Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Patient-Reported Quality-of-Life Trajectories in NMIBC Patients Undergoing Repeat Transurethral Resection: A 12-Month Multicenter Cohort Study

İlker Çelen

drilkercelen@yahoo.com

Manisa Celal Bayar University

Serkan Akan

Fatih Sultan Mehmet Training and Research Hospital

Murat Akgül

University of Health Sciences Turkey, Ümraniye Training and Research Hospital

Serdar Çelik

University of Health Sciences, Izmir City Hospital

Hasan Hüseyin Tavukçu

Istanbul Medipol University

Research Article

Keywords: non-muscle-invasive bladder cancer; repeat transurethral resection; health-related quality of life; patient-reported outcome measures

Posted Date: September 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9483696/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Patient-Reported Quality-of-Life Trajectories in NMIBC Patients Undergoing Repeat Transurethral Resection: A 12-Month Multicenter Cohort Study

Abstract

Purpose: To characterize 12-month patient-reported quality-of-life (QoL) trajectories in NMIBC patients undergoing repeat transurethral resection (re-TURBT), and to describe patterns of early burden and subsequent recovery.

Methods: Retrospective multicenter cohort (n = 90 NMIBC patients, 5 centres, 5 time points over 12 months). SF-36, HADS, and EORTC QLQ-NMIBC24 scored per official manuals. BH-adjusted Wilcoxon comparisons (per-subscale, family of four); Cohen's d_z with bootstrap 95% CI (5,000 iterations); GEE sensitivity. A conservative no-change imputation scenario served as missing-data sensitivity.

Results: A triphasic pattern was observed in most subscales: transient T1 early burden, T2 – T3 recovery, T4 improvement. Physical Functioning showed a significant transient T1 decline (65.8 vs. T0: 72.4; p_adj < 0.001) with recovery to baseline by T4 (d_z =− 0.01; 95% CI: − 0.21 to +0.20). Role-Physical improved by +30.4 points (d_z = +0.67; 95% CI: +0.43 to +0.96); 58.1% exceeded the 10-point improvement threshold (n = 50/86). HADS anxiety and depression declined below baseline (d_z =− 0.37 and − 0.39). Future Worries showed the largest improvement among the six pre-defined subscales (d_z =− 0.84; − 1.13 to − 0.60), significant from T2 onwards. Conservative no-change sensitivity confirmed directional robustness for all primary findings.

Conclusion: Trajectories were consistent with transient early burden followed by recovery across most domains at 12 months. The proportion of patients with probable anxiety fell from 35.6% right after re-TURBT to 6.7% at 12 months. Observational design without a concurrent control group limits causal inference. Early psychosocial support is hypothesis-generating and warrants prospective evaluation.

Keywords: non-muscle-invasive bladder cancer; repeat transurethral resection; health-related quality of life; patient-reported outcome measures

Plain English Summary

Bladder cancer patients sometimes need a second surgery (re-TURBT) to remove remaining cancer cells. We followed 90 patients from five Turkish hospitals to assess how quality of life changed over 12 months. Physical functioning showed a temporary worsening right after re-TURBT (at the first post-operative assessment), then recovered to baseline by 12 months with no lasting change. Most other areas of quality of life improved by 12 months. More than half

of patients (58.1%) exceeded the pre-defined 10-point improvement threshold for physical role functioning. The proportion of patients with probable anxiety fell from 35.6% right after re-TURBT to 6.7% at 12 months. Future Worries showed the largest improvement among the six pre-defined subscales. These findings come from an observational study without a comparison group; they cannot confirm that re-TURBT caused these changes. Providing psychological support during the first few months may help patients, but this requires testing in a controlled trial.

