The Recurrent Urinary Tract Infection Symptom Scale (RUTISS): responsiveness to antibiotic treatment and the minimal clinically important difference

Urology 210 (2026) 64–70

Contents lists available at ScienceDirect

Urology

journal homepage: www.sciencedirect.com/journal/urology

Health Services Research

The Recurrent Urinary Tract Infection Symptom Scale: Responsiveness to Antibiotic Treatment and the Minimal Clinically Important Difference

Abigail F. Newlands[a,1], Lydia V. Tidmarsh[a,2], Melissa Kramer[a,b,3], Molly Bradbury[a,b,4], Sarah Snuggs[a,5], Katherine A. Finlay[a,*,6]

[a] School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire, UK
[b] Live UTI Free Ltd, Dublin, Ireland

A B S T R A C T

Objective: To evaluate the responsiveness of the Recurrent Urinary Tract Infection Symptom Scale (RUTISS) to antibiotic treatment and determine the minimal clinically important difference (MCID).

Methods: Female adults with recurrent UTI (N = 108, mean age = 53.6 years) completed a repeated-measures, longitudinal, 40-day naturalistic follow-up study. For participants initiating and completing antibiotic treatment within this period, RUTISS scores were evaluated at baseline (2 days pre-treatment), day 1 of treatment, day 3, and at follow-up (7 days post-treatment). Responsiveness was assessed using effect sizes pre- and post-treatment (rank-biserial correlation), Spearman's rank correlations with anchor measures (Patient Global Impression of Change, PGIC, and UTI Symptom Assessment, UTISA), and receiver operating characteristic analysis. The MCID was triangulated via anchor- and distribution-based methods.

Results: The RUTISS total score demonstrated excellent responsiveness between all timepoints and was strongest between baseline and day 3 (rank-biserial r = 0.52), with strong correlations with the PGIC (Spearman's ρ = 0.73) and UTISA (Spearman's ρ = 0.63). Receiver operating characteristic analysis indicated excellent discrimination for detecting even minimal improvement (AUC = 0.82). The recommended MCID was 6.5 points, exceeding the standard error of measurement. All subscales demonstrated good to excellent responsiveness.

Conclusion: The RUTISS demonstrates excellent responsiveness to antibiotic treatment providing a validated MCID, enabling precise interpretation of treatment effects. As the first fully validated PROM for recurrent UTI, the RUTISS addresses a critical unmet need to assess patient symptom reporting. By providing a responsive and interpretable patient-centered endpoint, the RUTISS enables adequately powered clinical trials and evidence-based symptom monitoring, directly supporting improved patient care and intervention development.

Recurrent urinary tract infection (rUTI) affects approximately 100 million people worldwide each year,[1] generating substantial personal and healthcare burdens with high symptom morbidity and repeated clinical presentations.[1-5] rUTI profoundly affects quality of life, causing anxiety about recurrence, disrupted intimate relationships and work productivity, and reduced wellbeing.[2,5-7] Despite its prevalence and impact, outcome assessment in rUTI management and clinical research has traditionally focused on limited microbiological information which may fail to capture patient-experienced symptom severity and treatment response perceptions.[8,9]

Patient-reported outcome measures (PROMs) provide essential perspectives that complement traditional clinical endpoints. While microbiological culture remains important for pathogen identification and treatment decisions, symptom burden does not reliably correlate with culture results.[9,10] Patient-centered evidence is increasingly crucial in evaluating treatment outcomes,[8,11,12] recognizing that patient perspectives represent a critical dimension to monitor treatment efficacy alongside microbiological assessment. However, few validated measures are available specifically for assessing rUTI, limiting rUTI clinical trial design and real-world treatment evaluation.[11]

The Recurrent UTI Symptom Scale (RUTISS) is currently the only PROM specifically developed to assess rUTI symptom and pain severity.[13,14] Developed following gold-standard COnsensus-based

* Address correspondence to: Katherine A. Finlay, Ph.D., C.Psychol., A.F.B.P.sS., School of Psychology and Clinical Language Sciences, University of Reading, Reading, Berkshire RG6 7BE, UK.

E-mail address: katherine.finlay@reading.ac.uk (K.A. Finlay).

[1] 0000-0002-4718-0075
[2] 0009-0004-0613-1607
[3] 0000-0002-9242-5203
[4] 0009-0005-9125-0818
[5] 0000-0001-5191-9517
[6] 0000-0002-8997-2652

https://doi.org/10.1016/j.urology.2026.01.020
Received 19 November 2025; Accepted 16 January 2026

Standards for the selection of health Measurement INstruments (COSMIN) guidelines and alongside the Recurrent UTI Impact Questionnaire (RUTIIQ),[15,16] the RUTISS is a psychometrically robust assessment tool that is both patient-generated and approved by clinical expertise. A 15-item PROM, the RUTISS quantifies patient-perceived rUTI symptom frequency, change, and severity.[13,14] This rapid assessment tool demonstrates excellent content validity, structural validity, internal consistency, and test-retest reliability.[13,14] However, its responsiveness, the ability to detect change over time following intervention, and minimal clinically important difference (MCID) need to be established to facilitate use as a treatment evaluation tool or clinical trial endpoint.

