Research and Reports in Urology

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Psychometric Validation of the QLQ-NMIBC24 in Low Grade, Intermediate Risk Non-Muscle Invasive Bladder Cancer

Charles C Peyton[1], Rahul Dhanda[2], Tom Burke[3,4], Victoria Tsurutis[2], Nikky Ugwuoke[2], Anagha Nadkarni[2], Brent Burger[2], Veronica Lee[2], Michael J Louie[2], Angela M Stover[5,6]

[1]Department of Urology, University of Alabama at Birmingham, Birmingham, AL, USA; [2]Urogen Pharma, Princeton, NJ, USA; [3]Prime HCD, London, UK; [4]Faculty of Health and Social Care, University of Chester, Chester, UK; [5]Department of Health Policy and Management, University of North Carolina, Chapel Hill, NC, USA; [6]Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA

Correspondence: Charles C Peyton, Department of Urology, Urologic Oncology, University of Alabama at Birmingham, 1720 2nd Ave. South, FOT 1120, Birmingham, AL, 35294-3411, USA, Tel +1 205 996 4315, Email cpeyton@uabmc.edu

Purpose: This study aimed to evaluate the psychometric properties of the EORTC QLQ-NMIBC24 and to propose distribution-based minimal clinically important difference (MCID) thresholds in patients with low-grade, intermediate-risk NMIBC (LG-IR-NMIBC).

Patients and Methods: Patients with LG-IR-NMIBC from two phase 3 trials (ATLAS, n = 270; ENVISION, n = 240) completed the EORTC QLQ-C30 (ENVISION only) and QLQ-NMIBC24 questionnaires at baseline, Week 6, and Month 3. Psychometric evaluation of the QLQ-NMIBC24 in the LG-IR-NMIBC population included internal consistency, item convergence, known-groups validity, test-retest reliability and responsiveness to clinical change.

Results: Item-item (0.40–0.96) and item-scale correlations (0.44–0.96) indicated good convergent validity across most domains. Internal consistency was acceptable for all multi-item domains (Cronbach's α 0.78–0.93) except Malaise, which showed lower reliability (Cronbach's α 0.34–0.67). Known-groups comparisons supported the instrument's ability to distinguish patients by physical function (effect sizes up to 1.63) and sex-related domains. Test-retest reliability was generally good for Urinary Symptoms and Sexual Function (ICC 0.79–0.82). Selected domains (Urinary Symptoms, Sexual Intimacy, Malaise) demonstrated responsiveness to clinical change. Distribution-based MCID estimates varied across domains (4.37–16.29), providing potential thresholds for meaningful changes in HRQoL scores. Anchor-based analyses using QLQ-C30 change scores were explored, but correlations with QLQ-NMIBC24 domains (r = –0.37 to 0.01) did not meet the minimum threshold (0.37).

Conclusion: The QLQ-NMIBC24 demonstrates valid, reliable, and interpretable measurement properties in patients with LG-IR-NMIBC. Although sensitivity varied across domains, the findings support its use in clinical trials and potentially routine practice. Distribution-based MCID thresholds provide guidance for interpreting meaningful HRQoL changes, though further studies using anchor-based methods are warranted.

Keywords: bladder cancer, LG-IR-NMIBC, HRQoL, EORTC questionnaires, psychometric validation, minimal clinically important difference

Introduction

Non–muscle-invasive bladder cancer (NMIBC) accounts for approximately 75% of newly diagnosed bladder cancer cases.[1,2] NMIBC is classified by histological grade as either low grade (LG) or high grade (HG),[3] and by risk of recurrence/progression as low, intermediate, or high risk.[2] LG intermediate-risk (LG-IR) NMIBC includes LG tumors that do not meet the criteria for low risk, such as multiple or large (>3 cm) LG Ta tumors, or recurrent LG Ta tumors without carcinoma in situ (CIS).[4,5]

LG-IR-NMIBC is typically managed with transurethral resection of the bladder tumor (TURBT)[6–8] which can be followed by a single dose of adjuvant intravesical chemotherapy, or office based biopsies and fulgurations.[9–11] Recently,

https://doi.org/10.2147/RRU.S591097
Received: 24 December 2025
Accepted: 22 February 2026
Published: 18 March 2026

UGN-102, an intravesical mitomycin-containing reverse thermal gel, has been approved as an additional chemoablative treatment option for recurrent LG-IR-NMIBC.[12]

Patient-reported outcomes (PROs) provide direct measures of patient based disease impact perspectives, including symptom burden, physical and social functioning, and overall health-related quality of life (HRQoL).[13,14] The European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire core 30 (QLQ-C30) is a widely used questionnaire to assess HRQoL in cancer patients across several tumor types.[15] To address NMIBC-specific concerns, the EORTC developed the QLQ-NMIBC24, a complementary module tailored for NMIBC patients to be administered alongside the QLQ-C30.[16]

Psychometric validation of the QLQ-NMIBC24 has been conducted in UK patients with high- and intermediate-risk NMIBC,[17] in the Netherlands in patients with high-risk NMIBC,[18] and in Denmark and South Korea in patients with NMIBC receiving TURBT,[19,20] demonstrating its reliability and validity across populations. Psychometric properties of PRO instruments may differ in LG-IR-NMIBC compared with higher-risk populations due to differences in symptom burden, treatment and changes in health status, which could influence PROs item relevance, score distribution and responsiveness.[21,22] Furthermore, establishing minimal clinically important difference (MCID) thresholds for LG-IR-NMIBC is clinically relevant, as they provide guidance for interpreting whether observed changes in QLQ-NMIBC24 scores reflect meaningful improvements or worsening from the patient's perspective.

However, the psychometric properties of the QLQ-NMIBC24 have not yet been specifically evaluated in patients with LG-IR-NMIBC. This study evaluates psychometric properties (validity, reliability, and responsiveness) of the EORTC QLQ-NMIBC24 in patients with LG-IR-NMIBC and proposes distribution-based MCID thresholds for this patient population.

Materials and Methods

Quality of Life Questionnaires

The EORTC QLQ-C30 is a widely used questionnaire to assess HRQoL in cancer patients across several tumor types.[15] QLQ-C30 consists of 30 items covering functional domains, symptom scales, and global health status, with established scoring procedures and validated properties.[15,23,24]

The QLQ-NMIBC24 was developed to address NMIBC-specific concerns and is administered alongside the QLQ-C30.[16] The QLQ-NMIBC24 includes 24 items spanning 11 domains, such as urinary symptoms, malaise, intravesical treatment-related issues, sexual function, and future worries, with separate scoring for functional and symptom scales.[16]

In both QLQ-NMIBC24 and QLQ-C30, items are scored using a 1 (not at all) to 4 (very much) Likert scale, except the global health status/QoL in QLQ-C30, which has a 1–7 scale. Scales and single-item measures' scores are linearly transformed to range from 0 to 100. In the QLQ-C30, higher functional scales scores indicate higher level of functioning ("better") and in symptom scales higher scores indicate more symptoms ("worse").[25] In the NMIBC24, higher scores indicate lower level of functioning ("worse") in functional scales and more symptoms in symptom scales ("worse").