Introduction

Bladder cancer ranks among the most common malignancies in men worldwide, with approximately 614,000 new cases and 220,000 deaths reported annually.[1] NMIBC accounts for roughly 75% of all newly diagnosed bladder tumors and encompasses stages Ta, T1, and carcinoma in situ (CIS).[2,3]

Following initial transurethral resection (TUR), the risk of residual tumor ranges from 33% to 55%, particularly in T1 lesions.[4,5] The EAU guidelines recommend re-TURBT within 2 – 6 weeks for patients with high-risk NMIBC, incomplete initial resection, or absence of muscularis propria.[3,6,7,8] Re-TURBT also carries procedural burdens, including repeated anaesthesia and postoperative discomfort. Re-TURBT has been shown to enhance recurrence-free survival and enable more accurate risk stratification.[5,9]

HRQOL assessment has emerged as an increasingly valued outcome in oncological care.[10] In bladder cancer, QoL is affected by the disease, surgical interventions, intravesical therapy, and surveillance burden.[11] Data addressing patient-reported outcomes specific to NMIBC across the peri-re-TURBT pathway remain limited.[12,13]

The SF-36 is a widely validated generic QoL instrument.[14] The EORTC QLQ-NMIBC24 is a validated disease-specific module for NMIBC.[15] The HADS is a validated screening tool for anxiety and depressive symptoms in medical populations.[16,17]

The target estimand is the observed 12-month patient-reported trajectory across the NMIBC treatment pathway. The primary objective was to characterize whether early patient-reported burden is transient or sustained. The secondary objective was to examine stage-related differences at 12 months. Because baseline assessments preceded the index TUR, the design captures the peri-re-TURBT treatment continuum rather than isolating a single causal re-TURBT effect.

Materials and Methods

Study Design and Patient Selection

Retrospective multicenter longitudinal cohort, Turkish Urooncology Society Geriatric Urooncology and Quality of Life Working Group. PRO questionnaires were administered prospectively as part of routine clinical care (2016 – 2023); records were retrospectively extracted, pooled, and analysed after ethics approval (2025). No prospective protocol or pre-registration was established; SF-36 RP was designated as the primary endpoint prior to analysis.

Clinical eligibility criteria: (1) histologically confirmed NMIBC following initial TUR; (2) re-TURBT indication per EAU guidelines[3]; (3) age ≥ 18. Exclusion: (1) muscle-invasive disease; (2) distant metastasis; (3) documented active psychiatric disorder; (4) concurrent secondary malignancy.

Of 127 screened records, 112 met clinical eligibility. Of these, 90 had complete PRO data at all five time points and constituted the primary complete-case analytic cohort (Fig. 1). The 22 not included (10 unable to contact, 7 incomplete questionnaires, 5 withdrew consent) did not differ from completers in age ($p = 0.43$), sex ($p = 0.67$), or stage ($p = 0.51$); baseline PRO data were unavailable for excluded patients. Reported in accordance with STROBE.[18]

Among the 52 Ta patients, re-TURBT was indicated based on incomplete resection, high-grade suspicion, or absence of detrusor muscle in the initial specimen, consistent with EAU criteria; tumour grade and detrusor data were not uniformly available across centres and are reported as a limitation.

Ethics Approval and Consent

Performed in accordance with the Declaration of Helsinki. PRO questionnaires were collected as part of routine clinical care; retrospective research use was subsequently approved by Manisa Celal Bayar University Faculty of Medicine Non-Interventional Clinical Research Ethics Committee (December 11, 2025 / No. 1498) with waiver of individual re-consent. Patients had provided verbal informed consent for QoL assessment at routine follow-up visits.

Surgical Procedure

Standard TUR; re-TURBT performed 2 – 6 weeks after the index procedure (T1 assessments obtained at the first post-operative visit, approximately 2 – 4 weeks post re-TURBT). Both conventional and bipolar resection techniques were employed; complications classified per Clavien–Dindo.[19] Intravesical therapy type, schedule, and regimen heterogeneity were not systematically captured; their contribution to T1 – T2 trajectory changes cannot be formally quantified.

