD \times \sqrt{1-R}$, where SD is the standard deviation of the person estimates, and R is the reliability index of the scale (PSI and alpha estimates are both reported). The MDCs were calculated with the formula: $MDC = SEM \times 1.96 \times \sqrt{2}$ [40]. The MDC is a distribution-based responsiveness indicator based on data from a single timepoint. The MDC value indicates a score change that can be interpreted as a real change (for a person) in the construct that is being measured.

The 0–100 transformed metric scores were also used to establish concurrent validity correlations with both the IPOS-Renal and eGFR values, where a strong correlation was proposed between the KSB-Q and the IPOS-Renal symptom and total scores, a moderate correlation proposed between with the IPOS-Renal non-symptom score, and a weak-moderate correlation proposed between the KSB-Q and eGFR. Correlation values were classified as 0.1–0.39 = weak; 0.4–0.69 = moderate; and >0.7 = strong [41].

Cognitive interviews

Cognitive interviews were conducted by an experienced interviewer with expertise in measure development and Rasch methodology (A.S.). Participants who consented to take part in the cognitive interviews were asked to review a copy of the KSB-Q, to evaluate face validity of the items. During the interview, participants completed the KSB-Q using a think-aloud procedure [42]. The interviews explored understanding and comprehensiveness (the extent to which a PROM captures all relevant aspects of the patient's experience with a health condition), as well as clarity of the design and structure of the questionnaire, including response options, recall periods, and format [22]. Interviews were recorded and analysed using a thematic approach.

Patient and public involvement (PPI)

Throughout the design and development of the project we considered it important to include patient partners. Two PPI members with lived experience of CKD (A.B. and R.V.) joined the study management group as formal members and were involved in the study design (recruitment strategy and patient-facing materials), as well as development of the conceptual framework and KSB-Q domains/items.

RESULTS

Participant characteristics

A total of 1464 surveys were posted out to patients and 419 responses were received (29% response rate). Participant characteristics are summarized in Table 3. The sample included 252 (60.1%) male respondents. The mean eGFR of the sample (those not on dialysis) was 26.10 ml/min 1.73 m^2 (SD 20.66, range 3–90). Most of the people in the sample were over 50 (72.2%), 119 respondents reported receiving dialysis (28.4%), and 127 (30.3%) reported a functioning kidney transplant.

Item pool refinement

Review of the internal pilot data ($n = 23$) indicated that several respondents had queried the absence of items related to shortness of breath. This item was initially excluded during the item generation phase by the study management group as it failed to meet criteria 3, i.e. it had not been included in established CKD core outcome sets. After consultation with the clinical and patient representatives on the study management group, a shortness of breath domain containing 11 items was added to the survey. No other changes were made to the survey before formal field testing.

Content validity

Most survey respondents indicated that the items across the pool were relevant (81.6%, 279/342), comprehensible (97.4%, 335/344), and comprehensive (74.7%, 248/332). Participants involved in cognitive interviews ($n = 5$) did not raise any substantial issues with face validity, response or recall options, the design and clarity of the KSB-Q, or comprehension of the items. None of the participants identified any missing symptoms, although one participant suggested patients who had been recently diagnosed might find the range of issues covered daunting to read. The cognitive interviews suggested no modifications to the KSB-Q were required.

Rasch analysis

Rasch analysis included both pilot and formal field test data. Although data were received from 419 respondents, three records did not contain any information for the item pool, therefore the final sample for analysis was $n = 416$. The KSB-Q subjected to Rasch analysis was formed from the nine root items covering the following domains: fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs, and shortness of breath. These items displayed good fit to the model across almost all assessment criteria (see Table 4 for details). The overall scale fit was good (chi-square $P = .12$), with good reliability (PSI = 0.80, Cronbach's alpha = 0.87), and was unidimensional (2.3% significant $P = .05$ t-tests). The scale was well-targeted, although there was a small floor and ceiling effect, indicating that individuals that lie towards the very extreme ends of the measurement range were not quite as well covered (Fig. 3). There was no apparent item bias (DIF) for sex, age group, or ethnicity, indicating that the scale is stable and unbiased across these groups. Overall, the 'easiest' item to score higher on (lowest location) was the fatigue item, meaning that this was the most commonly reported symptom with the great-

Table 3: Participant characteristics. eGFR: estimated glomerular filtration rate.