This study aimed to: (1) evaluate the responsiveness of the RUTISS total score and individual subscales in detecting change following rUTI antibiotic treatment, and (2) determine the MCID for the total score. Detailed in "Statistical Analysis," it was hypothesized a priori that the total score and all subscales would: (H1) demonstrate good responsiveness when comparing to a criterion anchor (AUC > 0.70), (H2) statistically significantly decrease between pre- and post-antibiotic treatment with at least moderate effect sizes ($r \geq 0.3$), and (H3) demonstrate at least moderate convergent validity of responsiveness with related anchor measures ($\rho > 0.30$).[15,17,18]

MATERIALS AND METHODS

Study Design

A repeated-measures, longitudinal observational diary study employed a 40-day design, with female adults with rUTI completing daily online surveys. For participants who initiated antibiotics during the observation period, four assessment timepoints were identified: (1) Baseline, 2 days pre-treatment, (2) day 1 of antibiotic treatment, (3) day 3 of antibiotic treatment, and (4) follow-up, 7 days after treatment completion. The 3-day assessment window aligns with UK-based National Institute for Health and Care Excellence (NICE) guidelines for antibiotic prescription,[19] while the follow-up time period captures sustained treatment response when symptoms typically resolve.[4] This pragmatic design captured naturalistic symptom patterns and treatment decisions, allowing retrospective identification of antibiotic treated episodes occurring at varying timepoints within the 40-day observation period.

Participants and Sampling

A total of 108 adults meeting the European Association of Urology diagnostic criteria for rUTI (≥2 UTIs in 6 months or ≥3 UTIs in 12 months)[20] initiated and completed UTI treatment during the 40-day observation period (see Table 1 for demographic characteristics). Standard urine culture misses up to 58% of infections, potentially leading to dismissal or misdiagnosis.[9,10] Therefore, diagnosis was based on participant-reported history of clinically diagnosed UTI episodes rather than prospective urine testing at enrollment. This approach aligns with the patient-reported focus of PROM validation and is consistent with pragmatic methodologies that balance scientific rigor with real-world applicability and generalizability across diverse clinical experiences.[21] All participants were female, reflecting the primary rUTI population.[1] Exclusion criteria comprised non-rUTI diagnosis, current diagnosis of interstitial cystitis, and pregnancy. This sample exceeds the COSMIN-recommended minimum of 100 participants required for responsiveness.[15]

Participants were recruited via multiple UTI stakeholders from the Alliance for Patient-centred UTI Research (APUR). A snowball sampling approach maximized alternative sources by encouraging participants and UTI clinician networks to share the study information. Aiming to reduce attrition bias and in line with NIHR payment guidance for researchers,[22] retention over the 40-day observation period was supported by reminder emails and incentivized with a $100 (USD) online shopping voucher prize draw for two randomly selected participants.

Table 1
Participant characteristics.

Characteristic	*n*	%
Ethnicity		
Black (including Black British, African American)	3	2.78
Hispanic and/or Latin American	2	1.85
Mixed ethnicity	4	3.70
White (including White Australia, Canada, Europe, Ireland, UK, USA)	99	91.7
Relationship status		
Single	28	25.9
In a relationship	80	74.1
Highest level of education		
Did not graduate high school	2	1.85
High school graduate	18	16.7
Technical/vocational program	4	3.70
Associate's or foundation degree	9	8.33
Bachelor's degree or equivalent	32	29.6
Master's degree or equivalent	37	34.3
Doctoral degree or equivalent	4	3.70
Other	2	1.85
Employment status		
Full-time employed or self-employed	23	21.3
Part-time employed or self-employed	26	24.1
Retired	37	34.3
Homemaker	6	5.56
Student	2	1.85
Other	2	1.85
Not currently working	12	11.1

$N = 108$. Mean age = 53.6 years ($SD = 18.3$).

Procedure and Materials

Following informed e-consent provision, participants completed an online survey comprising a questionnaire about socio-personal demographics and current UTI-related antibiotic use. Participants completed the RUTISS daily, alongside the Patient Global Impression of Change (PGIC) and UTI Symptom Assessment (UTISA) as anchor measures for responsiveness analysis.[13-15,23,24] Email reminders facilitated daily survey completion. Daily survey completion was encouraged but not mandatory; handling of missing data (8.95%) is detailed in Statistical Analysis.