Data

This study used data from two phase 3 clinical studies conducted in patients with LG-IR-NMIBC: ATLAS[26] and ENVISION.[27]

ATLAS was a randomized, controlled, open-label phase 3 study. Participants with LG-IR-NMIBC were recruited across 72 sites in 10 countries and randomized to either (i) UGN-102 (mitomycin) for intravesical solution, administered once weekly for 6 weeks, with or without TURBT (UGN-102 was the primary treatment, and TURBT was performed in cases of residual disease at the 3 month assessment) or (ii) TURBT alone. After randomization, participants had follow-up clinic visits every three months. Assessment of patient reported outcome (PRO) measures was included as a secondary study objective. ATLAS received Institutional Review Board approval (IRB No. BL006) and was registered with ClinicalTrials.gov (NCT04688931).[28]

ENVISION was a single arm phase 3 study. Participants with LG-IR-NMIBC were recruited across 56 locations in 10 countries. Participants received UGN-102 intravesical instillations once weekly for six weeks, followed by follow-up

clinic visits every three months from the start of the treatment. Assessment of PRO measures was included as an exploratory study objective. ENVISION received Institutional Review Board approval (IRB Tracking Number: 20216408) and was registered with ClinicalTrials.gov (NCT05243550).[29] ENVISION included both the QLQ-C30[15] and the QLQ-NMIBC24[16] questionnaires, whereas ATLAS included only the QLQ-NMIBC24.[16] Questionnaires were administered at baseline and at treatment and follow-up visits.

In ENVISION, missing values were handled following the EORTC scoring manual:[25] (i) if at least 50% of the items in a scale were answered, the scale scores were calculated according to the standard equations given on the manual (any items with missing values will be ignored); (ii) if less than 50% of the items were answered, the score was set to missing; (iii) in single-item measures the score was set to missing. In ATLAS, scale scores were only calculated if all items had non-missing answers. The difference in missing data handling reflects variations in the respective study protocols and statistical analysis plans, which were developed independently for each trial.

This study used all available data, regardless of the participants' adherence to treatment, collected at baseline, 6 weeks (UGN-102 arms only) and three months. Although psychometric analyses in this study focused on these time points, PROs data were also collected at later follow-up visits as part of the trial protocols.[26,27] All data analyses were performed using STATA version 18.[30]

Psychometric Evaluation Methods

Convergent Validity

Convergent validity was evaluated using polychoric correlations at two levels. First, item-item correlations were examined to assess the association between individual items within the same scale. Second, item-scale correlations were calculated using item-omitted correlations, which correlate each item with the total score of its own scale after removing that item. Correlations ≥ 0.40 were considered indicative of convergent validity, in line with previous studies.[17,18,20]

Internal Consistency

Internal consistency was assessed using Cronbach's alpha, which measures the average strength of association among all possible pairs of items within a domain.[31] A threshold of 0.7 was used, with higher values indicating internal consistency, in line with previous studies.[17–20]

Known-Groups Validity

Two groups based on the baseline score on the QLQ-C30's physical function scale (PFS) were defined: PFS < 90 and PFS ≥ 90.[17] Participants in the PFS ≥ 90 group were expected to be "healthier", with higher scores on functional scales and lower scores on symptoms scales in the QLQ-C30, and lower scores in the NMIBC24 domains. Additionally, analysis by sex (male vs female)[17] was performed to assess the absence of differences across general domains and to explore potential differences in domains more likely to be sex-related.

Differences in mean scores were examined using *t*-tests assuming unequal variances. Effect sizes comparing the difference in means for the two groups defined by PFS were calculated; values of 0.5 and 0.8 were considered indicative of moderate and large effects, respectively.[32]

Test-Retest Reliability

Test-Retest Reliability was evaluated using QLQ-NMIBC24 scores at baseline ("test") and week 6/month 3 ("re-test") from participants who reported no change in the QLQ-C30 global QoL domain between time points (test and re-test). Individual intraclass correlation coefficients (ICCs) were calculated using subjects as targets and visits as raters. ICC values were interpreted based on previously published thresholds: < 0.5 indicated poor, 0.5–0.75 moderate, 0.75–0.9 good and > 0.9 excellent reliability.[33]

Responsiveness to Change

To investigate whether QLQ-NMIBC24 scores reflect clinical change, analysis of covariance (ANCOVA) models were applied, using the change from baseline in QLQ-NMIBC24 domains as the dependent variable, and a clinical indicator of

change (0 = no; 1 = yes) and the baseline domain score as explanatory variables. Specifically, the clinical indicator of change was Complete Response (CR; ie, having no detectable disease and having received no further treatment) at month 3.

Meaningful Change

Meaningful change from the patient's perspective, referred to as the minimal clinically important difference (MCID) or minimal important change (MIC), can be estimated using distribution-based and anchor-based approaches.[34]

The distribution-based approach uses estimates derived from the observed distribution of scores, including half of the standard deviation (SD) and the standard error of measurement (SEM), calculated as:

$$SEM = SD \times \sqrt{(1 - r)}$$

where r is the ICC obtained from test-retest reliability.[34]

The anchor-based approach uses an external indicator, either clinical or patient-reported, to identify patients who experience minimal change (eg, "a little better" or "a little worse") and estimates the MCID from their mean PRO change.[34,35] Anchors should assess a construct similar to that of the PRO of interest, and their correlation should be sufficient (≥0.37).[35–37] For the QLQ-NMIBC24, the QLQ-C30 global QoL domain is a suitable candidate anchor. The suitability of anchor-based MCID estimation was evaluated by examining correlations between changes in QLQ-NMIBC24 domains and the anchor QLQ-C30, and anchor-based MCID estimates were only derived when this criterion was met.

Results

Patient Characteristics at Baseline

A total of 270 patients in ATLAS and 240 in ENVISION were enrolled to receive treatment, following screening of 396 and 318 patients, respectively. Baseline demographic characteristics for patients treated in the ATLAS and enrolled in the ENVISION studies are summarized in Table 1.

In ATLAS (n = 270), the mean age was 66.5 years (SD 10.5), and most participants were male (70.4%). Almost all patients were White (98.9%), with two (0.7%) identifying as Asian and one (0.4%) with race not reported (Table 1).

In ENVISION (n = 240), the mean age was 68.8 years (SD 11.6), and the majority were male (61.3%). Most participants were White (97.5%), with two Asian (0.8%), three Black or African American (1.3%), and one with race not reported (0.4%, Table 1).

At baseline, all patients in both studies completed at least one item of the patient-reported outcome questionnaires, and most had complete datasets (96.3% in ATLAS and 93.8% in ENVISION). At week 6, the proportion of fully completed questionnaires declined in ATLAS (46.3%), reflecting the absence of a Week 6 visit for patients in the TURBT-alone arm, but remained high in ENVISION (95.4%). At month 3, completion rates were 88.9% and 89.6% in ATLAS and ENVISION, respectively (Table 1).

Item Correlations and Consistency of QLQ-NMIBC24 Domains

Analysis of within-scale correlations for the QLQ-NMIBC24 questionnaire demonstrated good convergent validity across all multi-item domains in the ATLAS study cohort. Item-item correlations ranged from 0.40 to 0.96 across domains and timepoints, indicating moderate to strong associations between items within each domain (Table 2). Item-omitted correlations ranged from a minimum of 0.44 for Urinary Symptoms at baseline to a maximum of 0.96 for Sexual Function at Month 3, supporting good convergent validity (Table 2).

