Evaluating the quality of life impact of recurrent urinary tract infection: validation and refinement of the Recurrent UTI Impact Questionnaire (RUTIIQ)

Received: 29 August 2023 | Accepted: 5 February 2024

DOI: 10.1002/nau.25426

CLINICAL ARTICLE

Neurourology Urodynamics | SUFU Society of Urodynamics, Female Pelvic Medicine & Urogenital Reconstruction | WILEY

Evaluating the quality of life impact of recurrent urinary tract infection: Validation and refinement of the Recurrent UTI Impact Questionnaire (RUTIIQ)

Abigail F. Newlands BSc, MSc[1] | Melissa Kramer BBus, BA[2] | Lindsey Roberts BSc, MSc, PhD, CPsychol[3] | Kayleigh Maxwell BSc, MSc[4] | Jessica L. Price[2] | Katherine A. Finlay PhD, CPsychol, AFBPsS[1]

[1]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[2]Live UTI Free Ltd, Sandyford, Dublin, Ireland

[3]School of Psychology, University of Buckingham, Buckingham, UK

[4]Department of Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, UK

Correspondence
Katherine A. Finlay, PhD, CPsychol, AFBPsS, School of Psychology and Clinical Language Sciences, University of Reading, Reading, RG6 7BE, UK.
Email: katherine.finlay@reading.ac.uk

Abstract

Background and Aims: Recurrent urinary tract infection (rUTI) has significant negative consequences for a wide variety of quality of life (QoL) domains. Without adequate validation and assessment of the unique insights of people living with rUTI, clinical results cannot be fully understood. The Recurrent UTI Impact Questionnaire (RUTIIQ), a novel patient-reported outcome measure of rUTI psychosocial impact, has been robustly developed with extensive patient and clinician input to facilitate enhanced rUTI management and research. This study aimed to confirm the structural validity of the RUTIIQ, assessing its strength and bifactor model fit.

Methods: A sample of 389 adults experiencing rUTI (96.9% female, aged 18–87 years) completed an online cross-sectional survey comprising a demographic questionnaire and the RUTIIQ. A bifactor graded response model was fitted to the data, optimizing the questionnaire structure based on item fit, discrimination capability, local dependence, and differential item functioning.

Results: The final RUTIIQ demonstrated excellent bifactor model fit (RMSEA = 0.054, CFI = 0.99, SRMSR = 0.052), and mean-square fit indices indicated that all included items were productive for measurement (MNSQ = 0.52–1.41). The final questionnaire comprised an 18-item general "rUTI QoL impact" factor, and five subfactor domains measuring "personal wellbeing" (three items), "social wellbeing" (four items), "work and activity interference" (four items), "patient satisfaction" (four items), and "sexual

Abbreviations: CFI, comparative fit index; DIF, differential item functioning; EFA, exploratory factor analysis; GRM, graded response model; IRT, item response theory; MHRM, Metropolis-Hastings Robbins-Monro (estimation method); MNSQ, mean square; QoL, quality of life; RMSEA, root mean square error of approximation; rUTI, recurrent urinary tract infection; RUTIIQ, Recurrent Urinary Tract Infection Impact Questionnaire; RUTISS, Recurrent Urinary Tract Infection Symptom Scale; SRMSR, standardized root mean square residual.

wellbeing" (three items). Together, the general factor and five subfactors explained 81.6% of the common model variance. All factor loadings were greater than 0.30 and communalities greater than 0.60, indicating good model fit and structural validity.

Conclusions: The 18-item RUTIIQ is a robust, patient-tested questionnaire with excellent psychometric properties, which capably assesses the patient experience of rUTI-related impact to QoL and healthcare satisfaction. Facilitating standardized patient monitoring and improved shared decision-making, the RUTIIQ delivers the unique opportunity to improve patient-centered care.

KEYWORDS

bifactor model, chronic pain, item response theory, patient experience, patient-reported outcomes, psychosocial outcomes, women's health

1 | INTRODUCTION

Recurrent urinary tract infection (rUTI) is defined as experiencing two or more UTIs in six months or three or more in a year,[1] and affects more than 100 million people worldwide annually.[2] Given considerable negative impact to a broad range of quality of life (QoL) domains,[3–6] and significant socioeconomic implications,[4,7–9] there is an urgent need to validate the unique rUTI patient experience and incorporate QoL assessment into clinical management and research.[5,10–12] Without the inclusion of rUTI-specific patient-reported outcome measures (PROMs), evaluation of clinical testing outcomes and symptoms are limited in their real-world application. Such tools are required to improve patient monitoring, shared decision-making, and rUTI management.[13,14]

The Recurrent UTI Impact Questionnaire (RUTIIQ) is a new PROM evaluating rUTI-related impact to QoL.[15] The RUTIIQ was developed in accordance with gold-standard PROM development recommendations by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative,[16,17] with extensive, international expert clinician and patient input (see Figure 1 for methodology; Stages I–IV have been published in Newlands et al., 2023).[15] A five-factor structure was identified by exploratory factor analysis (EFA) of pilot data, comprising: "personal wellbeing," "social wellbeing," "work and activity interference," "patient satisfaction," and "sexual wellbeing."[15] The RUTIIQ demonstrates excellent psychometric properties, including strong test–retest reliability (intraclass correlation coefficient, ICC = 0.66–0.91; computed based on a single-rating, absolute-agreement, two-way mixed effects model[18]), internal consistency (Cronbach's $\alpha = 0.81–0.96$), content validity (item content validity indices, I-CVI = 0.75–1.00), and concurrent validity with related QoL measures (Spearman's $\rho = 0.69–0.76$).[15]

As the required next step in the PROM development process, in preparation for ongoing work to determine the questionnaire's clinical responsiveness to intervention, the current study aimed to build on preliminary testing of the RUTIIQ by confirming its structural validity and identifying areas for refinement.[16] A bifactor structure was hypothesized, expecting a general "rUTI QoL impact" factor that explains the common variance between all items, and five specific, uncorrelated subfactors aligning with the factors identified through EFA, each explaining the unique influence of a specific construct beyond the general factor.[19,20] Bifactor modelling enables the evaluation of general scores based on all items represented by a general trait, as well as individual domain scores for items represented by specific traits.[19–21]

2 | MATERIALS AND METHODS

2.1 | Study design and participants

Adults meeting the diagnostic criteria for rUTI,[1] with advanced or proficient fluency in English based on the Common European Framework of Reference for Languages,[22] participated in an online cross-sectional survey ($N = 389$, 96.9% female biological sex; see Table 1 for characteristics). Definition of rUTI was presented to participants using the Recurrent UTI Symptom Scale (RUTISS),[23,24] with participants required to report at