The RUTISS is a 15-item validated PROM assessing symptom frequency (3 items), global change in symptom severity (1 item), and symptom severity (11 items, comprising four subscales: "urinary symptoms", "urinary presentation", "UTI pain and discomfort", and "bodily sensations"). Higher total scores (maximum 100) indicate greater severity. The full 11-item scale and individual subscales demonstrate excellent psychometric performance, with high internal consistency (Cronbach's $\alpha > 0.82$), strong content validity, test-retest reliability (ICC > 0.73), and concurrent validity with related measures (Spearman's $\rho > 0.60$). The RUTISS is available from InSpired Health Outcomes or from the corresponding author.[13,14]

The PGIC is a single-item assessment of patient-perceived change in health over time.[23] The 11-point scale ranges from –5 ("very much worse") to +5 ("very much better"). The 11-point PGIC demonstrates strong discriminative ability and test-retest reliability.[23] A change of ≥2 points is considered clinically meaningful.[23]

The UTISA is a validated 14-item PROM assessing acute UTI symptom severity, demonstrating good reliability and validity.[24] Only the 7-item severity subscale, aligned with RUTISS constructs, was employed (total score range 0-21).[24]

Statistical Analysis

Data Handling and Missing Data

All analyses were computed in R. Statistical assumptions were examined prior to analysis (see Supplementary Material 1), with the linearity assumption and sample size adequacy met, and no problematic outliers detected. Given mild-moderate non-normality (Shapiro-Wilk *P*

$< .05$) and mild violation of the homogeneity assumption (Levene's test $F(2,95) = 3.31$, $P = .041$), non-parametric statistical methods were employed. Missing data (8.95%) were "missing completely at random" (Little's test: $P = .26$).[25] Given non-normal distribution of RUTISS scores, multiple imputation by chained equations was therefore performed using Predictive Mean Matching (PMM) for all variables, generating 10 imputed datasets and pooling results using Rubin's rules.[25-27] A complete case analysis (CCA) was conducted as a sensitivity check ($n = 48$, 44.4% of sample), replicating all analyses using identical methods.

For all approaches, analyses focused on three assessment intervals: Baseline–day 3, day 1–day 3, and Baseline–Follow-up. Intervals from interim timepoints to Follow-up (eg, day 3–Follow-up) were excluded due to symptom resolution restricting variance.

Aim 1: Responsiveness

Adhering to COSMIN recommendations for conducting responsiveness analysis, three complementary approaches were used: (A) Receiver Operating Characteristic (ROC) curve analysis, (B) effect sizes for comparing scores pre- and post-antibiotic treatment, and (C) correlations with existing outcome measures as external anchors.[15]

A) ROC curve analysis
ROC curve analysis assessed the ability of RUTISS change scores to discriminate between participants who improved vs those who did not, using PGIC as the criterion anchor.[15] Participants reporting PGIC scores $\geq +1$ were classified as having experienced at least minimal improvement, as a clinically relevant threshold, and were compared to those with PGIC $< +1$ (no improvement or worsening). Optimal thresholds were calculated using Youden's index to balance sensitivity and specificity.[15,28] Area under the curve (AUC) values quantified the discriminative ability for detecting change, with higher values demonstrating better responsiveness.[17] AUC values greater than 0.70 represent acceptable discrimination, and greater than 0.80 indicate excellent discrimination.[15,17]

B) Effect sizes (pre-post antibiotic treatment)
RUTISS total scores and subscale scores were compared across the three assessment intervals. This approach captures cumulative change during the acute treatment response between Baseline and day 3, and sustained improvement between Baseline and Follow-up.[15] Wilcoxon signed-rank tests provided rank-biserial correlations as an effect size estimate ($r \geq 0.1$ = small effect, ≥ 0.3 = moderate, ≥ 0.5 = large).[29]

C) Correlations with PGIC and UTISA
Correlations between RUTISS change scores (total and subscales) and PGIC ratings and UTISA change scores were computed for the three assessment intervals.[15] This approach assesses the strength and direction of the association between the change detected by the new PROMs and changes in established external anchors. This provides evidence of convergent validity of responsiveness, with stronger correlations indicating that the RUTISS is more sensitive to changes in symptom severity. Spearman's rank correlation coefficients (ρ) were interpreted as $\rho \geq 0.20$ = moderate, $\rho \geq 0.4$ = relatively strong, $\rho \geq 0.60$ = strong.[30]

Aim 2: MCID Identification

The MCID was determined for the RUTISS total score using established methods recommended by the COSMIN initiative.[17,18] The Baseline–day 3 interval was prioritized for MCID derivation as it captures acute treatment response when symptomatic improvement is most expected, aligning with NICE antibiotic course recommendations. This interval also demonstrated the strongest discriminative ability (AUC = 0.82). A triangulation approach incorporated both anchor-based and distribution-based methods to identify the smallest change patients perceive as beneficial.

Anchor-based methods employed the PGIC as the external criterion across three approaches, each applied to the Baseline–day 3 assessment interval (acute treatment phase): (1) mean change in RUTISS total score among participants reporting minimal improvement (PGIC = +1), (2) median change among the same group, and (3) ROC-derived threshold representing the optimal cut-off point for discriminating minimal improvement. Following best-practice, these anchor-based methods were prioritized during triangulation since they reflect patient-centered perspectives on meaningful change.

Distribution-based methods comprised: (1) halving the baseline standard deviation of RUTISS total scores, representing a moderate effect size approximating clinical importance; (2) computing the standard error of measurement (SEM) by multiplying standard deviation by $\sqrt{(1 - \text{Cronbach's } \alpha)}$; and (3) calculating the minimal detectable change at 95% confidence (MDC95) by multiplying 1.96 by $\sqrt{2}$ by SEM, representing the threshold for statistically reliable change.