The Malaise domain exhibited lower internal consistency, with Cronbach's alpha values of 0.34 at baseline, 0.55 at Week 6, and 0.67 at Month 3, all below the commonly accepted threshold of 0.70, indicating limited reliability of this subscale in this population. Cronbach's alpha coefficients for the other multi-items domains ranged from 0.78 to 0.90 at baseline, demonstrating good to strong internal consistency (Table 2). Similar values were observed at Week 6 and Month 3, indicating that most domains maintained stable and acceptable internal consistency over time. The only exception was the Bloating and Flatulence domain, which showed a borderline Cronbach's alpha value of 0.69 at Month 3 (Table 2).

Table 1 Baseline Demographic Characteristics of Patients in the ATLAS and ENVISION Studies

Variable	All Patients – ATLAS[a] n = 270	All Patients – ENVISION[b] n = 240
Age, year		
Mean (SD)	66.5 (10.5)	68.8 (11.6)
Median (IQR)	68 (61–74)	70 (61.5–77.5)
Range	23 – 88	30 – 92
Sex, n (%)		
Female	80 (29.6)	93 (38.8)
Male	190 (70.4)	147 (61.3)
Race, n (%)		
Asian	2 (0.7)	2 (0.8)
Black or African American	0 (0)	3 (1.3)
White	267 (98.9)	234 (97.5)
Not reported	1 (0.4)	1 (0.4)
Baseline completion, n (%)		
All items completed	260 (96.3)	225 (93.8)
At least one item completed	270 (100.0)	240 (100)
Week 6 completion, n (%)		
All items completed	125 (46.3)	229 (95.4)
At least one item completed	129 (47.8)	233 (97.1)
Month 3 completion, n (%)		
All items completed	240 (88.9)	215 (89.6)
At least one item completed	249 (92.2)	226 (94.2)

Notes: [a]ATLAS included patients from: US, Europe, Israel. [b]ENVISION included patients from: US, Austria, Bulgaria, Estonia, Georgia, Latvia, Lithuania, Poland, Serbia, Spain.
Abbreviations: SD, standard deviation; IQR, inter quartile range; n, number.

Differentiation Between Patient Subgroups

To assess known-groups validity, we compared QLQ-C30 and QLQ-NMIBC24 scores between patient subgroups defined by physical function scores (PFS). Patients with lower physical function scores (PFS < 90) from the ENVISION dataset reported significantly worse level of functioning and more symptoms across most QLQ-C30 scales and QLQ-NMIBC24 domains, compared to patients with higher physical function scores (PFS ≥ 90), supporting known-groups validity ($p < 0.05$, Table 3). As expected, the largest differences were observed for Physical Functioning ($p < 0.0001$, effect size = −1.63), Fatigue ($p < 0.0001$, effect size = 0.97), and Global Quality of Life ($p < 0.0001$, effect size = −0.79, Table 3).

Most scales and items were similar between males and females, except that females reported significantly more problems than males in the QLQ-NMIBC24 Future Worries ($p = 0.006$, effect size = 0.39), Sexual Function ($p < 0.0001$, effect size = 1.01) and Sexual Enjoyment ($p = 0.0085$, effect size = 0.73) domains, and in the QLQ-C30 Emotional Function scale ($p = 0.0223$, effect size = - 0.31; Table 3).

Table 2 Convergent Validity and Internal Consistency of QLQ-NMIBC24 Domains

ATLAS	QLQ-NMIBC24 Domains with at least Two ITEMS					
	Urinary Symptoms	Malaise[c]	Future Worries	Bloating and Flatulence[c]	Sexual Function[c]	Male Sexual Problems[c]
Baseline						
Item-item correlation[a] (min-max)	0.40–0.83	0.48–0.48	0.71–0.83	0.82–0.82	0.86–0.86	0.78–0.78
Item-omitted correlation[b] (min-max)	0.44–0.79	0.48–0.48	0.75–0.83	0.82–0.82	0.86–0.86	0.78–0.78
Cronbach's alpha	0.86	0.34	0.90	0.78	0.86	0.78
Week 6						
Item-item correlation[a] (min-max)	0.59–0.92	0.73–0.73	0.82–0.92	0.88–0.88	0.90–0.90	0.91–0.91
Item-omitted correlation[b] (min-max)	0.64–0.85	0.73–0.73	0.84–0.91	0.88–0.88	0.90–0.90	0.91–0.91
Cronbach's alpha	0.91	0.55	0.93	0.85	0.88	0.89
Month 3						
Item-item correlation[a] (min-max)	0.53–0.86	0.82–0.82	0.73–0.91	0.78–0.78	0.96–0.96	0.89–0.89
Item-omitted correlation[b] (min-max)	0.53–0.77	0.82–0.82	0.79–0.87	0.78–0.78	0.96–0.96	0.89–0.89
Cronbach's alpha	0.89	0.67	0.91	0.69	0.93	0.86

Notes: [a]Polychoric correlation between items in each domain. [b]Polychoric correlation between each individual item and domain total raw score, omitting the item. [c]Domain with two items.

Stability of Patient-Reported Scores Over Time

Test-retest reliability, assessed on ENVISION data via intraclass correlation coefficient (ICC) between baseline and Week 6, and between baseline and Month 3, indicated good reliability (0.75–0.9)[33] for the Urinary Symptoms and Sexual Function at both Week 6 (ICC = 0.79 for both domains) and Month 3 (ICC = 0.80 and 0.82, respectively; Table 4). The Male Sexual Problem domain showed good reliability at Week 6 (ICC = 0.78) but moderate reliability at Month 3 (ICC = 0.63).

The Future Worries and Risk of Contaminating Partner domains demonstrated moderate reliability (0.5–0.75)[33] between baseline and both Week 6 (ICC = 0.70 and 0.61, respectively) and Month 3 (ICC = 0.58 and 0.70, respectively; Table 4).

The Malaise and Sexual Intimacy domains showed moderate reliability at Week 6 (ICC = 0.66 and 0.64, respectively) but low (ICC = 0.47) and poor reliability (ICC = −0.13) at Month 3, respectively (Table 4). In contrast, the Intravesical Treatment Issues, Bloating and Flatulence, and Sexual Enjoyment domains showed low reliability at Week 6 (ICC = 0.46, 0.45 and 0.44, respectively), but moderate reliability at Month 3 (ICC = 0.55, 0.61 and 0.71 respectively; Table 4).

Responsiveness to Change

Changes from baseline in QLQ-NMIBC24 domain scores were assessed using ANCOVA models adjusted for baseline values on the ENVISION data, comparing patients with and without complete response (CR) after 3 months of treatment with UGN-102. Adjusted coefficients and 95% confidence intervals at Week 6 and Month 3 are presented in Figure 1 and Supplementary Table 1.

The QLQ-NMIBC24 Urinary Symptoms (coefficient = 5.04, 95% CI: 0.61 to 9.47) and Sexual Intimacy (coefficient = −7.58, 95% CI: −15.06 to −0.10) domains demonstrated sensitivity to symptom changes at Week 6 (Figure 1, Supplementary Table 1), while the Malaise domain (coefficient = −3.81, 95% CI: −6.96 to −0.66) showed sensitivity to symptom changes at Month 3 (Figure 1, Supplementary Table 1).