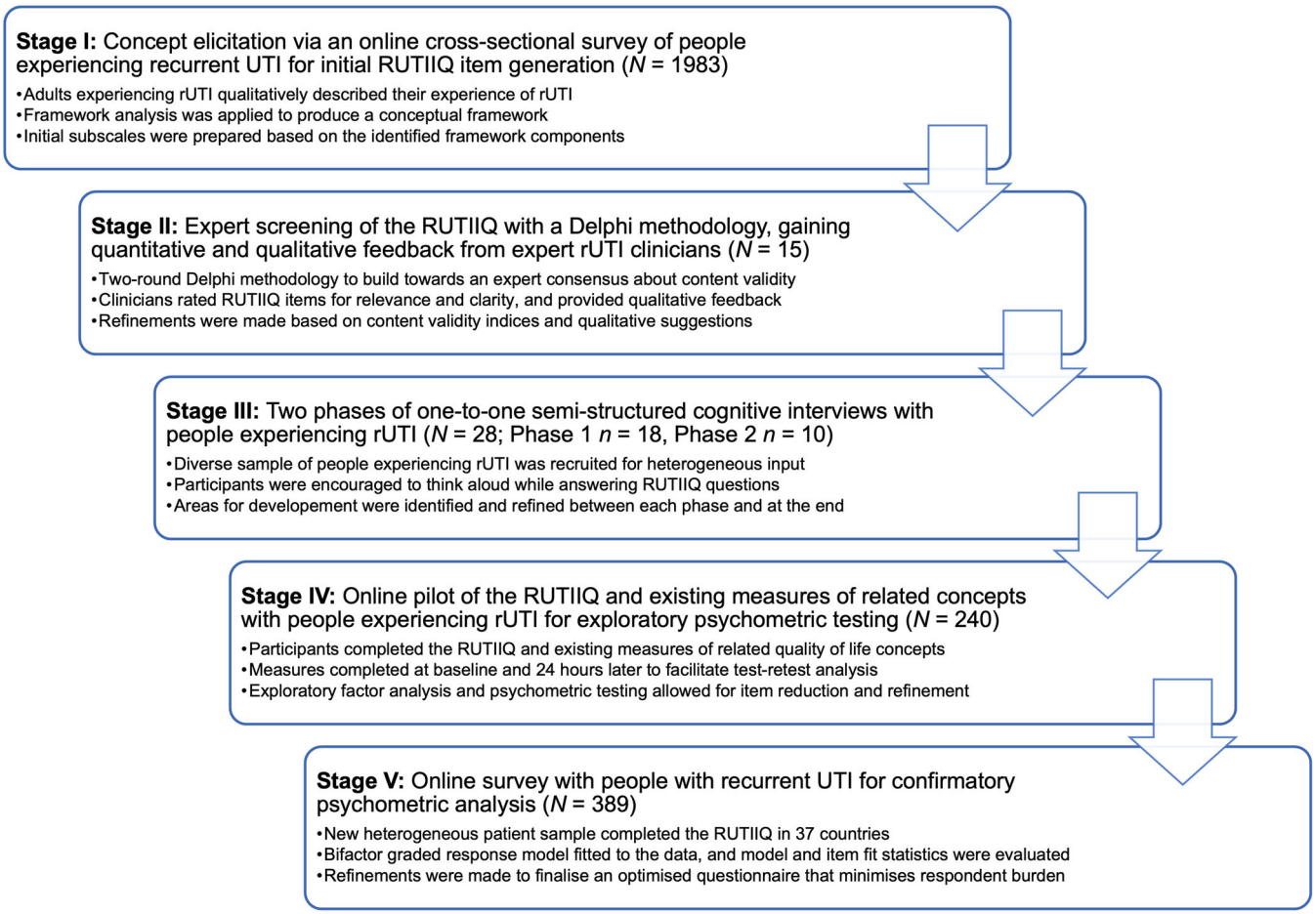

FIGURE 1 Methodology employed to develop and validate the Recurrent Urinary Tract Infection Impact Questionnaire (RUTIIQ). The current study reports the methodology and findings from Stage V. Results from Stages I–IV are published in Newlands et al., 2023.[15]

least two symptomatic episodes of UTI in the past six months or at least three in the past year.[1] Inclusion of participants based on self-report of symptoms reflected recent microbiological findings that standard urine culture misses up to 58% of true infections and published recommendations regarding the need to prioritize patient symptom reporting.[25–27] This approach enabled access to the larger, more heterogeneous sample required for item response theory (IRT) analysis. Exclusion criteria comprised a current diagnosis of interstitial cystitis. Aiming to validate the confirmatory factor structure and make refinements to minimize respondent burden, RUTIIQ data were collected for psychometric analyses including IRT (see Section 2.3).[19,28,29]

At least 250 participants were required to facilitate multidimensional IRT analysis with a graded response model (GRM) for ordered polytomous data[30]; sampling adequacy was exceeded (see Figure S1 for sampling flow diagram). Participants were mostly recruited via newsletters and social media posts from a key UTI stakeholder group, Live UTI Free (84.8%, $n = 300$), participants and clinicians sharing the study information on social media or by word of mouth (9.51%, $n = 37$), and online UTI-focused support groups (5.66%, $n = 22$).

2.2 | Procedure

After reviewing the study information and ethical considerations, participants gave electronic consent and completed a screening questionnaire to confirm eligibility to participate (see Figure S1). Eligible participants completed the preliminary RUTIIQ, a 30-item self-report questionnaire assessing rUTI-related impact to QoL with five domains: "personal wellbeing," "social wellbeing," "work and activity interference," "patient satisfaction" with UTI-related medical care, and "sexual wellbeing" (optional domain, preceded by a pre-qualifying question: "Do you feel your UTI(s) has/have impacted your sex life in the past 2 weeks?").[15] Participants used an 11-point Likert scale ranging from 0 ("strongly disagree") to 10 ("strongly agree") to rate their level of agreement with

TABLE 1 Participant demographic characteristics.

Characteristic	*n*	%
Biological sex		
Female	377	96.9
Male	12	3.08
Gender		
Female	374	96.1
Male	12	3.08
Nonbinary	2	0.51
Prefer not to say	1	0.26
Country of residence		
United Kingdom	153	39.3
United States	147	37.8
Canada	26	6.68
Australia	8	2.06
Ireland	5	1.29
Greece	4	1.03
India	4	1.03
Spain	4	1.03
Other[a]	38	9.77
Ethnicity		
Asian (including Asian American, Asian British)	10	2.57
Black (including African, African American, Caribbean, Black British)	3	0.77
Hispanic or Latino American	5	1.29
Mixed ethnicity or multiple ethnic groups	4	1.03
Native Hawaiian or other Pacific Islander	2	0.51
White (including Caucasian, White British, White European)	340	87.4
Other ethnicity	4	1.03
Prefer not to say	21	5.40
Fluency in English		
Native or bilingual	337	86.6
Advanced or proficient	52	13.4
Relationship status		
Married or in a civil partnership	199	51.2
In a relationship (unmarried)	117	30.1
Single	44	11.3
Divorced	14	3.60
Widowed	6	1.54
Separated	5	1.29
Other	2	0.51
Prefer not to say	2	0.51
Highest level of education		
Some high school/secondary school	7	1.80
High school/secondary school	65	16.7
Bachelor's degree or equivalent	168	43.2
Master's degree or equivalent	104	26.7
Doctoral level training or equivalent	16	4.11
Other professional qualification(s)	22	5.66
Prefer not to say	7	1.80
Annual household income (GBP)		
No current income	12	3.08
£1–£9999	15	3.86
£10 000–£24 999	37	9.51
£25 000–£49 999	102	26.2
£50 000–£74 999	48	12.4
£75 000–£99 999	40	10.3
£100 000 or more	59	15.2
Prefer not to say	76	19.5

Note: $N = 389$.