Estimates across all five methods were compared, ultimately recommending an MCID that prioritises convergence across patient-centered anchor-based methods (*Mdn*). Distribution-based estimates were used as supportive evidence and benchmarks for interpreting the clinical significance and reliability of the recommended MCID.[17,18]

RESULTS

Sample Characteristics

Participants were aged between 19 and 82 years ($M = 53.6$, $SD = 18.3$). All participants reported at least two UTI episodes within the past 6 months, and a third reported more than three ($n = 38$, 35.2%). The average RUTISS total score at Baseline was 23.0 ($SD = 18.8$), with scores ranging from 0 to 78.2 (maximum possible score = 100). This wide range reflects heterogeneous rUTI symptom presentation; some individuals experience few but severe symptoms while others report multiple milder symptoms. The ability to capture the variation across the full spectrum of rUTI severity indicates the high sensitivity of the RUTISS. At day 3 and Follow-up, average total scores reduced to 13.6 ($SD = 15.6$) and 11.2 ($SD = 14.2$), respectively. The average Baseline UTISA total score was 6.76 ($SD = 4.29$), with scores ranging between 0 to 18 (maximum possible score = 21). UTISA scores similarly reduced to 4.34 ($SD = 3.99$) at day 3 and 4.09 ($SD = 3.90$), at Follow-up. Analysis of PGIC scores indicated that 37.0% of participants ($n = 40$) reported at least minimal improvement (PGIC $\geq +1$) in symptom severity between Baseline–day 3, while the same proportion indicated no change ($n = 40$, 37.0%). All RUTISS, UTISA and PGIC descriptive statistics for each timepoint are available in Supplementary Material 2.

Aim 1: Responsiveness

ROC Curve Analysis

Analysis demonstrated that RUTISS change scores effectively discriminated between participants with at least minimal improvement (PGIC $\geq +1$) from those without improvement across most scales and time intervals (see Table 2).

For RUTISS total scores, AUC values were excellent across all three intervals, exceeding the pre-specified threshold of 0.70 and supporting H1 ($M = 0.79$, $SD = 0.04$). The highest discrimination was observed for the Baseline–day 3 interval which indicated excellent responsiveness with an optimal threshold of 6.26 points (AUC = 0.82).

For subscales, discrimination varied by domain. The "urinary symptoms" subscale demonstrated excellent discrimination across all intervals (AUC $M = 0.74$, $SD = 0.03$), and the "UTI pain and/or discomfort" subscale demonstrated similarly strong performance across all intervals (AUC $M = 0.76$, $SD = 0.08$). The "urinary presentation" subscale demonstrated strong discrimination between Baseline–day 3 and Baseline–Follow-up (AUC $M = 0.74$, $SD = 0.02$), however less strong discriminative ability between day 1–day 3 (AUC = 0.57). The

Table 2
RUTISS receiver operating characteristic curve analysis.

Interval	AUC (95% CI)	Optimal Threshold	Sensitivity	Specificity	Youden's Index
RUTISS total score					
Baseline–day 3	0.82 (0.73,0.91)*	6.36	0.70	0.86	0.55
Day 1–day 3	0.74 (0.64,0.84)*	3.64	0.71	0.75	0.45
Baseline–Follow-up	0.81 (0.69,0.93)*	10.5	0.63	0.88	0.50
Urinary symptoms					
Baseline–day 3	0.77 (0.68,0.86)*	2.50	0.61	0.81	0.42
Day 1–day 3	0.71 (0.60,0.81)*	1.50	0.68	0.70	0.38
Baseline–Follow-up	0.75 (0.62,0.89)*	4.50	0.58	0.83	0.42
Urinary presentation					
Baseline–day 3	0.71 (0.61,0.81)*	2.50	0.50	0.91	0.41
Day 1–day 3	0.57 (0.46,0.67)	0.50	0.97	0.16	0.13
Baseline–Follow-up	0.75 (0.62,0.88)*	0.50	0.67	0.71	0.38
UTI pain and/or discomfort					
Baseline–day 3	0.82 (0.73,0.90)*	0.50	0.77	0.79	0.55
Day 1–day 3	0.67 (0.55,0.78)	1.50	0.53	0.82	0.35
Baseline–Follow-up	0.80 (0.68,0.92)*	3.50	0.54	0.96	0.50
Bodily sensations					
Baseline–day 3	0.58 (0.47,0.69)	0.50	0.45	0.74	0.18
Day 1–day 3	0.68 (0.56,0.79)	0.50	0.53	0.80	0.33
Baseline–Follow-up	0.65 (0.50,0.81)	0.50	0.96	0.33	0.29

AUC, area under the curve.
Optimal thresholds reflect the magnitude of RUTISS score reduction in points that best discriminates respondents with at least minimal improvement since starting treatment from those without any improvement.
Timepoint interpretation: Baseline = 2 days before starting antibiotic treatment; day 1 = day 1 of antibiotic treatment; day 3 = day 3 of antibiotic treatment; Follow-up = 7 days after completing antibiotic treatment.
*AUC greater than 0.70 criterion set for acceptable discrimination.