Table 3 Mean QLQ-C30 and QLQ-NMIBC24 Scores by Physical Function Category and Sex

Scale/Item	PF < 90, n = 105, mean (SD)	PF ≥ 90, n = 135, mean (SD)	*p* value (*t*-test)	Effect Size[a]	Female, n = 93, mean (SD)	Male, n = 147, mean (SD)	*p* value (*t*-test)	Effect Size[a]
QLQ-C30: Functional scales[b]								
Physical	75.7 (13.8)	98.3 (2.9)	< 0.0001	−1.63	88.7 (13.7)	88.3 (15.2)	0.8248	0.03
Role	84.8 (20.9)	97.3 (9.2)	< 0.0001	−0.60	93.2 (14.6)	90.9 (17.8)	0.2847	0.15
Cognitive	86.7 (17.2)	93.3 (12.1)	0.0009	−0.39	89.6 (16.8)	90.9 (13.5)	0.5238	−0.08
Emotional	79.4 (17.8)	87.1 (14.2)	0.0003	−0.43	80.7 (15.6)	85.6 (16.5)	0.0223	−0.31
Social	86.8 (17.9)	94.6 (11.1)	0.0001	−0.43	91.4 (14.5)	91.0 (15.3)	0.8564	0.02
Global QoL	66.6 (18.0)	80.9 (15.3)	< 0.0001	−0.79	74.2 (16.5)	74.9 (18.9)	0.7647	−0.04
QLQ-C30: Symptom scales[c]								
Fatigue	25.7 (16.4)	9.8 (12.7)	< 0.0001	0.97	17.6 (17.6)	16.3 (15.7)	0.5586	0.07
Pain	14.8 (18.2)	5.2 (11.4)	< 0.0001	0.52	9.9 (17.4)	9.1 (14.2)	0.7155	0.05
Nausea/Vomiting	1.7 (6.5)	0.5 (2.8)	0.0678	0.19	1.4 (5.3)	0.8 (4.5)	0.3362	0.12
Constipation	14.0 (21.1)	5.9 (15.7)	0.0013	0.38	12.2 (20.7)	7.7 (17.0)	0.0831	0.22
Diarrhea	4.4 (12.3)	2.5 (10.5)	0.1896	0.16	3.2 (11.1)	3.4 (11.5)	0.9063	−0.02
Sleep	21.9 (25.2)	10.1 (17.9)	0.0001	0.47	17.2 (23.4)	14.1 (21.3)	0.2951	0.13
Dyspnea	13.7 (20.5)	3.2 (10.7)	< 0.0001	0.51	7.9 (15.8)	7.7 (17.0)	0.9354	0.01
Appetite	8.3 (17.2)	2.5 (9.7)	0.0023	0.34	4.7 (13.5)	5.2 (13.9)	0.7594	−0.04
Financial	9.8 (19.0)	5.7 (14.4)	0.0638	0.22	6.1 (16.3)	8.4 (16.9)	0.2954	−0.14
QLQ-NMIBC24 Domains[c]								
Urinary Symptoms[d]	19.2 (16.1)	9.7 (11.2)	< 0.0001	0.59	14.8 (16.2)	13.2 (13.1)	0.4389	0.10
Malaise[d]	6.3 (10.7)	2.1 (6.3)	0.0004	0.40	4.3 (10.1)	3.7 (7.8)	0.6245	0.06
Intravesical Treatment Issues[d]	17.1 (19.7)	10.4 (17.5)	0.0067	0.34	15.9 (19.4)	11.8 (18.2)	0.1017	0.21
Future Worries[d]	42.1 (22.6)	30.8 (20.5)	0.0001	0.50	40.6 (20.4)	32.7 (22.8)	0.0060	0.39
Bloating and Flatulence[d]	11.9 (17.3)	7.0 (13.0)	0.0155	0.29	11.4 (18.3)	7.7 (12.7)	0.0911	0.20
Male Sexual Problems[e]	29.4 (37.1)	15.3 (27.3)	0.0116	0.38	NA	NA	NA	NA
Sexual Intimacy[f]	14.3 (24.9)	7.6 (14.1)	0.2505	0.27	10.5 (15.9)	8.8 (17.9)	0.6911	0.11
Risk of Contaminating Partner[f]	25.4 (36.4)	14.1 (23.4)	0.1944	0.31	17.5 (23.2)	16.7 (28.5)	0.8909	0.04
Female Sexual Problems[g]	11.1 (19.2)	16.7 (27.2)	0.6934	−0.29	NA	NA	NA	NA
Sexual Function[d]	87.9 (18.8)	73.9 (27.2)	< 0.0001	0.75	90.2 (16.4)	73.7 (27.1)	< 0.0001	1.01
Sexual Enjoyment[f]	46.0 (26.8)	39.9 (22.0)	0.3497	0.23	54.4 (22.8)	37.7 (22.2)	0.0085	0.73

Notes: [a]Glass's delta, standard deviations of the two groups are different. [b]A higher score means better function. [c]A higher score means more symptoms or worse problems. [d]Data for 134 patients with PF ≥ 90 and 105 patients with PF < 90. Data for 92 female and 147 male patients. [e]Data for 83 patients with PF ≥ 90 and 64 patients with PF < 90. [f]Data for 66 patients with PF ≥ 90 and 21 patients with PF < 90. Data for 19 female and 68 male patients. [g]Data for 16 patients with PF ≥ 90 and 3 patients with PF < 90.
Abbreviations: NA, not available; PF, physical function; SD, standard deviation.

Meaningful Change

The possible Minimal Clinically Important Difference (MCID) ranges for each QLQ-NMIBC24 domain at Week 6 and Month 3 are shown in Figure 2. Improvements are depicted on the negative side and worsening on the positive side. Overall, the figure highlights the variability of MCID ranges across domains and over time, showing that some domains,

Table 4 Test-Retest Reliability of QLQ-NMIBC24 Domains by Intraclass Correlation Coefficients (ICC) Between Band Follow-Up Visits (Week 6 and Month 3)

QLQ-NMIBC24 Domains	Baseline – Week 6				Baseline – Month 3			
	N	ICC	LB	UB	N	ICC	LB	UB
Urinary Symptoms	98	0.79	0.70	0.85	83	0.80	0.71	0.87
Malaise	98	0.66	0.53	0.76	83	0.47	0.28	0.62
Intravesical Treatment Issues	98	0.46	0.28	0.60	83	0.55	0.38	0.69
Future Worries	98	0.70	0.58	0.79	83	0.58	0.41	0.70
Bloating and Flatulence	98	0.45	0.27	0.59	83	0.61	0.45	0.73
Male Sexual Problems	55	0.78	0.65	0.87	48	0.63	0.42	0.77
Sexual Intimacy	25	0.64	0.34	0.82	25	−0.13	−0.51	0.29
Risk of Contaminating Partner	25	0.61	0.30	0.81	26	0.70	0.44	0.85
Female Sexual Problems	5	-	-	-	6	0.62	−0.18	0.93
Sexual Function	98	0.79	0.70	0.86	83	0.82	0.72	0.88
Sexual Enjoyment	25	0.44	0.06	0.71	25	0.71	0.44	0.86

Abbreviations: N, number of patients; ICC, Intraclass correlation coefficient; LB, Lower 95% confidence interval (CI) bound; UB, Upper 95% CI bound.

such as Malaise and Urinary Symptoms, show narrower consistent ranges after 6 weeks and 3 months, while others, such as Sexual Intimacy, show variable ranges across time (Figure 2). The detailed results of distribution-based metrics, including one-half SD and SEM, performed on ENVISION data are shown in Table 5. The one-half SD at baseline ranged from 4.37 for Malaise to 16.29 for Male Sexual Problems, indicating variability in baseline scores across domains. The SEM was generally comparable to or slightly higher than one-half SD, ranging from 5.13 (Malaise, Baseline–Week 6) to 19.84 (Male Sexual Problems, Baseline–Month 3).