[a]Other countries where $n \leq 3$ comprise the following 29 countries listed alphabetically: Angola, Argentina, Austria, The Bahamas, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Iceland, Israel, Italy, Jersey, Malawi, Mexico, Netherlands, New Zealand, Nigeria, Norway, Romania, Serbia, Slovakia, South Africa, Sweden, Thailand, Turkey, and Ukraine.

statements about rUTI impact. After completing the RUTIIQ, a debrief form signposted participants to support resources.

2.3 | Data handling and statistical analysis

The final sample comprised 389 participants (Figure S1). Eighty-nine participants did not complete the optional "sexual wellbeing" questions, responding "no" or "prefer not to say" to the pre-qualifying question. This left a total of 300 included datapoints for multidimensional IRT analysis. Responses to the "patient satisfaction" questions were reverse scored, thus higher scores indicated higher levels of rUTI QoL impact for all domains. As outlined in this section, a preliminary model was identified, making refinements based on item and model fit, before reconducting IRT analysis with a final model. Statistical terminology definitions are available in Table S1.

2.3.1 | Preliminary model identification

IRT analysis was conducted in R using the *mirt* package,[31] fitting a confirmatory bifactor model to assess the plausibility of assessing and scoring a general, overarching factor as well as individual domains (subfactors).[19,21] The confirmatory model specified one general factor onto which all 30 items were expected to load ("rUTI QoL impact"), and five orthogonal subfactors aligning with the factors identified during EFA[15]: "personal wellbeing" (items A1–A4), "social wellbeing" (items B1–B5), "work and activity interference" (items C1–C7), "sexual wellbeing" (items D1–D4), and "patient satisfaction" (items E1–E10). Recommendations for IRT modelling with polytomous scales (i.e. scales with more than two response options) by Toland et al. and by Reeve and Fayers were followed, identifying a suitable approach based on the RUTIIQ's ordered, polytomous Likert-type scale which successively increases from 0 ("strongly disagree") to 10 ("strongly agree").[19,30,32] A bifactor GRM for ordered polytomous scales was thus fitted, which is also specifically recommended for PROM evaluation.[30,32,33] The Metropolis-Hastings Robbins-Monro (MHRM) estimation method, the mathematical algorithm recommended to estimate multidimensional IRT models with more than three expected factors, was employed to estimate item and model parameters.[19]

2.3.2 | Model assumption checks

The intercept parameters, which govern the choice of the next response category over the previous one (e.g., responding 10 vs. 9), were examined to assess participants' use of the 11-point Likert scale.[19] It was expected that the parameters would successively decrease as the response categories (therefore the latent trait of QoL impact) increased.[19]

IRT analysis assumes that there is local independence of items after controlling for a latent construct. In other words, after controlling for the factor influencing a respondent's choice of response, there should be no statistically significant association or correlation between items.[34–36] Item pairs that do not meet this assumption are said to exhibit local item dependence, or LID. Yen's Q_3 statistics were computed for each item pair within the bifactor model, with values above 0.50 indicating LID.[34–36] It was expected that the Yen's Q_3 statistics should be less than 0.50 for all item pairs except those in which both items measure the same subfactor or "testlet" trait (e.g., A1–A2, B1–B2, etc.).[19,37]

Finally, it is assumed that an item should be interpreted in the same way across different subgroups, known as item invariance.[38] Likelihood ratio χ^2 analysis was conducted to check for item invariance, or the absence of differential item functioning (DIF).[38] This examined the extent to which each item performs differently within the model based on biological sex (female vs. male), age (older vs. younger than the median, 42 years old), household income (> £25 000 vs. < £25 000), level of education (university degree or above vs. school or lower), and current antibiotic use (yes vs. no).[38] Model parameters were freely estimated across categorical groups and p values were adjusted using the Bonferroni correction.[39] A statistically significant group difference (χ^2, $p_{adj} < 0.05$) indicated the presence of DIF.[38]

2.3.3 | Model fit and performance

Standardized item factor loadings were expected to be greater than 0.30 and communalities greater than 0.60, indicating good fit and structural validity.[40] Mean-square (MNSQ) outfit statistics were examined to evaluate item fit, with values between 0.50 and 2.00 indicating acceptability for measurement.[41] Item slope (discrimination) parameters (α), which can be interpreted similarly to factor loadings in classical factor analysis,[42] suggested which items performed best within the model in terms of differentiating between respondents' level of QoL impact.[19,43] Minimum $\alpha = 0.65$ was expected to suggest at least "moderate" discrimination capability, with higher values indicating better performance.[43]

Overall model fit was evaluated by computing the C_2 statistic of goodness of fit for ordinal data, with a non-statistically significant result suggesting good model fit.[44] This test is sensitive to sample size, thus making model fit inferences based on the following indices is usual: root mean square error of approximation ($RMSEA_{C2}$; "good fit" ≤0.06), Comparative Fit Index (CFI; "good fit" ≥0.95), and standardized root mean square residual (SRMSR; "good fit" ≤0.06).[44,45]

2.3.4 | Model refinement

The RUTIIQ was thus refined and finalized applying the following strategy.[19,34–41,43–45] First, an item was proposed for deletion if it: (i) demonstrated statistically significant DIF ($p_{adj} < 0.05$), (ii) showed poor item fit (MNSQ < 0.50 or >2.00), (iii) indicated low discrimination capability ($\alpha < 0.65$), (iv) demonstrated poor factor loading (<0.30), or (v) contributed insufficiently to the common model variance (communality, $h^2 < 0.60$). While some level of LID was expected due to related items within subfactors or

testlets (e.g., items A1–A4),[19,37] in cases of LID ($Q_3 > 0.50$) between item across different subfactors (e.g., items A1–B1), one item from the pair was considered for deletion based on which demonstrated stronger performance and fit. Finally, the RMSEA, CFI, and SRMSR model fit indices were assessed.

IRT analysis was re-conducted iteratively after making each proposed deletion until a confirmed model was reached. The final, refined version of the RUTIIQ was created according to this (see Table 2 for included items; the full RUTIIQ and scoring instructions are available from the corresponding author).