	Respondents ($n = 419$)	
Characteristics	Number	Percentage
Age (years)		
18–30	11	2.6
31–40	28	6.7
41–50	56	13.4
51–60	84	20.0
61–70	91	21.7
71+	128	30.5
rather not say	7	1.7
missing	14	3.3
Sex:		
female	153	36.5
male	252	60.1
rather not say	2	0.5
missing	12	2.9
Ethnicity:		
White	295	70.4
Asian or Asian British	42	10.0
Black, African, Caribbean, or Black British	55	13.1
Mixed or multiple ethnic groups	11	2.6
missing	16	3.8
eGFR ml/min 1.73 m^2[a]:		
mean	26.10, range 3–90	SD 20.66
receiving dialysis:		
yes	119	28.4
no	284	67.8
missing	16	3.8
Dialysis type ($n = 119$):		
in-centre	50	42.0
home	19	16.0
peritoneal	13	10.9
missing	37	31.1
Time on dialysis ($n = 119$):		
<1 year	26	21.8
1–2 years	33	27.7
3+ years	57	47.9
missing	3	2.5
In receipt of a kidney transplant:		
yes	127	30.3
no	266	63.5
missing	26	6.2
Transplant type ($n = 127$):		
deceased donor	53	41.7
living donor (relative, friend)	27	21.3
living donor (anonymous)	6	4.7
missing	41	32.3
Time since transplant ($n = 127$):		
<1 year	18	14.2
1–2 years	10	7.9
3+ years	58	45.7
missing	41	32.3

[a]Not collected for patients on dialysis.

Table 4: Summary of psychometric properties for the KSB-Q scale, comparing the original scoring to two rescoring options.

Rasch metric	Description	KSB-Q, initial version	KSB-Q, rescore 1 (response categories 'Rarely' and 'Sometimes' merged)	KSB-Q, rescore 2 (response categories 'Not at all' and 'Rarely' merged)	Target values[a]
total n − (extremes) = valid n	Total sample	416 − (23) = 393	416 − (23) = 393	416 − (41) = 375	
Number of items	No. of questions	9	9	9	
Overall scale fit (Chi-square P)	Fit to the Rasch model	$P = .12$	$P = .11$	$P = .64$	$P > .01$
Individual item fit residuals		1/9 out of range	1/9 out of range	All items within range	within ±2.5
Individual item fit (Chi-square P)		All items fit	All items fit	All items fit	$P > .05$ (Bonferroni adj)
Person fit residuals		3.1% outside range	3.1% outside range	1.3% outside range	within ±3.0
Unidimensionality	Determines if items measure a single underlying construct	2.30%	2.56%	1.61%	percentage of significant t-tests < 5%
Local dependency	Determines the independence of each item	1/36 pairwise significant	1/36 pairwise significant	1/36 pairwise significant	<0.2 above average Q3
Targeting	Explores if the scale and sample distributions are well-aligned	Good alignment	Good alignment	Slightly skewed	aligned
Percentage of the sample at the floor (min score)	Helps determine whether the scale adequately captures the full range of patient outcomes	5.1%	5.1%	9.4%	
Percentage of the sample at the ceiling (max score)		0.5%	0.5%	0.5%	
Percentage of the sample at the floor and ceiling (combined)		5.5%	5.5%	9.9%	<15%
PSI	Reliability indices	0.8	0.8	0.72	>0.85
Cronbach's alpha		0.87	0.83	0.86	>0.85
Response category threshold ordering	Assessed to ensure that item response categories were operating as intended	All disordered	3/9 disordered	2/9 disordered	All ordered
DIF-by-sex	Used to check for item bias across different subgroups	No DIF	No DIF	No DIF	$P > .05$ (Bonferroni adj)
DIF-by-age group		No DIF	No DIF	No DIF	$P > .05$ (Bonferroni adj)
DIF-by-ethnicity		No DIF	No DIF	No DIF	$P > .05$ (Bonferroni adj)
SEM on 0–100 values (based on alpha)	Determines the amount of error in the measurement	5.57	7.51	6.67	Smaller is better
SEM on 0–100 values (based on PSI)		6.90	8.15	9.43	Smaller is better
MDC on 0–100 values (based on alpha)	The smallest change in a PROM score that reflects a real difference, beyond measurement error	15.43	20.82	18.49	Smaller is better
MDC on 0–100 values (based on PSI)		19.14	22.59	26.15	Smaller is better
Test–retest ICC ($n = 48$)	Test–retest Reliability indices	0.82 (0.69–0.89)	0.82 (0.69–0.89)	0.81 (0.69–0.89)	>0.7
Test–retest correlation (Pearson) ($n = 48$)		0.83	0.83	0.82	>0.7
Test–retest correlation (Spearman) ($n = 48$)		0.88	0.85	0.88	>0.7

[a]Target values taken from [33].