The correlations between the change from baseline (CFB) in QLQ-NMIBC24 domains and the CFB in the QLQ-C30 QoL domain did not reach the minimum correlation threshold, neither for CFB to Week 6 (correlations between −0.37 and −0.06) nor to Month 3 (correlations between −0.26 and 0.01). Therefore, no potential MCID could be derived from anchor-based approaches.

Discussion

This study provides the first comprehensive psychometric evaluation of the EORTC QLQ-NMIBC24 in patients with LG-IR-NMIBC within global clinical trial settings, demonstrating psychometric properties largely consistent with previous reports.[17–20] The module showed good convergent validity across all the multi-item domains, confirming that items designed to measure the same construct were appropriately correlated, in line with prior findings from the UK,[17] Netherlands,[18] Denmark,[19] and Korea.[20]

Internal consistency was generally good, with Cronbach's alpha values exceeding the commonly accepted threshold of 0.7[17–20] for most domains. The only exception was the Malaise domain, which demonstrated lower internal consistency. These findings are in line with previous reports, where reduced internal consistency in the Malaise domain at baseline and during follow-up visits has been observed across different cultural and clinical settings.[17,18,20] Consistent observations across studies suggests that the two items comprising the Malaise domain ("feeling unwell" and "feeling tired") may not form a robust unidimensional scale in this population. Accordingly, the utility of the Malaise domain as a standalone scale warrants further psychometric evaluation in future research.

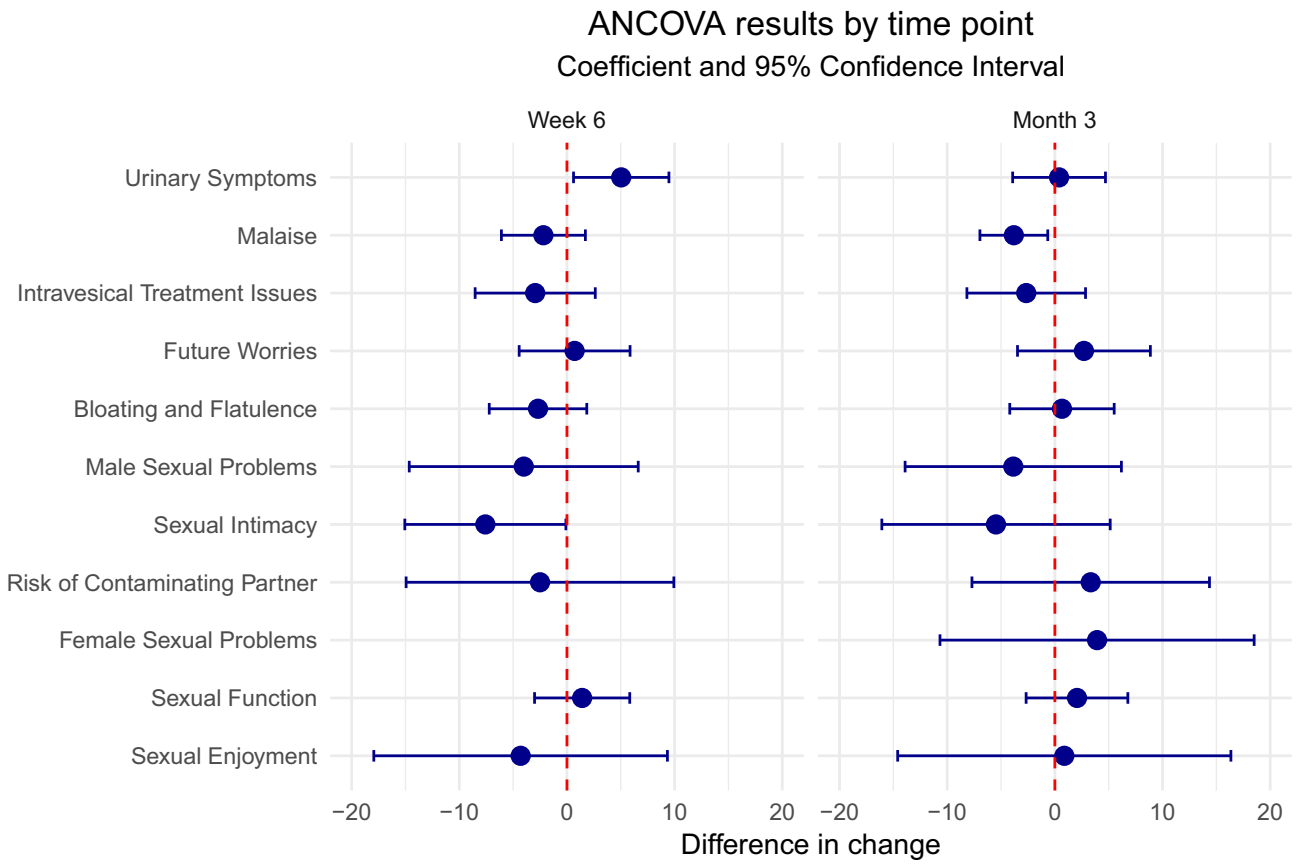

Figure 1 Differences in QLQ-NMIBC24 domain scores by clinical response at month 3, at Week 6 and Month 3. Estimated differences in change from baseline between patients with and without complete response (CR) at month 3 across QLQ-NMIBC24 domains. Results from Week 6 are shown on the left panel, results from Month 3 on the right panel. Differences were estimated using ANCOVA models adjusting for baseline domain scores. Points represent the estimated coefficients, and horizontal error bars indicate 95% confidence intervals. Confidence intervals excluding zero indicate statistically significant differences in changes from baseline associated with CR.

Known-groups comparisons showed that patients reporting lower physical functioning in the QLQ-C30 also had worse functional scores and more symptoms across QLQ-NMIBC24 domains. These results reinforce the sensitivity of the instrument to clinical differences, in line with previous validation studies.[17,20] While no sex-related differences were observed across general domains, female patients reported greater impairment in the Sexual Function and Sexual Enjoyment domains. Sex-specific variation in sexual QLQ-NMIBC24 domains has been previously observed by Blazeby et al[17] and Park et al,[20] with the trend in Blazeby et al consistent with our findings, whereas the trend in Park et al was opposite to that observed here.

Test-retest reliability, assessed via ICCs, was generally good, particularly for the Urinary Symptoms and Sexual Function domains. However, negative or low ICCs were observed in domains with very small sample sizes, such as Sexual Intimacy at Month 3, which may be due to limited data.[38] This finding highlights the importance of adequately powered samples, especially in domains where missing data can be common.

Responsiveness to change was observed in specific domains of the QLQ-NMIBC24 following ANCOVA analysis. In particular, the Urinary Symptoms, Sexual Intimacy, and Malaise domains demonstrated sensitivity to change, while other domains did not exhibit significant changes. This pattern of variable domain-level responsiveness has been reported in previous validation studies of the QLQ-NMIBC24, where some domains showed significant change over time while others remained stable.[17,20] These results suggest that the QLQ-NMIBC24 can detect meaningful changes in selected domains, although responsiveness may vary across different aspects of patients' health-related quality of life.