2.3.5 | Reliability

The internal consistency of the final RUTIIQ was evaluated by computing Cronbach's alpha (α) coefficients for the general "rUTI QoL impact" factor and for each subfactor, with $\alpha > 0.70$ indicating acceptable reliability.[46]

2.3.6 | Readability

The minimum literacy level for comprehension of the final RUTIIQ was estimated with the Automated Reliability Index (ARI), a readability assessment tool suitable for nonnarrative text such as PROMs.[47]

3 | RESULTS

3.1 | Participants

Most participants reported female biological sex (96.9%, $n = 377$; Table 1), and were aged between 18 and 87 years old ($M = 45.4$, SD = 17.1). Participants resided in 37 countries, mainly the United Kingdom (39.3%, $n = 153$) and United States (37.8%, $n = 147$). Approximately three-quarters (74.0%, $n = 288$) reported a bachelor's degree or higher, and approximately a third (37.8%, $n = 147$) reported an annual household income above £50 000. Participants reported an average of 3.62 UTIs in the past 6 months (SD = 2.90), and 7.06 in the past year (SD = 5.91).

3.2 | Preliminary bifactor model

The preliminary bifactor model, composed of one general factor and five orthogonal subfactors, converged successfully with MHRM estimation.

TABLE 2 List of final 18 items included in the Recurrent Urinary Tract Infection Impact Questionnaire (RUTIIQ).

Updated item number	Item
	Because of my UTI(s)...
A1	I have experienced feelings of anxiety.
A2	I have experienced feelings of low mood or depression.
A3	I have felt hopeless about the future.
B1	I have avoided socializing more than I used to.
B2	I have felt embarrassed in social situations.
B3	I have felt that I am no longer close to anyone.
B4	I have felt anxious in social situations.
C1	I regularly missed full or partial days of work, home responsibilities or studying.
C2	The kind or amount of work I could do was limited.
C3	It was more difficult than usual to concentrate on my work.
C4	It was more difficult than usual to handle my workload.
	Thinking about my UTI-related medical care...
D1	I have felt confident about being able to get the medical care I need.
D2	I have felt like my medical concerns are taken seriously.
D3	I have felt I could access UTI testing and treatment quickly enough.
D4	I have had easy access to the medical specialists I need.
E1*	I have avoided sexual activity to minimize risk of developing or worsening UTI symptoms.
E2*	I have felt unable to enjoy sexual activity due to my UTI(s).
E3*	I have been concerned about the impact of my UTI(s) on my sex life and/or sexual relationship(s).

Note: The full RUTIIQ and scoring instructions are available from the corresponding author.

*Questions about sexual wellbeing (items E1–E3) are optional and only to be answered by respondents who report that their UTI(s) has/have impacted their sex life in the past 2 weeks. The "Thinking about my UTI-related medical care" prompt only applies to items D1-D4, and not to items E1-E3.

3.2.1 | Preliminary model assumption checks

All model assumption checks were passed: (i) successively decreasing model intercept parameters confirmed consistent use of the 11-point Likert scale (see Table S2); (ii) no

LID was identified between items within different subfactors, meeting the assumption of local independence of items ($Q_3 < 0.50$), (iii) no DIF was identified based on age, biological sex, household income, education level, or current antibiotic use, meeting the assumption of item invariance (χ^2, $p_{adj} > 0.05$; see Table S3).

3.2.2 | Preliminary model fit and performance

The common variance was strongly represented, with model and item fit suggesting necessary areas for refinement. The general "rUTI QoL impact" factor and the five subfactors collectively accounted for 79.5% of the common variance, with the general factor accounting for 35.6% and subfactors accounting for between 2.00% and 19.1% each (Table S2). All items loaded onto the general factor and one subfactor with standardized factor loadings above 0.30 (range = 0.32–0.84, Table S2), indicating strong fit and structural validity, except item A4 assessing sleep disruption (general factor loading = 0.69, subfactor loading = 0.11). All items contributed sufficiently to the common model variance, demonstrating communalities above 0.60 (Table S2), except item A4 ($h^2 = 0.49$) and item C3 assessing self-pressure to work despite illness ($h^2 = 0.57$).

All item MNSQ fit statistics were between 0.50 and 2.00 for good fit, except item E4 assessing confidence in treatment decisions (MNSQ = 2.54; see Table S2). All except one item (A4) indicated at least "moderate" discrimination capability, demonstrated by slope parameters $\alpha > 0.64$, with 25 items (83.3%) indicating at least "high" discrimination ($\alpha > 1.35$).[43]

The CFI suggested good model fit (0.97). As the C_2 goodness of fit test ($C_2(375, N = 300) = 1124.26$, $p < 0.05$), RMSEA (0.082, 95% CI [.076, 0.087]) and SRMSR (0.064) suggested inadequate model fit, results indicated it was necessary to refine the model.

3.2.3 | Preliminary model refinement

The refinement strategy outlined in Section 2.3 was applied, removing poor fitting items one at a time, and re-running the analysis after each proposed deletion to assess the impact on model and item fit. Overall, 12 items were removed: A4, B1, C2, C3, C6, D4, E1, E4, E6, E8, E9, and E10 (see Table S4).

3.3 | Final bifactor model

A final 18-item confirmatory bifactor model was identified (see Figure 2), comprising one general "rUTI QoL impact" factor and five orthogonal subfactors. The final 18-item RUTIIQ was created in line with this model with updated item numbering (see Table 2 for final included items; the full RUTIIQ and scoring instructions are available from the corresponding author). This briefer measure demonstrates strong psychometric properties while seeking to minimize respondent burden.

As detailed in the scoring instructions that accompany the RUTIIQ (available from the corresponding author), an overall RUTIIQ impact score and five individual domain scores for specific QoL traits may be computed. Individual domain scores are computed by summing the item scores relevant to each domain,[48] with a maximum possible score range of 0–30 for the "personal wellbeing" and "sexual wellbeing" domains, and 0–40 for the "social wellbeing," "work and activity interference," and "patient satisfaction" domains. There may be no individual score for the "sexual wellbeing" domain since it is optional. The overall RUTIIQ impact score, a transformed score with maximum range 0–100 may be computed to simplify interpretation and comparison, especially in cases where respondents have opted not to answer the optional "sexual wellbeing" items.[48] To calculate a RUTIIQ impact score, administrators may sum each individual domain score, divide this sum by the number of completed items, and then multiply this by 10. The observed scores in this sample ($N = 389$) demonstrated heterogeneity in rUTI-specific QoL impact, ranging the full breadth of possible scores (Table 3).