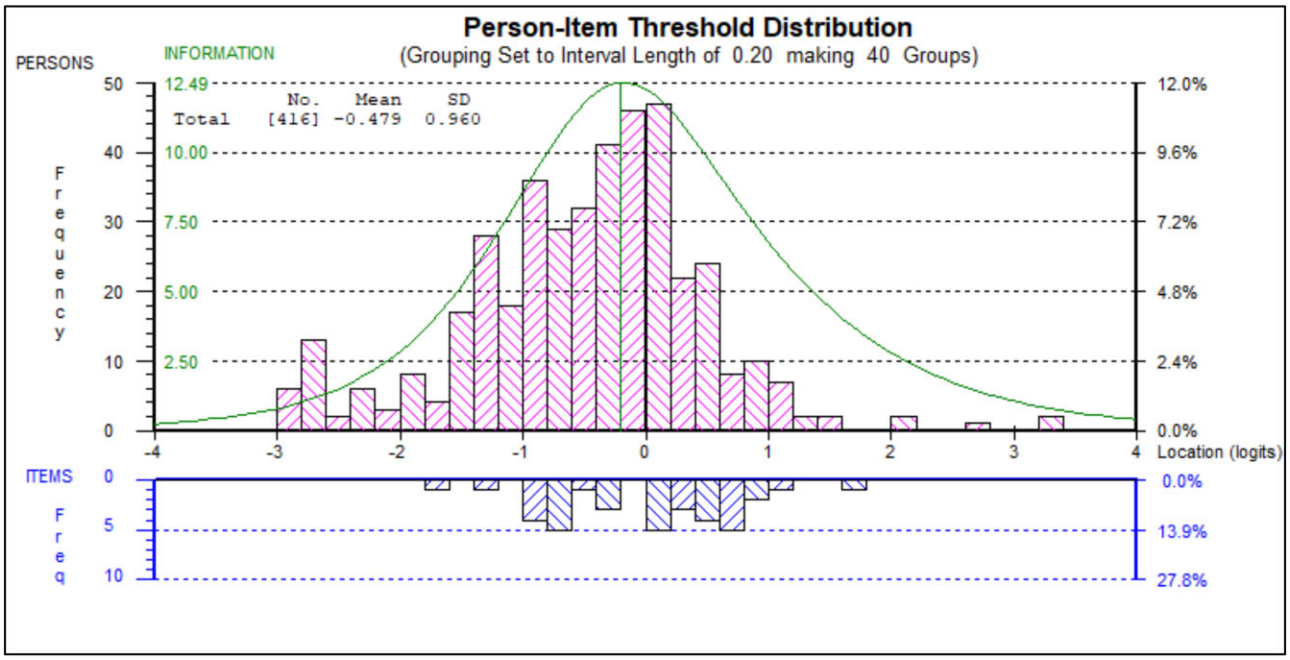

Figure 3: Targeting plot showing the relative location distribution of people (above the x-axis) and items (below the x-axis) on the logit scale. The blue bars below the x-axis represent the distribution of measurement points provided by the scale items. The pink bars above the x-axis represent the distribution of where the patients (persons) are located on the same scale. Note that most of the people fall under the green curve, which represents the useful measurement information that is provided by the scale items.

est impact on daily life. The dizziness item had the highest location, meaning that, overall, this was the least commonly reported symptom to affect daily life.

See Table 5 for all item fit details. One item (skin problems) displayed a high fit residual (2.98), indicating a slight underdiscrimination, but the chi-square fit for this item was good ($P = .17$). Additionally, there was one indication of an apparent pairwise local dependency between the 'fatigue' and 'shortness of breath', although the magnitude of this was not large (Q3 = 0.07 above the criterion cut point).

Despite the overall good fit, the response category structure displayed disordered thresholds, indicating that this structure did not operate as intended. Across all items, the implied probability distribution curves consistently showed the response category 'Rarely' to be dysfunctional, with it never emerging as the most likely response, at any level of the underlying trait. To address this, two generic exploratory post-hoc recodes were implemented across the item pool, and the impact of this was assessed. In rescore 1, the adjacent response categories 'Rarely' and 'Sometimes' were merged to form a four-response structure. In rescore 2, the adjacent response categories 'Not at all' and 'Rarely' were merged to form a four-response structure. For both rescore options, the response category functioning improved, with six out of nine operational items for rescore 1, and seven out of nine operational items for rescore 2. Results are summarized in Table 4, with fit indices remaining good across both options. However, the differing rescore options present a trade-off. For example, rescore 2 appeared to display better fit, but at the cost of slightly skewing the targeting, meaning that a larger floor effect was seen, along with a reduction in the reliability (PSI).

To examine the impact of the rescoring on theta person estimates, the original theta person estimates were correlated against the rescored scale theta person estimates [43]. This indicated that there was strong correlation between the originally scored theta estimates and both the rescore 1 (0.99) and rescore 2 (0.98) estimates. Since this showed that the rescoring had very little effect on the ordering of the person estimates, and as a satisfactory fit was found with the original model, it was decided to preserve the original scoring of the KSB-Q to retain maximum information, and to allow flexibility in future development and validation. Furthermore, the test–retest ICC (based on the original scoring) was 0.82 (95% CI = 0.69–0.89), indicating a good test–retest reliability.

The KSB-Q correlated (Spearman's) at 0.88 with the IPOS-Renal symptom scale; 0.86 with the IPOS-Renal total scale; and at 0.68 with the IPOS-Renal non-symptom scale, which all aligned with the hypothesized values. The 0–100 score MDC of the KSB-Q was 19.14 based on PSI reliability, and 15.43 based on alpha reliability.