"bodily sensations" subscale showed moderate discriminative ability (AUC M = 0.64, SD = 0.05), likely reflecting the broader, less UTI-specific nature of these symptoms.

Optimal thresholds varied by scale and time interval, given different score ranges and response patterns within each domain. Sensitivity and specificity were generally well-balanced, with most thresholds (90%) achieving sensitivity > 0.50 and specificity > 0.70.

Effect Sizes (Pre-Post Antibiotic Treatment)

As hypothesized (H2), RUTISS total scores decreased significantly between all timepoints with moderate-large effect sizes (see Table 3). The RUTISS total score therefore demonstrated excellent responsiveness to change before and after antibiotic treatment during the acute treatment phase (Baseline–day 3) and sustained improvement (Baseline–Follow-up).

The "urinary symptoms," "urinary presentation," and "UTI pain and/or discomfort" subscale scores also statistically significantly decreased between all timepoints as hypothesized (rank-biserial r M = 0.42, SD = 0.10). The "bodily sensations" subscale scores significantly decreased from Baseline–day 1 and Baseline–Follow-up (r M = 0.30, SD = 0.08), and non-significantly between day 1–day 3 (r = 0.15, P = .14), consistent with the gradual resolution of systemic symptoms following acute infection.

Correlations With PGIC and UTISA

The RUTISS total change scores for Baseline–day 3, day 1–day 3, and Baseline–Follow-up significantly correlated moderately-strongly with both UTISA change scores (see Table 3; Spearman's ρ M = 0.72, SD = 0.08) and PGIC ratings (ρ M = –0.58, SD = 0.12), meeting H3 and demonstrating excellent responsiveness to antibiotic treatment-related improvement.

All four subscale change scores for all intervals significantly correlated moderately-strongly with the UTISA change scores as hypothesized (ρ M = 0.54, SD = 0.18). All four subscale change scores demonstrated meaningful correlations with PGIC ratings (ρ M = 0.43, SD = 0.15). Most coefficients (10/12) exceeded the pre-specified 0.30 threshold, supporting H3.

Aim 2: MCID Identification

The MCID for the RUTISS total score was computed via triangulation of six estimation methods (see Table 4).

Anchor-based Methods

The three methods demonstrated strong agreement. The mean change among participants reporting minimal improvement was 6.36 points, while the median change was 4.09 points. The ROC-derived optimal threshold from the Baseline–day 3 interval, which best discriminated minimal improvement (Youden's index = 0.55), was 6.36 points. The median was therefore 6.36 points.

Distribution-based Methods

Providing statistical context to support interpretation, the half standard deviation criterion (9.10 points) represented a moderate effect size. The SEM (5.75 points) reflected measurement precision. The MDC95 (16.0) indicated the threshold for change that can be distinguished from measurement error with 95% confidence.

Triangulation

Prioritizing patient-centered anchor-based methods facilitated recommendation of a final MCID for the RUTISS total score of 6.5 points. This estimate is supported by the convergence of: the mean change in minimally improved patients (6.36 points), the ROC-derived optimal threshold (6.36), and proximity to the SEM (5.75). On average, with a mean reduction of 9.76 points between Baseline and day 3 (SD = 16.4), this sample perceived clinically meaningful improvement following antibiotic treatment.

The MCID exceeds the SEM (5.75), suggesting it captures change beyond typical measurement variability as expected.[18] The MCID and MDC95 serve complementary purposes in interpreting RUTISS change. The MCID of 6.5 points represents patient-perceived meaningful change and is appropriate for group-level research applications, treatment effectiveness evaluation, and clinical screening.[18] For individual patient monitoring where maximum statistical confidence is desired, the MDC95 of 16 points provides a more conservative threshold ensuring 95% certainty that observed changes exceed measurement error.[18]

Table 3
RUTISS responsiveness: comparison before and after treatment, and comparison with other measures.