3.3.1 | Final model assumption checks

All model assumption checks were passed: (i) evidence of successively decreasing intercept parameters and no disordered thresholds (Table 4); (ii) no LID was identified between items within different subfactors, meeting the assumption of local independence of items ($Q_3 < 0.50$); (iii) the assumption of item invariance was satisfied, evidenced by finding no DIF based on age, biological sex, household income, education, or current antibiotic use, meeting the assumption of item invariance (χ^2, $p_{adj} > 0.05$; Table S5).

3.3.2 | Final model fit and performance

The common variance was strongly represented, and both model and item fit were excellent. The general 'rUTI QoL impact' factor and five subfactors collectively accounted for 81.6% of the common variance (Table 4). The general factor accounted for 41.2% of the common variance, and the subfactors accounted for between 3.8%

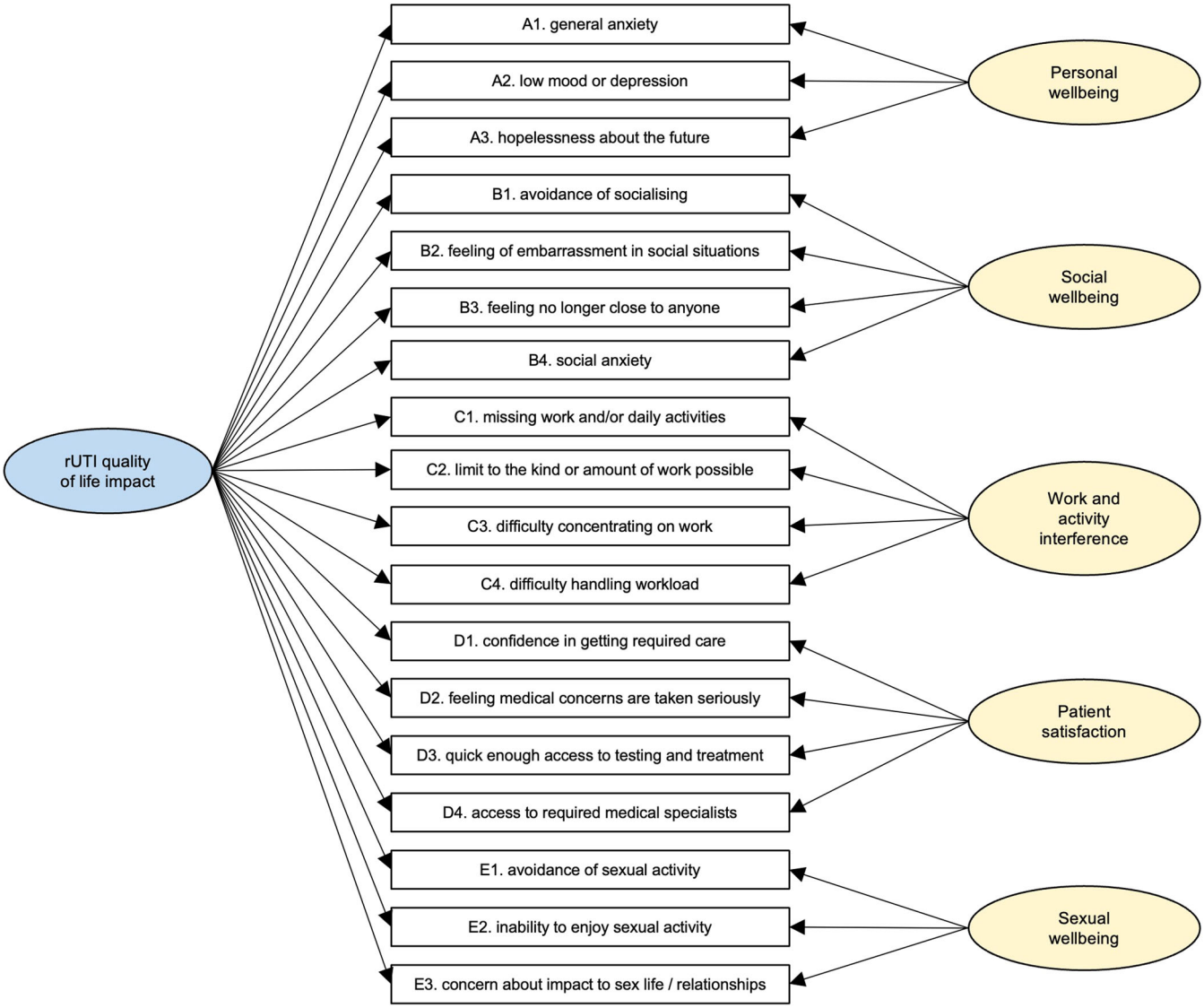

FIGURE 2 This diagram demonstrates the bifactor structure represented by the 18 items included in the Recurrent Urinary Tract Infection Impact Questionnaire (RUTIIQ). All items load onto the general factor colored in blue on the left: "rUTI quality of life impact." Each item also loads onto a subfactor colored in yellow on the right, each assessing a specific recurrent urinary tract infection (rUTI) quality of life trait. Standardized factor loadings (>0.30), communalities (>0.60), and model fit indices (RMSEA = 0.054, CFI = 0.99, SRMSR = 0.052) indicate excellent fit and structural validity (see Table 4).

and 13.2% each. All standardized factor loadings and communalities were greater than 0.30 and 0.60, respectively, indicating excellent fit and structural validity (Table 4). All item MNSQ fit statistics indicated productivity for measurement, falling between 0.52 and 1.41 ($M = 0.81$, SD = 0.24; see Table 4). All items indicated at least 'moderate' discrimination capability ($\alpha > 0.64$), with 15 (91.7%) demonstrating at least 'high' capability and performance ($\alpha > 1.35$; Table 4).[43]

The model RMSEA (0.054, 95% CI [.042, 0.064]), SRMSR (0.052) and CFI (0.99) demonstrated excellent model fit.[44,45] While the C_2 goodness of fit test produced a statistically significant result ($C_2(117, N = 300) = 217.87$, $p < 0.05$), model fit inferences were based on the RMSEA, SRMSR and CFI due to the test's sensitivity to sample size.[44,45]

3.3.3 | Reliability

Internal consistency was excellent for the 18-item general rUTI QoL impact' factor ($\alpha = 0.92$).[46] Reliability was similarly strong for the five subfactors, with findings ranging from $\alpha = 0.80$–0.93 (Table S6).

TABLE 3 Observed Recurrent Urinary Tract Infection Impact Questionnaire (RUTIIQ) scores.

Score	Items*	*N*	*M*	SD	Observed range
Overall RUTIIQ impact score	A1–E3	389	61.4	23.3	0–100
Individual domain scores					
Personal wellbeing	A1–A3	389	20.1	8.94	0–30
Social wellbeing	B1–B4	389	17.8	13.0	0–40
Work and activity interference	C1–C4	389	21.2	13.7	0–40
Patient satisfaction[a]	D1–D4	389	28.3	10.6	0–40
Sexual wellbeing[b]	E1–E3	300	26.0	5.63	0–30

Note: Higher scores indicate greater level of rUTI-related impact to quality of life. All observed scores ranges align with the maximum possible ranges for each score type.