The correlation (Pearson's) between the KSB-Q and eGFR was −0.12, which was lower than hypothesized. Previous studies have indicated that while eGFR is a critical clinical indicator, it may not fully capture the multidimensional nature of symptom burden and HRQOL in CKD patients [44]. We purport that this may explain the weak correlation seen between the KSB-Q and eGFR, while acknowledging the strong correlation seen between the KSB-Q and the symptom-focused IPOS-Renal PROM.

Readability

Readability was measured using the web-based application Hemmingway editor [45]. The American Medical Association and the National Institutes of Health recommend that the readability of patient materials should not exceed a sixth grade reading level [46]. The KSB-Q items recorded a fifth

Table 5: Individual item fit and local dependency Q3 residual correlations of KSB-Q items (before rescoring).

Item	Item	Location	SE	Fit residual	X^2	DF	*P* value	Local dependency Q3 residual correlations							
								Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
1	Fatigue	−0.715	0.05	−1.471	10.97	6	.089								
2	Pain	−0.234	0.049	−0.697	8.69	6	.192	−0.029							
3	Memory and concentration	0.197	0.051	−0.57	7.43	6	.283	−0.078	−0.107						
4	Sleep	−0.339	0.048	1.698	2.29	6	.891	−0.120	−0.186	−0.098					
5	Skin problems	0.115	0.047	2.976[a]	9.16	6	.165	−0.248	−0.143	−0.225	−0.120				
6	Gastrointestinal problems	0.197	0.049	−0.167	5.73	6	.454	−0.107	−0.152	−0.103	−0.186	−0.099			
7	Dizziness	0.477	0.053	−0.062	10.45	6	.107	−0.137	−0.124	−0.098	−0.174	−0.111	−0.062		
8	Restless legs	0.21	0.047	1.625	5.82	6	.444	−0.157	−0.092	−0.072	−0.131	−0.207	−0.191	−0.156	
9	Shortness of breath	0.093	0.079	−0.523	5.60	6	.470	0.140[b]	−0.091	−0.156	−0.228	−0.140	−0.080	−0.011	−0.289

DF, degrees of freedom; Location, item location estimate (in logits); SE, standard error; average Q3 residual correlation = −0.127; LD criterion (average Q3 + 0.2) = 0.073;
[a]Misfit;
[b]LD indicated at criterion level.

grade reading level, equivalent to year 6 (age 10–11) in the UK.

DISCUSSION

Principal findings

The KSB-Q represents a short, accessible, symptom PROM with evidence of strong psychometric properties. Rasch analysis indicated that items representing nine key symptom areas (fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs, and shortness of breath) formed a valid, well-targeted, reliable, unidimensional measure of somatic kidney symptom burden. Both cognitive debriefing and item pool survey responses provided evidence of content validity encompassing relevance, comprehensiveness, and clarity.

There has been considerable heterogeneity in the PROMs used to measure symptoms in CKD and relatively few have demonstrated sufficient psychometric properties according to established guidelines developed by the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) group [47]. In a large-scale systematic review conducted in 2017 by Aiyegbusi and colleagues [21], the Kidney Disease Quality of Life Short Form (KDQOL-SF) (80 items) [48], and KDQOL-36 (36 items) [49] were the only measures deemed to have sufficient psychometric evidence for use in patients either not on KRT or on dialysis, and the End-Stage Renal Disease Symptom Checklist Transplantation Module (ESRD-SCL-TM) (43 items) [50] the only measure with sufficient evidence for use in renal transplant recipients. It was acknowledged, however, that these tools were still missing evidence to support important measurement properties including content validity, reliability, measurement error, structural validity, and responsiveness. Similarly, the IPOS-Renal (28 items), a newer measure that was not included in the Aiyegbusi review but is now commonly used in practice, is missing evidence underpinning its content validity, measurement error, structural validity, and responsiveness [31].

The KSB-Q offers several advantages over existing tools with regards to somatic symptom measurement. First, its short length and readability level may enhance patient acceptability and long-term compliance. This is important, as the utility of routine symptom monitoring for patients with CKD, who require life-long treatment, is predicated on sustained engagement from end-users [51]. Second, the KSB-Q achieves a positive rating across most COSMIN criteria, indicting strong measurement properties (Table 6). Third, the KSB-Q items demonstrate good targeting alignment in a broad sample population including those not receiving KRT, individuals receiving dialysis and those with a functioning kidney transplant. Fourth, the KSB-Q has been assessed using Rasch methods, demonstrating a high level of internal construct validity and providing evidence confirming that all items are relevant and contribute to a unidimensional total score [34, 35]. This fit of the Rasch model allows for the provision of an interval-level symptom burden score, which can be easily interpreted through the 0–100 conversion [36]. The KSB-Q overall score may be used quantify symptom burden at the individual or aggregate level, both in research and routine clinical practice. Moreover, patient responses to individual questionnaire items may provide valuable information regarding specific symptoms. In practice, ongoing longitudinal provision of KSB-Q data by patients may help map changes over

Table 6: COSMIN Rating of measurement properties.