Anchor-based MCID estimation requires a minimum correlation (≥0.37) between the anchor and the PRO to ensure the anchor is sufficiently related to the outcome. In this study, correlations between the QLQ-C30 global QoL domain and

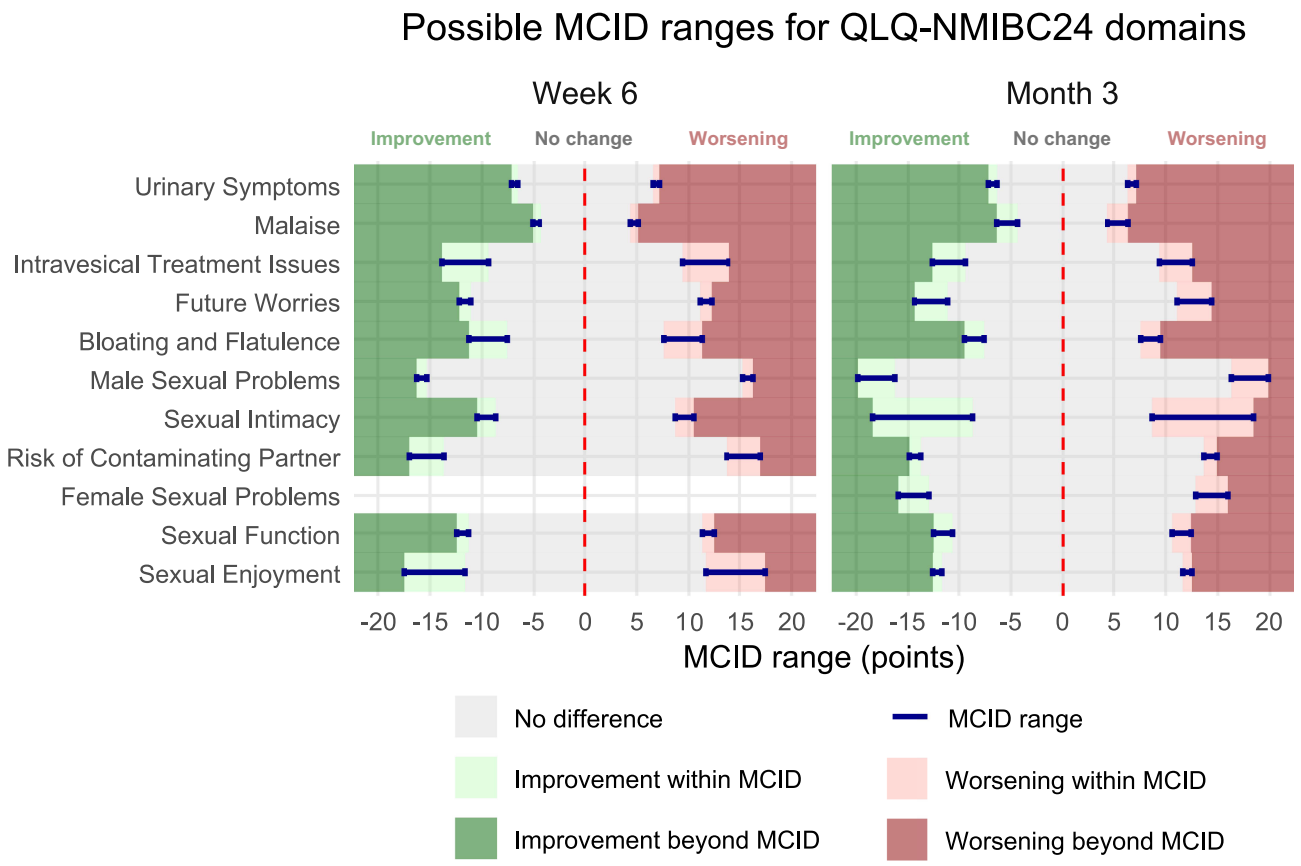

Figure 2 Possible Minimal Clinically Important Difference (MCID) ranges for QLQ-NMIBC24 domains at Week 6 and Month 3. Bars on the negative side represent improvements, and bars on the positive side represent worsening, with shading indicating whether changes are within or beyond the MCID (Table 5). The grey area indicates the change is not clinically relevant/no change. Domains with unavailable MCID estimates (Sexual Function at week 6) are indicated by missing bars.

QLQ-NMIBC24 domains did not meet this threshold. The reasons for these low correlations are unclear. One possibility is that, although the QLQ-C30 is a widely used cancer-specific HRQoL instrument, it may not fully capture disease-specific concerns in LG-IR-NMIBC. Consequently, distribution-based methods were used to estimate MCID. These estimates demonstrated variability across QLQ-NMIBC24 domains, with lower threshold values ranging from 4.37 for Malaise to 16.29 for Male Sexual Problems. While these thresholds are context-specific, a change equal to or exceeding the MCID can be considered meaningful in that domain, with the direction of the change (positive or negative) indicating improvement or worsening. The analysis of meaningful change suggests that not all domains of the QLQ-NMIBC24 are equally sensitive to detecting change. Multi-item domains tended to capture smaller and more nuanced shifts in health status, reflecting the finer granularity of their scoring. By contrast, single-item domains required larger changes before these could be interpreted as meaningful. Taken together, these findings could aid in the clinical interpretation of HRQoL in patients with LG-IR-NMIBC and enable clinicians to evaluate the impact of different treatment interventions on patients with LG-IR-NMIBC.

Recently, a new NMIBC-specific patient-reported outcome measure, the NMIBC Symptom Index (NMIBC-SI), was published.[39] Developed to assess patient-reported symptom burden and treatment experience,[39] its relationship to the QLQ-NMIBC24 is not yet established. Future research could explore whether the NMIBC-SI may have a role as a complementary instrument to existing measures, including its potential use as a source of anchor questions in NMIBC clinical trials.

Although this study provides valuable insights into the psychometric properties of the QLQ-NMIBC24 in the LG-IR-NMIBC population, it has some limitations. In the ATLAS study, participants completed only the QLQ-NMIBC24 and

Table 5 Estimated Difference Distribution-Based Approach, QLQ-NMIBC24 Domains

QLQ-NMIBC24 Domains [Number of Items in Domain]	One-half SD[a]	SEM		Possible MCID Range	
		Baseline Week 6	Baseline Month 3	Baseline Week 6	Baseline Month 3
Urinary Symptoms [7]	7.17	6.57	6.34	[6.57; 7.17]	[6.34; 7.17]
Malaise [2]	4.37	5.13	6.38	[4.37; 5.13]	[4.37; 6.38]
Intravesical Treatment Issues [1]	9.38	13.85	12.55	[9.38; 13.85]	[9.38; 12.55]
Future Worries [4]	11.08	12.20	14.40	[11.08; 12.20]	[11.08; 14.40]
Bloating and Flatulence [2]	7.59	11.27	9.48	[7.59; 11.27]	[7.59; 9.48]
Male Sexual Problems [2]	16.29	15.30	19.84	[15.30; 16.29]	[16.29; 19.84]
Sexual Intimacy [1]	8.69	10.48	18.43	[8.69; 10.48]	[8.69; 18.43]
Risk of Contaminating Partner [1]	13.66	16.99	14.92	[13.66; 16.99]	[13.66; 14.92]
Female Sexual Problems [1]	12.87	NA	15.97	NA	[12.87; 15.97]
Sexual Function [2]	12.43	11.29	10.61	[11.29; 12.43]	[10.61; 12.43]
Sexual Enjoyment [1]	11.64	17.47	12.52	[11.64; 17.47]	[11.64; 12.52]

Notes: [a] SD calculated at baseline.
Abbreviations: NA, not available; MCID, minimal clinically important difference; SD, standard deviation; SEM, standard error of the measurement.

not the QLQ-C30, which limited the availability of patient-reported data. Score changes reflecting meaningful differences were estimated using distribution-based approaches, as the correlations between the QLQ-C30 QoL domain and the QLQ-NMIBC24 domains were insufficient to support anchor-based methods, which are typically given greater weight when available.