Abbreviations: *M*, mean; SD, standard deviation.

[a]Reverse scored.

[b]The "sexual wellbeing" domain is optional, therefore not all participants responded to these questions.

*Item numbers as per the refined, final version of the RUTIIQ (see Table 2; the full RUTIIQ and scoring instructions are available from the corresponding author).

3.3.4 | Readability

The ARI for the final RUTIIQ is 7.0, demonstrating that this PROM is appropriate for people with a reading age of 12 years old or above (approximately US 7th grade, UK Key Stage 3/year 8).[47]

4 | DISCUSSION

The RUTIIQ is a psychometrically strong measure of patient-reported QoL outcomes and rUTI healthcare experience. A rigorous PROM development methodology following best-practice recommendations by COSMIN maximized international patient and clinician input throughout,[16,17] with exploratory analysis indicating excellent internal consistency, test–retest reliability, content validity, construct validity, and structural validity.[15] The high-quality statistical approaches utilized to refine the RUTIIQ demonstrated the strength of its factor structure.

The final, optimized 18-item RUTIIQ demonstrates a well-fitted bifactor structure that minimizes respondent burden. A general factor evaluates "rUTI QoL impact," with five subfactors measuring "personal wellbeing," "social wellbeing," "work and activity interference," "patient satisfaction," and "sexual wellbeing." Simple scoring and administration instructions are provided with the questionnaire, available from the corresponding author. The RUTIIQ can be utilized within a number of clinical and research contexts, including providing rapid quantitative insights into key QoL domains impacted by rUTI, assessing longitudinal change in patient QoL outcomes in response to new and existing interventions, exploring patient-specific responses to antibiotic treatment approaches, and identifying underlying domains that may benefit from targeted medical and/or psychosocial management and intervention. It is recommended that in both clinical and research settings, the RUTIIQ is administered alongside a validated rUTI-specific patient-reported outcome measure of rUTI symptom presentation to capture the full breadth of the rUTI patient perspective, such as the 15-item Recurrent UTI Symptom Scale (RUTISS).[23,24]

The bifactor model and item fit statistics were excellent,[44,45] highlighting the strength of the RUTIIQ and its structural validity. The final 18 items each loaded highly onto a specific subfactor in addition to the general factor, demonstrating that they can also be assessed as separate, individual domains.[44,45] Internal consistency and reliability of the general factor and subfactors were high, meeting gold-standard recommendations.[17]

Further research would address certain limitations. It is recognized that most participants were Caucasian and reported a high level of education and household income, thus further research is necessary to establish cross-validation of the RUTIIQ and develop translations for non-English speaking populations or lower socio-economic status respondents. It is acknowledged that some RUTIIQ respondents may opt not to complete the sexual wellbeing items; this is in accordance with the APA Ethics Code and UK Government Social Research published recommendations on asking personal questions,[49,50] and the remaining subscales maintain validity. Although rUTI is significantly more prevalent in females,[8] additional testing of the RUTIIQ would be

TABLE 4 Bifactor graded response model item parameter estimates, fit statistics, and factor structure of the 18-item Recurrent Urinary Tract Infection Impact Questionnaire (RUTIIQ).

	Slope						Intercept										Standardized factor loading							
Item*	α^{G}	α^{S1}	α^{S2}	α^{S3}	α^{S4}	α^{S5}	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	G	S1	S2	S3	S4	S5	h^2	Item MNSQ outfit
A1	2.54	1.22					7.39	5.64	4.96	4.62	3.97	3.26	2.12	1.08	−0.32	−1.23	0.77	0.37					0.73	0.78
A2	5.51	3.84					13.0	10.5	8.70	7.56	6.34	4.87	3.69	1.67	−0.87	−3.05	0.80	0.56					0.94	0.56
A3	3.37	2.15					6.90	5.26	4.41	3.84	3.20	2.47	1.50	0.20	−1.04	−2.27	0.78	0.50					0.85	0.75
B1	4.01		1.87				5.80	4.48	3.52	2.93	2.50	1.57	0.68	−0.33	−1.94	−3.03	0.85		0.39				0.87	0.69
B2	2.34		2.44				3.52	1.88	1.20	0.57	0.24	−0.38	−1.02	−1.74	−2.75	−3.63	0.62		0.64				0.80	0.75
B3	1.79		1.61				1.44	0.50	0.01	−0.39	−0.89	−1.43	−1.91	−2.36	−3.09	−3.37	0.61		0.55				0.67	0.73
B4	4.45		3.32				5.70	3.85	2.56	1.95	1.23	0.36	−0.65	−2.00	−3.42	−4.54	0.77		0.57				0.92	0.60
C1	2.45			1.90			3.43	2.32	1.74	1.17	0.85	0.11	−0.50	−1.14	−1.79	−2.33	0.69			0.54			0.77	0.98
C2	3.69			3.15			5.94	4.64	3.72	2.93	2.18	1.36	0.72	−0.22	−1.54	−2.57	0.72			0.61			0.89	0.72
C3	3.72			2.95			6.92	5.85	4.64	3.74	2.76	1.98	1.29	0.50	−1.00	−2.05	0.74			0.59			0.89	0.71
C4	3.49			2.80			5.94	4.64	3.72	2.93	2.18	1.36	0.72	−0.22	−1.54	−2.57	0.73			0.58			0.87	1.03
D1	2.42				4.64		11.3	9.18	7.76	6.71	5.46	4.50	3.24	2.04	0.23	−2.14	0.44				0.84		0.90	0.72
D2	1.37				2.87		6.39	5.36	3.97	3.23	2.76	1.80	1.17	0.76	−0.59	−1.73	0.38				0.80		0.78	1.03
D3	1.30				2.03		5.14	4.36	3.67	3.15	2.64	2.07	1.47	0.82	0.05	−1.01	0.44				0.69		0.67	1.19
D4	1.26				2.39		6.17	5.49	4.24	3.27	3.04	2.12	1.59	0.86	−0.19	−1.21	0.40				0.75		0.72	0.88
E1	2.28					4.11	10.5	9.67	8.71	8.11	7.13	6.35	6.05	4.77	3.02	1.49	0.46					0.82	0.88	0.52
E2	2.55					4.27	9.84	9.34	8.53	7.65	6.91	5.75	5.03	4.06	2.66	0.85	0.49					0.81	0.90	0.55
E3	1.72					1.73	7.37	6.22	5.68	4.80	4.29	3.90	3.18	2.33	1.12	−0.13	0.58					0.58	0.67	1.41
ECV																	0.41	0.04	0.07	0.08	0.13	0.09		

Note: α = slope (or discrimination) parameters; higher slopes indicate greater discrimination. "moderate" discrimination capability: $\alpha = 0.65–1.34$; "high" discrimination capability: $\alpha = 1.35–1.69$; "very high" discrimination capability: $\alpha \geq 1.70$.[41] c_1–c_{10} = intercept parameters; these should successively decrease in value between c_1 and c_{10} to demonstrate consistent use of the 11-point scale.[19] Item MNSQ fit statistics between 0.50 and 2.00 are interpreted as acceptable for measurement, with statistics closer to 1.0 indicating best fit to the model with the least distortion.[40]

Abbreviations: ECV, explained common variance; G, general factor (rUTI quality of life impact); h^2, communality; MNSQ, mean square (item fit statistics); S1, sub-factor 1 (personal wellbeing); S2, sub-factor 2 (social wellbeing); S3, sub-factor 3 (work and activity interference); S4, sub-factor 4 (patient satisfaction); S5, sub-factor 5 (sexual wellbeing).