COSMIN criterion	Rating	Reason
Content validity	+	The KSB-Q was developed through a rigorous process, including a global systematic review, input from patients, clinicians, and psychometricians, and cognitive debriefing interviews. The items were found to be relevant, comprehensible, and comprehensive by most survey respondents.
Structural validity	+	Rasch analysis confirmed the unidimensionality of the KSB-Q, with good overall scale fit (chi-square $P = .12$) and no significant local dependency issues. The scale was well-targeted, and the items displayed good fit to the model.
Internal consistency	+	The KSB-Q demonstrated high internal consistency with a Cronbach's alpha of 0.87.
Cross-cultural validity/measurement invariance	+	The KSB-Q showed no DIF for sex, age group, or ethnicity. This suggests that the items perform consistently across different subgroups.
Reliability	+	The test-retest reliability of the KSB-Q was good, with an ICC of 0.82 (95% CI 0.69–0.89). This indicates that the measure is stable over time.
Measurement error	+	The SEM and MDC values were reported, with SEM based on Cronbach's alpha being 5.57 and MDC being 15.43. These values are acceptable and indicate low measurement error.
Criterion validity	?	The study did not seek to provide information on the correlation of the KSB-Q with a gold standard measure. Therefore, criterion validity cannot be fully assessed.
Hypotheses testing for construct validity	+	The KSB-Q showed strong correlations with the IPOS-Renal symptom scale (0.88) and total scale (0.86), and moderate correlation with the IPOS-Renal non-symptom scale (0.68). These correlations align with the hypothesized values, supporting construct validity.
Responsiveness	?	The study was not designed as a longitudinal study, so responsiveness could not be established. Further testing is needed to determine the responsiveness of the KSB-Q.

+, the measurement property meets the criteria for good quality; ?, there is not enough information reported to determine whether the measurement property meets the criteria; -, the measurement property does not meet the criteria for good quality.

time and support health care teams/providers to deliver patient-centred care.

Moving beyond the KSB-Q, the use of the Rasch model allows the further development of the full KSB item pool as a CAT, which would administer items based on patients' previous responses, thus tailoring the measure to their individual respondents, enhancing the accuracy of ability estimates and optimizing the information collected, while minimizing questionnaire burden [52]. The use of CAT systems in this way is key in ensuring patients are only asked to address items/domains that are relevant to them; highlighted as an important component in routine PROM monitoring [51]. This should also address concerns that newly diagnosed patients, who are just coming to terms with their disease, might be daunted by the range of symptoms presented in a larger questionnaire; an issue that was highlighted in our cognitive interviews. Our work on the development of a multi-symptom CAT, based on the complete KSB item pool, is ongoing and will be reported in a future paper.

Strengths and weaknesses

The development of the KSB-Q has been supported by a rigorous process, including: (i) concept elicitation derived from a global systematic review incorporating data from almost 200 000 patients across 62 countries [6]; (ii) design input from expert patients, clinicians, and psychometricians and (iii) calibration involving over 400 people with CKD. It has also been validated using Rasch methodology, which provides robust psychometric standards for fundamental measurement [53]. However, although it was shown to make very little difference to the person estimates, the five-response format of the KSB-Q did not function exactly as anticipated, with disordered thresholds being observed across items. These response categories are currently being retained in their original format, but further research may be warranted to identify the optimal format. As the item 'shortness of breath' was added following the pilot phase, the volume of data included in the analysis was lower compared to the other KSB-Q items ($n = 153$ respondents), this should be taken into account when interpreting the results. As our sample represented a predominantly older/white population, there is a need for more diverse sampling in future associated studies to ensure that our findings are generalizable across broader demographic groups. This study was not designed as a longitudinal study, so the responsiveness of the KSB-Q could not be established. The MDC has been reported, but further testing and triangulation with anchor-based, longitudinal methodologies is necessary to formally establish the responsiveness of the scale.

CONCLUSIONS

The KSB-Q provides a short and accessible measure of symptom burden for adults with CKD (stages 3–5), including those not receiving KRT, individuals on dialysis and those with a functioning kidney transplant. The measure has strong psychometric properties and provides an interval-level single summary symptom burden score spanning nine somatic symptom domains, including: fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs, and shortness of breath.

SUPPLEMENTARY DATA

Supplementary data are available at *Clinical Kidney Journal* online.

ACKNOWLEDGEMENTS

The authors thank the study participants for their vital input.

FUNDING

The study was funded by Kidney Research UK (Stoneygate Research Award, KS_RP_013_20180914).

DATA AVAILABILITY STATEMENT

Deidentified individual participant data that underlie the results reported in this article will be made available to share with researchers who provide a methodologically sound proposal, beginning 3 months and ending 36 months following article publication. Proposals should be directed to d.kyte@worc.ac.uk. To gain access, data requestors will need to sign a data access agreement.