Patient Global Impression of Change (PGIC) data, typically used to calculate anchor-based MCID estimates,[40,41] were not collected.

This study used phase 3 clinical trial data to establish responder thresholds, which may limit the generalizability of the findings.[37] In addition, the study population was relatively homogeneous, consisting predominantly of white and male participants, which may further restrict applicability to the broader LG-IR-NMIBC population. Thresholds derived from this context could be influenced by trial-specific efficacy outcomes, and future research in the broader LG-IR-NMIBC population will be necessary to validate and refine these estimates.

Sample size limitations, particularly in certain subgroups and domains, may have contributed to variability in reliability estimates and to the occurrence of low ICCs. Missing responses were most frequent for sex-related items, consistent with prior validation studies from the UK,[17] the Netherlands,[18] and Korea.[20] As a consequence, missing responses may have affected the results for sex-related domains. It should be noted that some questions were only administered to participants who were sexually active during the previous four weeks. Despite this, overall completion rates were high, supporting the feasibility and acceptability of the QLQ-NMIBC24 in this trial context. However, findings may not be generalizable to other patient groups (ie, other than LG-IR) within NMIBC, as MCIDs are specific to the data used to estimate them.[34]

Finally, no formal multiplicity adjustment was applied.

Conclusion

This study demonstrates the reliability and validity of the EORTC QLQ-NMIBC24 in patients with LG-IR-NMIBC, supporting its use as a valuable tool for assessing patient-reported outcomes in this population, both in clinical research and potentially in routine practice. The distribution-based MCID estimates generated in this analysis may further assist

clinicians in interpreting HRQoL scores and evaluating the impact of treatment interventions. Moreover, these MCID estimates provide foundational thresholds for contextualizing clinical trial outcomes within health technology assessments, value frameworks, and economic models. Some limitations, including the generalizability of findings, should be considered when interpreting these results. Also, the Malaise domain showed suboptimal performance, suggesting that scoring modifications or careful interpretation is advised when using this domain in LG-IR-NMIBC populations. Future studies could build on these findings by applying anchor-based methods, although identifying suitable anchors for disease-specific instruments in lower-risk populations may be challenging. Furthermore, evaluating responsiveness in broader NMIBC populations, including multi-center, cross-cultural, or international confirmatory factor analyses would improve generalizability.

Abbreviations

ANCOVA, Analysis of covariance; CFB, Change from baseline; CI, Confidence interval; CR, Complete response; EORTC, European Organization for Research and Treatment of Cancer; HG, High grade; HRQoL, Health-related quality of life; ICC, Intraclass correlation coefficients, IQR, Inter quartile range; IR, Intermediate risk; LB, Lower bound; LG, Low grade; MDC, Minimal detectable change; MCID, Minimal clinically important difference; NA, Not available; NMIBC, Non-muscle-invasive bladder cancer; PFS, Physical function scale; PGIC, Patient Global Impression of Change; PRO, Patient reported outcome; QLQ-C, Quality of Life Questionnaire Core 30; SD, Standard deviation; SE, Standard error; SEM, Standard error of measurement; TURBT, Transurethral resection of the bladder tumor; UB, Upper bound.

Ethics Approval and Informed Consent

The clinical trial from which data were derived was approved by the appropriate research ethics committee(s), and all participants provided written informed consent for use of their data for research purposes. The current psychometric validation study constitutes a secondary analysis of anonymized data and posed no additional risk to participants; therefore, no further ethical approval was required.

Acknowledgments

Under the direction of the authors, writing support was provided by Francesco Amadeo, María José Aragón and Idaira Rodríguez Santana of Prime HCD. An abstract related to this work was presented at the American Urology Association Annual Meeting 2025 as a poster presentation with interim findings. The poster's abstract was published in the Journal of Urology: https://doi.org/10.1097/01.JU.0001109788.99555.df.07.

Funding

This research was supported by UroGen Pharma Ltd., with sponsor employees serving as co-authors involved in the study design, analysis, interpretation, and manuscript drafting.

Disclosure

CCP has served as a consultant for UroGen Pharma and Janssen and served on an advisory board for Ferring Pharma. RD was a paid consultant of UroGen Pharma. TB is an employee of the University of Chester, Health and Social Care department and was employed by Prime HCD at the time of the study; Prime HCD received research funding from UroGen Pharma for this work. VT, NU, AN, BB, VL and MJL are employees of UroGen Pharma and own stock. AMS received research funding through her institution in the past 36 months from for-profit organizations (UroGen Pharma Ltd. and Pfizer). She also received a one-time consulting fee ($2k) to present findings at UroGen Pharma Ltd. in June 2024. The authors report no other conflicts of interest in this work.

References

1. Nielsen ME, Smith AB, Meyer AM, et al. Trends in stage-specific incidence rates for urothelial carcinoma of the bladder in the United States: 1988 to 2006. *Cancer*. 2014;120(1):86–95. doi:10.1002/cncr.28397