*Item numbers as per the refined, final version of the RUTIIQ (see Table 2; the full RUTIIQ and scoring instructions are available from the corresponding author). Items D1–D4 (patient satisfaction) have been reverse scored.

beneficial to better understand its psychometric properties in males. Early indications suggest that the RUTIIQ is accessible and appropriate for use by males living with rUTI, however it should be noted the small proportions of males recruited for this study mean that broader extrapolation of the RUTIIQ to male-specific rUTI presentation is currently limited, reflective of a paucity of male-specific literature in UTI. While the sample size was adequate, a larger sample may have further improved the model parameter estimates, thus future research could seek to validate the model within a larger patient sample and in a clinical context.[30] The next stage in PROM development, assessing the responsiveness and clinical interpretability of the RUTIIQ, is ongoing.[16]

5 | CONCLUSION

The RUTIIQ is a psychometrically valid 18-item questionnaire assessing patient-reported personal wellbeing, social wellbeing, work and activity interference, patient satisfaction, and sexual wellbeing. Its simple scoring facilitates standardized patient monitoring and quantification of QoL impact. This brief patient-reported outcome measure offers a unique opportunity to critically assess and prioritize the rUTI patient perspective, supplementing clinical management by improving shared decision-making and highlighting psychosocial challenges requiring intervention.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conceptualization and methodological design. Abigail F. Newlands undertook the study investigation, data collection, and project administration. Abigail F. Newlands and Katherine A. Finlay conducted formal data analysis and interpretation. Melissa Kramer and Jessica L. Price contributed to the study resources and participant recruitment. Abigail F. Newlands prepared the original draft manuscript, and all authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to thank Live UTI Free for supporting this study and recruitment.

CONFLICT OF INTEREST STATEMENT

Melissa Kramer is CEO of Live UTI Free Ltd.; however, no financial incentives have been received.

DATA AVAILABILITY STATEMENT

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

ETHICS STATEMENT

Ethical approval was granted by the School of Psychology and Clinical Language Sciences Research Ethics Committee, University of Reading (project reference no.: 2022-115-KF). All participants were provided with study information sheets and debrief forms, and electronically signed their consent to take part before participation.

ORCID

Abigail F. Newlands http://orcid.org/0000-0002-4718-0075
Melissa Kramer http://orcid.org/0000-0002-9242-5203
Lindsey Roberts http://orcid.org/0000-0001-5277-2377
Kayleigh Maxwell http://orcid.org/0000-0002-8747-7201
Jessica L. Price http://orcid.org/0000-0002-0487-0826
Katherine A. Finlay http://orcid.org/0000-0002-8997-2652

REFERENCES

1. Bonkat G, Bartoletti R, Bruyére F, et al. EAU Guidelines on Urological Infections 2020. *European Association of Urology Guidelines 2020 Edition*. European Association of Urology Guidelines Office; 2020.
2. Zeng Z, Zhan J, Zhang K, Chen H, Cheng S. Global, regional, and national burden of urinary tract infections from 1990 to 2019: an analysis of the global burden of disease study 2019. *World J Urol*. 2022;40(3):755-763. doi:10.1007/s00345-021-03913-0
3. National Health Service. Urinary tract infections (UTIs). 2022. Accessed January 4, 2023. https://www.nhs.uk/conditions/urinary-tract-infections-utis/
4. Wagenlehner F, Wullt B, Ballarini S, Zingg D, Naber KG. Social and economic burden of recurrent urinary tract infections and quality of life: a patient web-based study (GESPRIT). *Exp Rev Pharmacoeconom Outcome Res*. 2018;18(1):107-117. doi:10.1080/14737167.2017.1359543
5. Naber KG, Tirán-Saucedo J, Wagenlehner FME. Psychosocial burden of recurrent uncomplicated urinary tract infections. *GMS Infect Dis*. 2022;10:10. doi:10.3205/id000078
6. Boeri L, Capogrosso P, Ventimiglia E, et al. Six out of ten women with recurrent urinary tract infections complain of distressful sexual dysfunction—a case-control study. *Sci Rep*. 2017;7(1):44380. doi:10.1038/srep44380
7. Ciani O, Grassi D, Tarricone R. An economic perspective on urinary tract infection: the "costs of resignation". *Clin Drug Invest*. 2013;33(4):255-261. doi:10.1007/s40261-013-0069-x
8. Foxman B. Urinary tract infection syndromes. *Infect Dis Clin North Am*. Mar 2014;28(1):1-13. doi:10.1016/j.idc.2013.09.003
9. François M, Hanslik T, Dervaux B, et al. The economic burden of urinary tract infections in women visiting general practices in France: a cross-sectional survey. *BMC Health Serv Res*. 2016;16(1):365. doi:10.1186/s12913-016-1620-2
10. Moskowitz D, Lee U. Patient distress in women with recurrent urinary tract infections: how can physicians better meet patients needs? *Curr Urol Rep*. 2018;19(12):97. doi:10.1007/s11934-018-0849-1