CONFLICT OF INTEREST STATEMENT

D.K. reports grants from the National Institute for Health Research (NIHR) outside the submitted work. M.C. is Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation, Director of the Centre for Patient-Reported Outcomes Research and is a National Institute for Health and Care Research (NIHR) Senior Investigator. M.C. receives funding from the NIHR Birmingham Biomedical Research Centre, NIHR Surgical Reconstruction and Microbiology Research Centre and NIHR ARC West Midlands and NIHR Blood and Transplant Research Unit in Precision and Cellular Therapeutics at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust NIHR/UKRI., LifeArc, Health Data Research UK, Innovate UK (part of UK Research and Innovation), Macmillan Cancer Support, UCB Pharma, Janssen, Merck, GSK, and Gilead. M.C. has received personal fees from Astellas Aparito Ltd, CIS Oncology, Halfloop, Takeda, Merck, Daiichi Sankyo, Glaukos, GSK, Pfizer, and the Patient-Centered Outcomes Research Institute (PCORI) outside the submitted work. A.S. reports funding by the British Heart Foundation outside the submitted work. S.D., O.L.A., and M.C. are funded by the NIHR Applied Research Collaboration (ARC) West Midlands. N.A. is funded by the Health Education England (HEE)/NIHR Integrated Clinical Academic (ICA) Clinical Doctoral Research Fellowship (CDRF). O.L.A. receives funding from the NIHR Birmingham Biomedical Research Centre (BRC), NIHR Blood and Transplant Research Unit (BTRU) in Precision Transplant and Cellular Therapeutics, NIHR ARC West Midlands, LifeArc, UKRI, Health Foundation, Merck, Gilead Science Ltd, Anthony Nolan, Sarcoma UK, and GSK. He declares personal fees from Gilead Sciences Ltd, Merck, and GSK outside the submitted work. M.H. reports grants from NIHR, MRC, and European Huntington's Disease Network, and personal fees from Adelphi Outcomes and Proctor & Gamble outside the submitted work. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

REFERENCES

1. Go AS, Chertow GM, Fan D *et al*. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 2004;**351**:1296–305. https://doi.org/10.1056/NEJMoa041031
2. Hill NR, Fatoba ST, Oke JL *et al*. Global Prevalence of chronic kidney disease—a systematic review and meta-analysis. *PLoS ONE* 2016;**11**:e0158765. https://doi.org/10.1371/journal.pone.0158765
3. Chaker L, Falla A, van der Lee SJ *et al*. The global impact of non-communicable diseases on macro-economic productivity: a systematic review. *Eur J Epidemiol* 2015;**30**:357–95. https://doi.org/10.1007/s10654-015-0026-5
4. Sundström J, Bodegard J, Bollmann A *et al*. Prevalence, outcomes, and cost of chronic kidney disease in a contemporary population of 2·4 million patients from 11 countries: the CaReMe CKD study. *Lancet Regional Health—Europe* 2022;**20**:100438. https://doi.org/10.1016/j.lanepe.2022.100438
5. Pollock C, James G, Garcia Sanchez JJ *et al*. Healthcare resource utilisation and related costs of patients with CKD from the UK: a report from the DISCOVER CKD retrospective cohort. *Clin Kidney J* 2022;**15**:2124–34. https://doi.org/10.1093/ckj/sfac168
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. *PLoS Med* 2022;**19**:e1003954. https://doi.org/10.1371/journal.pmed.1003954
7. Perlman RL, Finkelstein FO, Liu L *et al*. Quality of life in chronic kidney disease (CKD): a cross-sectional analysis in the Renal Research Institute-CKD study. *Am J Kidney Dis* 2005;**45**:658–66. https://doi.org/10.1053/j.ajkd.2004.12.021
8. Ricardo AC, Goh V, Chen J *et al*. Association of sleep duration, symptoms, and disorders with mortality in adults with chronic kidney disease. *Kidney Int Rep* 2017;**2**:866–73. https://doi.org/10.1016/j.ekir.2017.05.002
9. Amro A, Waldum B, von der Lippe N *et al*. Symptom clusters predict mortality among dialysis patients in Norway: a prospective observational cohort study. *J Pain Symptom Manage* 2015;**49**:27–35. https://doi.org/10.1016/j.jpainsymman.2014.04.005
10. Tummalapalli SL, Mendu ML. Value-based care and kidney disease: emergence and future opportunities. *Adv Chronic Kidney Dis* 2022;**29**:30–39. https://doi.org/10.1053/j.ackd.2021.10.001
11. Calvert M, Kyte D, Price G *et al*. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ* 2019;**364**:k5267. https://doi.org/10.1136/bmj.k5267
12. FDA; guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes* 2006;**4**:79.
13. Basch E, Deal A, Kris M *et al*. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol* 2016;**34**:557–65 . https://doi.org/10.1200/JCO.2015.63.0830
14. Velikova G, Absolom K, Warrington L *et al*. *Phase III Randomized Controlled Trial of eRAPID (electronic patient self-Reporting of Adverse-events: Patient Information and advice)—An eHealth Intervention during Chemotherapy*. Alexandria, Virginia, USA: American Society of Clinical Oncology: 2020;.
15. Basch E, Schrag D, Henson S *et al*. Effect of electronic symptom monitoring on patient-reported outcomes among patients with metastatic cancer: a randomized clinical trial. *JAMA* 2022;**327**:2413–22. https://doi.org/10.1001/jama.2022.9265
16. Kyte D, Anderson N, Bishop J *et al*. Results of a pilot feasibility randomised controlled trial exploring the use of an