2. Holzbeierlein JM, Bixler BR, Buckley DI, et al. Diagnosis and treatment of non-muscle invasive bladder cancer: AUA/SUO guideline: 2024 Amendment. *J Urol*. 2024;211(4):533–538. doi:10.1097/JU.0000000000003846
3. Magers MJ, Lopez-Beltran A, Montironi R, Williamson SR, Kaimakliotis HZ, Cheng L. Staging of bladder cancer. *Histopathology*. 2019;74 (1):112–134. doi:10.1111/his.13734
4. Chang SS, Boorjian SA, Chou R, et al. Diagnosis and treatment of non-muscle invasive bladder cancer: AUA/SUO guideline. *J Urol*. 2020;196 (4):1021–1029. doi:10.1016/j.juro.2016.06.049
5. Tan WS, Steinberg G, Witjes JA, et al. Intermediate-risk non–muscle-invasive bladder cancer: updated consensus definition and management recommendations from the International Bladder Cancer Group. *Eur Urol Oncol*. 2022;5(5):505–516. doi:10.1016/j.euo.2022.05.005
6. Ben Muvhar R, Paluch R, Mekayten M. Recent advances and emerging innovations in transurethral resection of bladder tumor (TURBT) for non-muscle invasive bladder cancer: a comprehensive review of current literature. *Res Rep Urol*. 2025;17:69–85. doi:10.2147/RRU.S386026
7. Bree KK, Shan Y, Hensley PJ, et al. Management, surveillance patterns, and costs associated with low-grade papillary Stage Ta non-muscle-invasive bladder cancer among older adults, 2004-2013. *JAMA Network Open*. 2022;5(3):e223050. doi:10.1001/jamanetworkopen.2022.3050
8. Kim LHC, Patel MI. Transurethral resection of bladder tumour (TURBT). *Transl Androl Urol*. 2020;9(6):3056–3072. doi:10.21037/tau.2019.09.38
9. Filon M, Schmidt B. New treatment options for non-muscle-invasive bladder cancer. *Am Soc Clin Oncol Educ Book*. 2025;45(2):e471942. doi:10.1200/EDBK-25-471942
10. Matulewicz RS, Steinberg GD. Non-muscle-invasive bladder cancer: overview and contemporary treatment landscape of neoadjuvant chemoablative therapies. *Rev Urol*. 2020;22(2):43–51.
11. Matulay JT, Soloway M, Witjes JA, et al. Risk-adapted management of low-grade bladder tumours: recommendations from the International Bladder Cancer Group (IBCG). *BJU Int*. 2020;125(4):497–505. doi:10.1111/bju.14995
12. Food and Drug Administration. FDA approves mitomycin intravesical solution for recurrent low-grade intermediate-risk non-muscle invasive bladder cancer. Available from: https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-mitomycin-intravesical-solution-recurrent-low-grade-intermediate-risk-non-muscle. Accessed August 21, 2025.
13. Perry MB, Taylor S, Khatoon B, et al. Examining the effectiveness of electronic patient-reported outcomes in people with cancer: systematic review and meta-analysis. *J Med Internet Res*. 2024;26:e49089. doi:10.2196/49089
14. Balitsky AK, Rayner D, Britto J, et al. Patient-reported outcome measures in cancer care: an updated systematic review and meta-analysis. *JAMA Network Open*. 2024;7(8):e2424793. doi:10.1001/jamanetworkopen.2024.24793
15. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365–376. doi:10.1093/jnci/85.5.365
16. EORTC. Non-muscle-invasive bladder cancer EORTC QLQ - NMIBC24 questionnaire. Available from: https://qol.eortc.org/questionnaire/qlq-nmibc24/. Accessed February 23, 2026.
17. Blazeby JM, Hall E, Aaronson NK, et al. Validation and reliability testing of the EORTC QLQ-NMIBC24 questionnaire module to assess patient-reported outcomes in non–muscle-invasive bladder cancer. *Europ urol*. 2014;66(6):1148–1156. doi:10.1016/j.eururo.2014.02.034
18. Ripping TM, Westhoff E, Aaronson NK, et al. Validation and reliability of the Dutch version of the EORTC QLQ-NMIBC24 Questionnaire Module for patients with non-muscle-invasive bladder cancer. *J Patient-Reported Outcomes*. 2021;5(1):96. doi:10.1186/s41687-021-00372-4
19. Mogensen K, Christensen KB, Vrang M-L, Hermann GG. Hospitalization for transurethral bladder resection reduces quality of life in Danish patients with non-muscle-invasive bladder tumour. *Scand J Urol*. 2016;50(3):170–174. doi:10.3109/21681805.2015.1132762
20. Park J, Shin DW, Kim T-H, et al. Development and validation of the Korean Version of the European Organization for research and treatment of cancer quality of life questionnaire for patients with non-muscle invasive bladder cancer: EORTC QLQ-NMIBC24. *Cancer Res Treat*. 2018;50 (1):40–49. doi:10.4143/crt.2016.594
21. Longoni M, Scilipoti P, De Angelis M, et al. Contemporary outcomes in non-muscle-invasive bladder cancer: a large European multicentre study. *BJU Int*. 2025;136(5):826–834. doi:10.1111/bju.16780
22. McNall S, Hooper K, Sullivan T, Rieger-Christ K, Clements M. Treatment modalities for non-muscle invasive bladder cancer: an updated review. *Cancers*. 2024;16(10):1843. doi:10.3390/cancers16101843
23. Osoba D, Aaronson N, Zee B, Sprangers M, te Velde A. Modification of the EORTC QLQ-C30 (version 2.0) based on content validity and reliability testing in large samples of patients with cancer. The Study Group on Quality of Life of the EORTC and the Symptom Control and Quality of Life Committees of the NCI of Canada Clinical Trials Group. *Qual Life Res*. 1997;6(2):103–108. doi:10.1023/a:1026429831234
24. Cocks K, Wells JR, Johnson C, et al. Content validity of the EORTC quality of life questionnaire QLQ-C30 for use in cancer. *Eur J Cancer*. 2023;178:128–138. doi:10.1016/j.ejca.2022.10.026
25. Fayers P, Aaronson N, Bjordal K, et al. *The EORTC QLQ-C30 Scoring Manual*. 3rd ed. Brussels: European Organisation for Research and Treatment of Cancer; 2001.
26. UroGen Pharma Ltd. A phase 3 study of UGN-102 for low grade intermediate risk non-muscle invasive bladder cancer (ATLAS). Available from: https://www.clinicaltrials.gov/study/NCT04688931. Accessed February 23, 2026.
27. UroGen Pharma Ltd. A phase 3 single-arm study of UGN-102 for treatment of low grade intermediate risk non-muscle-invasive bladder cancer (ENVISION). Available from: https://www.clinicaltrials.gov/study/NCT05243550. Accessed February 23, 2026.
28. Prasad SM, Huang WC, Shore ND, et al. Treatment of low-grade intermediate-risk nonmuscle-invasive bladder cancer With UGN-102 ± transurethral resection of bladder tumor compared to transurethral resection of bladder tumor monotherapy: a randomized, controlled, phase 3 trial (ATLAS). *J Urol*. 2023;210(4):619–629. doi:10.1097/ju.0000000000003645
29. Prasad SM, Shishkov D, Mihaylov NV, et al. Primary chemoablation of recurrent low-grade intermediate-risk nonmuscle-invasive bladder cancer with UGN-102: a single-arm, open-label, phase 3 trial (ENVISION). *J Urol*. 2025;213(2):205–216. doi:10.1097/ju.0000000000004296
30. StataCorp. *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC.; 2023.
31. Association AP. APA dictionary of psychology. Updated April 19, 2018. Available from: https://dictionary.apa.org/cronbachs-alpha. Accessed August 13, 2025.
32. Cohen J. CHAPTER 2 - The t Test for means. In: Cohen J, editor. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press; 1977:19–74.
33. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Med*. 2016;15 (2):155–163. doi:10.1016/j.jcm.2016.02.012

34. 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(2):102–109. doi:10.1016/j.jclinepi.2007.03.012
35. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD*. 2005;2(1):63–67. doi:10.1081/COPD-200050663
36. Cohen J. CHAPTER 3 - The significance of a product moment rs. In: Cohen J, editor. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press; 1977:75–107.
37. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res*. 2018;27(1):33–40. doi:10.1007/s11136-017-1616-3
38. Williams B, FitzGibbon L, Brady D, Christakou A. Sample size matters when estimating test-retest reliability of behaviour. *Behav Res Methods*. 2025;57(4):123. doi:10.3758/s13428-025-02599-1
39. Rutherford C, Tait MA, Costa DSJ, et al. Development and psychometric evaluation of a patient-reported symptom index for patients with non-muscle invasive bladder cancer: the NMIBC-SI. *J Patient Rep Outcomes*. 2025;9(1):36. doi:10.1186/s41687-025-00864-7
40. Directorate-General for Health and Food Safety. Guidance on outcomes for joint clinical assessments; 2024.
41. Uysal S, Sadjadi R. The minimal clinically important difference and its use in neuromuscular disorder. *Practical Neurolog*. 2025;24(3):21–24.