11. Grigoryan L, Mulgirigama A, Powell M, Schmiemann G. The emotional impact of urinary tract infections in women: a qualitative analysis. *BMC Womens Health*. 2022;22(1):182. doi:10.1186/s12905-022-01757-3
12. Maxwell K, Roberts L, Kramer M, Price J, Newlands A, Finlay KA. Psychosocial burden and healthcare disillusionment in recurrent UTI: a large-scale international survey of patient perspectives. *Front Urol*. 2023;3:1264299. doi:10.3389/fruro.2023.1264299
13. Sosland R, Stewart JN. Management of recurrent urinary tract infections in women: how providers can improve the patient experience. *Urology*. 2021;151:8-12. doi:10.1016/j.urology.2020.06.059
14. Meadows KA. Patient-reported outcome measures: an overview. *Br J Commun Nurs*. 2011;16(3):146-151. doi:10.12968/bjcn.2011.16.3.146
15. Newlands AF, Roberts L, Maxwell K, Kramer M, Price JL, Finlay KA. Development and psychometric validation of a patient-reported outcome measure of recurrent urinary tract infection impact: the recurrent UTI impact questionnaire. *Qual Life Res*. 2023;32:1745-1758. doi:10.1007/s11136-023-03348-7
16. Mokkink LB, De Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-1179. doi:10.1007/s11136-017-1765-4
17. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3
18. Qin S, Nelson L, McLeod L, Eremenco S, Coons SJ. 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(4):1029-1033. doi:10.1007/s11136-018-2076-0
19. Toland MD, Sulis I, Giambona F, Porcu M, Campbell JM. Introduction to bifactor polytomous item response theory analysis. *J Sch Psychol*. 2017;60:41-63. doi:10.1016/j.jsp.2016.11.001
20. Chen FF, Hayes A, Carver CS, Laurenceau JP, Zhang Z. Modeling general and specific variance in multifaceted constructs: a comparison of the bifactor model to other approaches. *J Pers*. 2012;80(1):219-251. doi:10.1111/j.1467-6494.2011.00739.x
21. Dunn KJ, McCray G. The place of the bifactor model in confirmatory factor analysis investigations into construct dimensionality in language testing. *Front Psychol*. 2020;11:1357. doi:10.3389/fpsyg.2020.01357
22. Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press; 2001.
23. Newlands AF, Roberts L, Maxwell K, Kramer M, Price JL, Finlay KA. The Recurrent Urinary Tract Infection Symptom Scale: development and validation of a patient-reported outcome measure. *BJUI Compass*. 2023;4(3):285-297. doi:10.1002/bco2.222
24. Newlands AF, Kramer M, Roberts L, Maxwell K, Price JL, Finlay KA. Confirmatory structural validation and refinement of the recurrent urinary tract infection symptom scale. *BJUI Compass*. 2024;5(2):240-252. doi:10.1002/bco2.297
25. Brubaker L, Chai TC, Horsley H, Khasriya R, Moreland RB, Wolfe AJ. Tarnished gold—the "standard" urine culture: reassessing the characteristics of a criterion standard for detecting urinary microbes. *Front Urol*. 2023;3:1206046. doi:10.3389/fruro.2023.1206046
26. Pickard R, Chadwick T, Oluboyede Y, et al. Continuous low-dose antibiotic prophylaxis to prevent urinary tract infection in adults who perform clean intermittent self-catheterisation: the AnTIC RCT. *Health Technol Assess (Rockv)*. 2018;22(24):1-102. doi:10.3310/hta22240
27. Harding C, Mossop H, Homer T, et al. Alternative to prophylactic antibiotics for the treatment of recurrent urinary tract infections in women: multicentre, open label, randomised, non-inferiority trial. *BMJ*. 2022;376:e068229. doi:10.1136/bmj-2021-0068229
28. Immekus JC, Snyder KE, Ralston PA. Multidimensional item response theory for factor structure assessment in educational psychology research. *Front Educ*. 2019;4:45. doi:10.3389/feduc.2019.00045
29. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007;16(1):5-18. doi:10.1007/s11136-007-9198-0
30. Reeve BB, Fayers PM. Applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers PM, Hays RD, eds. *Assessing Quality of Life in Clinical Trials: Methods and Practice*. 2nd ed. Oxford University Press; 2005.
31. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw*. 2012;48(6):1-29. doi:10.18637/jss.v048.i06
32. Samejima F. The general graded response model. In: Nering M, Ostini R, eds. *Handbook of Polytomous Item Response Theory Models*. Routledge; 2010:77-107.
33. Thissen D, Reeve BB, Bjorner JB, Chang C-H. Methodological issues for building item banks and computerized adaptive scales. *Qual Life Res*. 2007;16(1):109-119. doi:10.1007/s11136-007-9169-5
34. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: identification of local dependence in the rasch model using residual correlations. *Appl Psychol Meas*. 2016;41(3):178-194. doi:10.1177/0146621616677520
35. Davidson M, Keating JL, Eyres S. A low back-specific version of the SF-36 Physical Functioning scale. *Spine*. 2004;29(5):586-594. doi:10.1097/01.brs.0000103346.38557.73
36. ten Klooster PM, TAAL E, van de Laar MAFJ. Rasch analysis of the Dutch health assessment questionnaire disability index and the health assessment questionnaire II in patients with rheumatoid arthritis. *Arthritis Care Res*. 2008;59(12):1721-1728. doi:10.1002/art.24065
37. DeMars CE. Application of the bi-factor multidimensional item response theory model to testlet-based tests. *J Educ Meas*. 2006;43(2):145-168. doi:10.1111/j.1745-3984.2006.00010.x
38. Schneider L, Strobl C, Zeileis A, Debelak R. An R toolbox for score-based measurement invariance tests in IRT models. *Behav Res Methods*. 2022;54(5):2101-2113. doi:10.3758/s13428-021-01689-0
39. Bland JM, Altman DG. Statistics notes: multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170.

40. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 6th ed. Pearson; 2012.
41. Linacre JM. What do infit and outfit, mean-square and standardized mean? *Rasch Meas Trans*. 2002;16(2):871-882.
42. Hsieh C-A, Von Eye AA, Maier KS. Using a multivariate multilevel polytomous item response theory model to study parallel processes of change: the dynamic association between adolescents' social isolation and engagement with delinquent peers in The National Youth Survey. *Multivar Behav Res*. 2010;45(3):508-552. doi:10.1080/00273171.2010.483387
43. Baker FB, Kim S. *The Basics of Item Response Theory Using R*. Springer International Publishing; 2017.
44. Cai L, Monroe SL A new statistic for evaluating item response theory models for ordinal data. *CRESST Report 839*. National Center for Research on Evaluation, Standards, and Student Testing; 2014.
45. Maydeu-Olivares A, Joe H. Assessing approximate fit in categorical data analysis. *Multivar Behav Res*. 2014;49(4):305-328. doi:10.1080/00273171.2014.911075
46. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334. doi:10.1007/BF02310555
47. Kincaid JP, Delionbach LJ. Validation of the automated readability index: a follow-up. *Hum Fact*. 1973;15(1):17-20. doi:10.1177/001872087301500103
48. Widaman KF, Revelle W. Thinking thrice about sum scores, and then some more about measurement and analysis. *Behav Res Methods*. 2023;55(2):788-806. doi:10.3758/s13428-022-01849-w
49. American Psychological Association. Ethical Principles of Psychologists and Code of Conduct. http://www.apa.org/ethics/code/index.html
50. Government Social Research. *Ethics in Social Research: The views of Research Participants*. 2007: P6124.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Newlands AF, Kramer M, Roberts L, Maxwell K, Price JL, Finlay KA. Evaluating the quality of life impact of recurrent urinary tract infection: validation and refinement of the Recurrent UTI Impact Questionnaire (RUTIIQ). *Neurourol Urodyn*. 2024;43:902-914. doi:10.1002/nau.25426