Time Interval	*M* Change	*SD* Change	*Mdn* Change	Wilcoxon *V*	Rank-Biserial Correlation *r* (95% CI)	Spearman's ρ Correlation With UTISA (95% CI)	Spearman's ρ Correlation With PGIC (95% CI)
RUTISS total score							
Baseline–day 3	−9.28	17.65	−4.55	979.6	0.52 (0.38,0.66)*	0.73 (0.59,0.84)*	−0.63 (−0.76, −0.46)*
Day 1–day 3	−4.22	10.2	−2.73	1087.9	0.43 (0.27,0.58)*	0.64 (0.48,0.77)*	−0.45 (−0.60, −0.28)*
Baseline–Follow-up	−11.39	18.89	−8.04	1068.9	0.53 (0.39,0.67)*	0.80 (0.71,0.87)*	−0.67 (−0.82, −0.44)*
"Urinary symptoms" subscale							
Baseline–day 3	−4.23	7.5	−2.00	411.5	0.52 (0.38,0.67)*	0.70 (0.57,0.81)*	−0.54 (−0.69, −0.36)*
Day 1–day 3	−1.67	4.51	−1.00	742.5	0.37 (0.21,0.54)*	0.63 (0.47,0.76)*	−0.34 (−0.51, −0.15)*
Baseline–Follow-up	−4.62	9.53	−2.00	137.0	0.50 (0.28,0.72)*	0.79 (0.64,0.88)*	−0.53 (−0.72, −0.25)*
"Urinary presentation" subscale							
Baseline–day 3	−2.73	6.34	0.00	329.0	0.44 (0.28,0.61)*	0.36 (0.15,0.55)*	−0.43 (−0.60, −0.24)*
Day 1–day 3	−1.54	3.67	0.00	292.0	0.44 (0.28,0.59)*	0.30 (0.09,0.48)*	−0.20 (−0.36, −0.01)*
Baseline–Follow-up	−3.50	6.61	0.00	30.0	0.56 (0.37,0.74)*	0.51 (0.27,0.71)*	−0.59 (−0.75, −0.35)*
"UTI pain and/or discomfort" subscale							
Baseline–day 3	−2.57	5.34	−1.00	391.0	0.48 (0.33,0.64)*	0.71 (0.58,0.82)*	−0.60 (−0.74, −0.45)*
Day 1–day 3	−0.67	3.21	0.00	721.0	0.25 (0.03,0.39)*	0.52 (0.34,0.66)*	−0.35 (−0.52, −0.15)*
Baseline–Follow-up	−1.90	5.79	−1.00	203.5	0.35 (0.10,0.60)*	0.79 (0.61,0.90)*	−0.63 (−0.79, −0.41)*
"Bodily sensations" subscale							
Baseline–day 3	−1.20	4.23	0.00	298.0	0.34 (0.16,0.51)*	0.36 (0.15,0.53)*	−0.22 (−0.42, −0.02)*
Day 1–day 3	−0.20	3.66	0.00	566.0	0.15 (−0.03,0.31)	0.41 (0.20,0.58)*	−0.32 (−0.50, −0.10)*
Baseline–Follow-up	−1.94	5.03	0.00	112.0	0.35 (0.11,0.60)*	0.44 (0.17,0.65)*	−0.36 (−0.60, −0.08)*

CI, confidence interval; *M*, mean; *Mdn*, median; PGIC, Patient Global Impression of Change; RUTISS, Recurrent Urinary Tract Infection Symptom Scale; *SD*, standard deviation; UTISA, Urinary Tract Infection Symptom Assessment (change score).
N = 108.
Interpretation of rank-biserial correlation (*r*): $r \geq 0.1$ = small effect, ≥ 0.3 = moderate, ≥ 0.5 = large.[29]
Interpretation of Spearman's rank correlation (ρ): $\rho \geq 0.10$ = negligible correlation, $\rho \geq 0.20$ = moderate, $\rho \geq 0.4$ = relatively strong, $\rho \geq 0.60$ = strong.[30]
Timepoint interpretation: Baseline = 2 days before starting antibiotic treatment; day 1 = day 1 of antibiotic treatment; day 3 = day 3 of antibiotic treatment; Follow-up = 7 days after completing antibiotic treatment.
Note that negative RUTISS change scores indicate symptom improvement (ie, reduction in RUTISS scores and therefore reduction in severity).
*$P < .001$.

Table 4
MCID triangulation.

Method	Estimate	Interpretation
Anchor-based		
1. Mean RUTISS change in participants with minimal improvement (PGIC = +1)	6.36	Average change in minimally improved group
2. Median RUTISS change in participants with minimal improvement (PGIC = +1)	4.09	Typical individual improvement in minimally improved group
3. ROC-derived threshold	6.36	Optimal discriminative threshold
Distribution-based		
1. 0.5 *SD* baseline	9.10	Moderate effect size threshold
2. SEM ($SD \times \sqrt{(1 - \text{Cronbach's } \alpha)}$)	5.75	Measurement precision estimate. Utilizing Cronbach's α = 0.90 (from Newlands et al.)
3. MDC95	16.0	Threshold for reliable change at 95% confidence
Triangulation		
Anchor-based median	6.36	Primary estimate: patient-centered
Recommended MCID	6.5 points	Exceeds measurement error, appropriate for group-level evaluation

MCID, minimal clinically important difference; MDC95, minimal detectable change at 95% confidence; PGIC, Patient Global Impression of Change; ROC, Receiver Operating Characteristic; RUTISS, Recurrent Urinary Tract Infection Symptom Scale; *SD*, standard deviation; SEM, standard error of measurement.
The recommended MCID is based on triangulation prioritizing patient-centered anchor-based methods in accordance with best-practice. The MCID exceeds the SEM, indicating change beyond typical measurement variability.

Complete Case Analysis

CCA yielded results highly consistent with multiple imputation, supporting robustness to missing data assumptions (see Supplementary Material 3).