electronic patient-reported outcome measure in the management of UK patients with advanced chronic kidney disease. *BMJ Open* 2022;**12**:e050610. https://doi.org/10.1136/bmjopen-2021-050610
17. Greenham L, Bennett PN, Dansie K *et al*. The Symptom Monitoring with Feedback Trial (SWIFT): protocol for a registry-based cluster randomised controlled trial in haemodialysis. *Trials* 2022;**23**:419. https://doi.org/10.1186/s13063-022-06355-0
18. Grove BE, Ivarsen P, de Thurah A *et al*. Remote follow-up using patient-reported outcome measures in patients with chronic kidney disease: the PROKID study—study protocol for a non-inferiority pragmatic randomised controlled trial. *BMC Health Serv Res* 2019;**19**:631. https://doi.org/10.1186/s12913-019-4461-y
19. Basch EM, Abernethy A, Mullins CD *et al*. Development of a guidance for including patient-reported outcomes (PROS) in post-approval clinical trials of oncology drugs for comparative effectiveness research (CER). *Value Health* 2011;**14**:A10. https://doi.org/10.1016/j.jval.2011.02.060
20. Aiyegbusi OL, Roydhouse J, Rivera SC *et al*. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat Commun* 2022;**13**:6026. https://doi.org/10.1038/s41467-022-33826-4
21. Aiyegbusi OL, Kyte D, Cockwell P *et al*. Measurement properties of patient-reported outcome measures (PROMs) used in adult patients with chronic kidney disease: a systematic review. *PLoS ONE* 2017;**12**:e0179733. https://doi.org/10.1371/journal.pone.0179733
22. Patrick DL, Burke LB, Gwaltney CJ *et al*. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value Health* 2011;**14**:967–77.
23. Wainer H, Dorans NJ, Flaugher R *et al*. *Computerized Adaptive Testing: A Primer*. Oxfordshire, England, UK: Routledge: 2000;. https://doi.org/10.4324/9781410605931
24. Linacre J. Sample size and item calibration stability. *Rasch Mes Trans* 1994;**7**:328.
25. Stucki G. International Classification of Functioning, Disability, and Health (ICF): a promising framework and classification for rehabilitation medicine. *Am J Phys Med Rehabil* 2005;**84**:733–40. https://doi.org/10.1097/01.phm.0000179521.70639.83
26. Evangelidis N, Sautenet B, Madero M *et al*. Standardised Outcomes in Nephrology—Chronic Kidney Disease (SONG-CKD): a protocol for establishing a core outcome set for adults with chronic kidney disease who do not require kidney replacement therapy. *Trials* 2021;**22**:612. https://doi.org/10.1186/s13063-021-05574-1
27. Tong A, Manns B, Hemmelgarn B *et al*. Establishing core outcome domains in hemodialysis: report of the standardized outcomes in nephrology-hemodialysis (SONG-HD) Consensus Workshop. *Am J Kidney Dis* 2017;**69**:97–107. https://doi.org/10.1053/j.ajkd.2016.05.022
28. Tong A, Sautenet B, Poggio ED *et al*. Establishing a core outcome measure for graft health: a standardized outcomes in nephrology-kidney transplantation (SONG-Tx) consensus workshop report. *Transplantation* 2018;**102**:1358–66. https://doi.org/10.1097/TP.0000000000002125
29. 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. *Am J Kidney Dis* 2019;**73**:372–84. https://doi.org/10.1053/j.ajkd.2018.10.007
30. Coles T, Plyler K, Hernandez A *et al*. Recalling what we thought we knew about recall periods: a qualitative descriptive study of how adults diagnosed with cancer use recall periods for patient-reported outcome items about physical function. *Qual Life Res* 2024;**33**:1819–28.
31. Raj R, Ahuja K, Frandsen M *et al*. Validation of the IPOS-Renal Symptom Survey in advanced kidney disease: a cross-sectional study. *J Pain Symptom Manage* 2018;**56**:281–7. https://doi.org/10.1016/j.jpainsymman.2018.04.006
32. Qin S, Nelson L, McLeod L *et al*. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Qual Life Res* 2019;**28**:1029–33. https://doi.org/10.1007/s11136-018-2076-0
33. Terwee CB, Bot SD, de Boer MR *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;**60**:34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012
34. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. San Diego, California, USA: MESA Press: 1993.
35. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud* 2009;**46**:380–93. https://doi.org/10.1016/j.ijnurstu.2008.10.007
36. Tennant A, Küçükdeveci AA. Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Front Rehabil Sci* 2023;**4**:1208670. https://doi.org/10.3389/fresc.2023.1208670
37. Andrich D, Sheridan B, Gao L; *RUMM 2030. 4.0 for Windows (upgrade 4600.0109)*. Perth, WA: RUMM Laboratory Pty Ltd. 2010;.
38. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q(3): identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Meas* 2017;**41**:178–94. https://doi.org/10.1177/0146621616677520
39. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;**3**:205–31.
40. Dontje ML, Dall PM, Skelton DA *et al*. Chastin SFM; reliability, minimal detectable change and responsiveness to change: indicators to select the best method to measure sedentary behaviour in older adults in different study designs. *PLoS ONE* 2018;**13**:e0195424. https://doi.org/10.1371/journal.pone.0195424
41. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg* 2018;**126**:1763–8. https://doi.org/10.1213/ANE.0000000000002864
42. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. *Qual Health Res* 1993;**3**:430–41. https://doi.org/10.1177/104973239300300403
43. Liegl G, Roorda LD, Terwee CB *et al*. Suitability of the animated activity questionnaire for use as computer adaptive test: establishing the AAQ-CAT. *Qual Life Res* 2023;**32**:2403–13. https://doi.org/10.1007/s11136-023-03402-4
44. So S, Brown MA, Li K. Factors associated with quality of life in patients with kidney failure managed conservatively and with dialysis: a cross-sectional study. *BMC Nephrol* 2023;**24**:322. https://doi.org/10.1186/s12882-023-03355-3