DISCUSSION

This study supports the Recurrent UTI Symptom Scale (RUTISS) as a responsive patient-reported outcome measure (PROM) for rUTI treatment evaluation. Analysis demonstrated strong effect sizes and correlations with established anchors, and excellent discriminative ability. Patient-centered triangulation identified an MCID of 6.5 points, validated by convergent evidence from multiple approaches. These findings enable interpretation of treatment effects and enhanced patient monitoring and establish the RUTISS as a valuable complement to traditional microbiological endpoints in rUTI research and clinical trials.

The RUTISS demonstrated excellent responsiveness across multiple assessment approaches. Effect sizes indicate meaningful symptom reduction during acute treatment, while strong correlations with PGIC and UTISA confirm convergent validity.[15,17] The high discrimination establishes the RUTISS as sensitive to patient-perceived improvement.[17] Subscale performance was generally high, with "urinary symptoms", "urinary presentation", and "UTI pain and discomfort" showing the strongest responsiveness. "Bodily sensations" demonstrated more modest discrimination, likely reflecting broader, less rUTI-specific symptoms that respond more gradually to treatment.

The MCID of 6.5 points provides a practical threshold for clinical decision-making and trial endpoints. Convergence of patient-centered methods strengthens confidence in this estimate.[18] The MCID represents meaningful patient-perceived improvement, suitable for treatment effectiveness evaluation and clinical screening. The MDC95 (16 points) offers a complementary threshold for individual patient monitoring, enabling clinicians to distinguish true clinical change from measurement variability with 95% confidence when making treatment decisions.[18]

Despite many strengths, several study limitations warrant consideration. Firstly, the MCID was determined in the context of antibiotic treatment, thus its applicability to other interventions requires evaluation. Further, while the sample exceeded COSMIN recommendations, it is acknowledged that all participants were female, and most identified as White, resided in more developed countries, and reported a high level of education. Additional assessment in diverse samples, including in males, non-White populations, and lower socioeconomic groups, would strengthen the application of the measure in broader demographic contexts. Future research will seek cross-cultural validation of the RUTISS, and translations into non-English languages and dialects continue to be developed. Further, future research is necessary to examine RUTISS performance across clinically defined subgroups, including collecting detailed information about menopausal status and comorbidities.

The naturalistic design prioritized ecological validity, enhancing generalizability. Microbiological confirmation, standardized antibiotic regimens, and assessment of urinary analgesic use were not directly captured in this study due to the community-based sample. Overlapping conditions (eg, genitourinary syndrome of menopause, bladder pain syndromes) cannot be excluded, and the impact of antibiotic treatment may be multifactorial, including anti-inflammatory and placebo effects. However, MCID and responsiveness testing assess the ability of the RUTISS to detect patient-perceived change, and is not contingent on the mechanisms or causes for that change.[15,17] MCID and responsiveness findings are indicative of minimal or greater change in symptoms, across the full range of the patient experience. It is established that standard urine culture misses up to 58% of true infections.[9,10] Indeed, in the UK, antibiotic treatment can be prescribed based on symptom reporting only.[19] While the naturalistic approach employed by this study reflects authentic patient experiences across diverse clinical contexts, essential for development of patient-centered tools, it necessarily precluded verification of culture-confirmed UTI episodes or standardized treatment regimens. However, this pragmatic methodology importantly captures the patient population and patient-led symptom reporting embodied within current guidelines. Controlled studies incorporating urine culture are ongoing and will complement these findings.

The RUTISS fills a critical gap in rUTI assessment by providing a validated, responsive PROM that captures patient-experienced symptom burden alongside microbiological endpoints. The established MCID facilitates simple interpretation of treatment responses and sample size calculations for clinical trials, addressing a key barrier to adequately powered studies. For clinical practice, the RUTISS offers a brief, validated tool for systematically evaluating symptom severity and monitoring patient outcomes, improving patient-clinician communication and shared decision-making.[8,12] An overview of key RUTISS applications is provided in Supplementary Material 4.

CONCLUSION

This study establishes the RUTISS as a responsive and clinically interpretable patient-reported outcome measure for rUTI symptom severity. Results from this study demonstrate the measure's excellent responsiveness to antibiotic treatment and a validated MCID of 6.5 points, with particular benefit for the assessment of patient-reported outcomes in females living with rUTI. The 15-item RUTISS is a robust, yet brief tool for evaluating treatment effectiveness from the patient perspective. Representing an important advancement in patient-centered rUTI care and research, the RUTISS has critical utility for clinical assessment and monitoring and is ready for implementation as clinical trial endpoints.

Disclosures

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Declaration

This study was approved by the University of Reading School of Psychology and Clinical Language Sciences Research Ethics Committee (reference: 2023-129-KF). All participants provided electronic informed consent prior to participation.