45. Hemingway Editor. *Hemingway App makes your writing concise and correct*. https://hemingwayapp.com/ (March 2025, date last accessed).
46. Weiss BD. *Help Patients Understand. Manual for Clinicians*. AMA Foundation 2007;.
47. Mokkink LB, Elsman EBM. Terwee CB; COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Qual Life Res* 2024;**33**:2929–39. https://doi.org/10.1007/s11136-024-03761-6
48. Hays RD, Kallich J, Mapes D *et al*. *Kidney Disease Quality of Life Short Form (KDQOL-SF™), Version 1.3: a Manual for Use and Scoring*. Santa Monica, CA, USA: Rand 1997;**7994**.
49. Hays RD, Kallich JD, Mapes DL *et al*. Development of the kidney disease quality of life (KDQOL) instrument. *Qual Life Res* 1994;**3**:329–38. https://doi.org/10.1007/BF00451725
50. Franke GH, Reimer J, Kohnle M *et al*. Quality of life in end-stage renal disease patients after successful kidney transplantation: development of the ESRD symptom checklist—transplantation module. *Nephron* 1999;**83**:31–39. https://doi.org/10.1159/000045470
51. Schick-Makaroff K, Levay A, Thompson S *et al*. An evidence-based theory about PRO use in kidney care: a realist synthesis. *Patient* 2022;**15**:21–38. https://doi.org/10.1007/s40271-021-00530-2
52. Paap MC, Born S, Braeken J. Measurement efficiency for fixed-precision multidimensional computerized adaptive tests: comparing health measurement and educational testing using example banks. *Appl Psychol Meas* 2018;:0146621618765719.
53. Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care* 2004;**42**:I7–16. https://doi.org/10.1097/01.mlr.0000103528.48582.7c

Received: 30.9.2024; Editorial decision: 3.4.2025

Click here for Prescribing Information

This is a promotional advertisement intended for UK HCPs only, produced and funded by MEDICE UK

Vafseo® vadadustat

Vafseo® is indicated for the treatment of symptomatic anaemia associated with chronic kidney disease (CKD) in adults on chronic maintenance dialysis[1]

A once-daily oral treatment for anaemia associated with CKD in patients receiving dialysis[1]

Vafseo® is now recommended by NICE, within its marketing authorisation, as an option for treating symptomatic anaemia caused by CKD in adults having maintenance dialysis[2]

Contact MEDICE UK now to be put in touch with your regional Vafseo® Key Account Manager

Contact Us

▼ This medicinal product is subject to additional monitoring. This will allow quick identification of new safety information. Adverse events should be reported. Reporting forms and information can be found at https://yellowcard.mhra.gov.uk/ or search for MHRA Yellow Card in Google Play or Apple App Store. Adverse events should also be reported to Medice UK Ltd, 0204 582 2845, medicalinformation@medice.co.uk

MEDICE THE HEALTH FAMILY

Abbreviations: CKD, chronic kidney disease; HCP, healthcare professional; NICE, National Institute of Health and Care Excellence. **References:** 1. Vafseo 300 mg film-coated tablets Summary of Product Characteristics (SmPC). Available at: https://www.medicines.org.uk/emc/product/15656/smpc (Accessed March 2025).
2. NICE Guidance TA1035. Vadadustat for treating symptomatic anaemia in adults having dialysis for chronic kidney disease. Available at: https://www.nice.org.uk/guidance/ta1035 (Accessed March 2025).

UK-VAF-2024-1963-v2
March 2025