CRediT Authorship Contribution Statement

Abigail F. Newlands: Writing – review & editing, Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Lydia V. Tidmarsh**: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Melissa Kramer**: Writing – review & editing, Resources, Project administration, Methodology, Investigation, Conceptualization. **Molly Bradbury**: Writing – review & editing, Resources, Project administration, Methodology, Investigation, Conceptualization. **Sarah Snuggs**: Writing – review & editing, Supervision, Conceptualization. **Katherine A. Finlay**: Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

Melissa L. Kramer: CEO of Live UTI Free Ltd, however, no financial incentives have been received. The other authors have no conflict of interest to declare.

Acknowledgment

The authors thank Live UTI Free, Chronic UTI Australia, CUTIC and Let's Talk UTI from the Alliance for Patient-centred UTI Research (APUR) for supporting this study and recruitment of participants.

Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.urology.2026.01.020.

References

1. 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;40:755–763. https://doi.org/10.1007/s00345-021-03913-0
2. 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). *Expert Rev Pharmacoecon Outcomes Res.* 2018;18:107–117. https://doi.org/10.1080/14737167.2017.1359543
3. 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;16:365. https://doi.org/10.1186/s12913-016-1620-2
4. Colgan R, Keating K, Dougouih M. Survey of symptom burden in women with uncomplicated urinary tract infections. *Clin Drug Investig.* 2004;24:55–60. https://doi.org/10.2165/00044011-200424010-00007
5. Maxwell K, Roberts L, Kramer M, Price JL, Newlands AF, Finlay KA. Psychosocial burden and healthcare disillusionment in recurrent UTI: a large-scale international survey of patient perspectives. *Front Urol.* 2023;3:1264299. https://doi.org/10.3389/fruro.2023.1264299
6. Newlands AF, Kramer ML, Maxwell K, Price JL, Finlay KA. The mediating role of coping in the relationship between perceived health and psychological wellbeing in recurrent urinary tract infection: the rUTI Illness Process Model. *Health Psychol Behav Med.* 2024;12:2420806. https://doi.org/10.1080/21642850.2024.2420806
7. 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;7:44380. https://doi.org/10.1038/srep44380
8. Sosland R, Stewart JN. Management of recurrent urinary tract infections in women: how providers can improve the patient experience. *Urology.* 2020;51:8–12. https://doi.org/10.1016/j.urology.2020.06.059
9. 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. https://doi.org/10.3389/fruro.2023.1206046
10. 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. https://doi.org/10.1136/bmj-2021-0068229
11. Harrison NL, Day AW, Tandogdu Z, Bonkat G, Somani BK. Atlas of patient-reported outcome measures, nomograms and scoring systems used in simple and complicated urinary tract infections: a systematic review. *Ther Adv Infect Dis.* 2025;12:20499361251328258. https://doi.org/10.1177/20499361251328258
12. van Horrik TM, Colliers A, Blanker MH, et al. Qualitative study on shared decision making in cystitis management in general practice. *BJGP Open.* 2024;8:BJGPO.2023.0179. https://doi.org/10.3399/bjgpo.2023.0179
13. 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;4:285–297. https://doi.org/10.1002/bco2.222
14. 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;5:240–252. https://doi.org/10.1002/bco2.297
15. Mokkink L.B., Prinsen C., Patrick D.L., et al. COSMIN Study Design checklist for patient-reported outcome measurement instruments. Amsterdam, The Netherlands. 2019;2019:1-32.
16. 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. https://doi.org/10.1002/nau.25426
17. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018;27:1147–1157. https://doi.org/10.1007/s11136-018-1798-3
18. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102–109. https://doi.org/10.1016/j.jclinepi.2007.03.012
19. National Institute for Health and Care Excellence. Urinary tract infection (lower): antimicrobial prescribing (NICE guideline NG109); 2018. Accessed 11 November, 2025. ⟨https://www.nice.org.uk/guidance/ng109⟩.
20. 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.
21. Scales Jr.CD, Penson DF. Clinical trials in urology. In: Dahm P, Dmochowski RR, eds. *Evidence-Based Urology.* Wiley; 2018:19–24.
22. Mummolo J, Peterson E. Demand effects in survey experiments: an empirical assessment. *Am Pol Sci Rev.* 2019;113:517–529. https://doi.org/10.1017/S0003055418000837
23. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther.* 2009;17:163–170. https://doi.org/10.1179/jmt.2009.17.3.163
24. Clayson D, Wild D, Doll H, Keating K, Gondek K. Validation of a patient-administered questionnaire to measure the severity and bothersomeness of lower urinary tract symptoms in uncomplicated urinary tract infection (UTI): the UTI Symptom Assessment questionnaire. *BJU Int.* 2005;96:350–359. https://doi.org/10.1111/j.1464-410x.2005.05630.x
25. Little RJA. A test of missing completely at random for multivariate data with missing values. *J Am Stat Assoc.* 1988;83:1198–1202. https://doi.org/10.1080/01621459.1988.10478722
26. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* John Wiley & Sons; 1987.
27. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med.* 2011;30:377–399. https://doi.org/10.1002/sim.4067
28. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3
29. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed Psychology Press; 1988.
30. Rea LM, Parker RA. *Designing and Conducting Survey Research: A Comprehensive Guide.* Jossey-Bass; 1992.