Assessment Instruments

SF-36 (36-item): scored per the official RAND Health algorithm.[14] Items were recoded using the RAND scoring table (item 7 non-linear recoding; item 8 = 6 − item 8 when item 7 > 1; special case: item 7 = 1 and item 8 = 1 → item 8 = 6). Role-Physical (RP) and Role-Emotional (RE) response value 3 was treated as missing: three patients had all RP items coded 3 at T0 (RP T0 n = 87) and one stage-T1 patient was missing all T4 SF-36 data; these four were excluded from the paired RP responder analysis (n = 86). Note: value 3 did not appear in RP items at T4; the T4 full-sample n is 89. For RP, Table 3 reports paired-subset means (n = 86); S2 reports full-sample floor/ceiling effects (T0 n = 87, T4 n = 89). Sensitivity analysis treating value 3 as 100 (no limitation): $d_z = 0.66$ vs 0.67; virtually identical. Validated Turkish version (KF-36) used.[20]

EORTC QLQ-NMIBC24: scored per official EORTC scoring manual.[15] The module comprises 6 multi-item subscales and 5 single items. Five multi-item subscales and single-item Q54 are reported in the main text: Urinary Symptoms (Q31 – Q37); Malaise (Q38 – Q39); Future Worries (Q41 – Q44); Sexual Functioning (Q47 – Q48, functional); Male Sexual Problems (Q49 – Q50); Female Sexual Problems (Q54, n = 10). The sixth subscale (Bloating and Flatulence, Q45 – Q46) and remaining single items (Q40, Q51, Q52, Q53) are reported in Supplementary Table S1c. These latter areas were not included in the pre-defined comparison families; results are exploratory (unadjusted). Symptom scales: higher = worse; Sexual Functioning: higher = better. TS = (RS − 1)/3 × 100 (symptom) or (4 − RS)/3 × 100 (functional).

HADS: 14-item screening tool (anxiety: odd items 1,3,5,7,9,11,13; depression: even items 2,4,6,8,10,12,14); each subscale 0 – 21.[16] Pre-computed subscale totals from the dataset were used. HADS is a screening instrument; threshold interpretations (probable case: ≥11) are indicative. Validated Turkish version.[17]

Data Collection

Assessments at: T0 (pre-index TUR baseline; T0 collected at initial clinical evaluation before the index TUR; only patients with subsequently confirmed NMIBC enrolled); T1 (post-re-TURBT, pre-intravesical therapy; approximately 2 – 4 weeks post re-TURBT); T2, T3, T4 (3, 6, 12 months). All follow-up assessments (T2, T3, and T4) were administered immediately before the scheduled surveillance cystoscopy visit. Face-to-face interview by trained staff.

Statistical Analysis

SF-36 RP was designated as the primary endpoint prior to analysis (power analysis: MID = 10 points[21]; within-subject SD ≈ 20; ICC = 0.5; Monte Carlo, 10,000 iterations; $\alpha = 0.05$; power = 0.80; required n = 82; enrolled n = 90). All other outcomes were designated secondary

or exploratory. Clinical significance for RP was assessed using the 10-point threshold[21] and the 0.5 SD criterion (SD = 43.3, so 0.5 SD ≈ 21.7 points); these criteria are not equivalent and are reported separately.

Longitudinal changes: Friedman tests; BH-adjusted Wilcoxon signed-rank comparisons (per-subscale family of four: T1, T2, T3, T4 vs. T0). To evaluate the triphasic recovery pattern, consecutive timepoint comparisons (T1 → T2, T2 → T3, T3 → T4) were conducted for the primary endpoint and selected secondary outcomes (unadjusted, exploratory). Effect sizes: Cohen's d_z (mean of paired differences / SD, ddof = 1); positive values indicate improvement for higher-is-better scales (SF-36), negative for higher-is-worse scales (HADS, BLS24 symptom). Bootstrap 95% CI: percentile method, 5,000 iterations. Responder analysis: >10-point change (MID threshold). For the primary endpoint (RP), consecutive comparisons showed: T0 → T1: $p = 0.318$ (non-significant early change); T1 → T2: $p = 0.034$ (recovery); T2 → T3: $p = 0.209$; T3 → T4: $p < 0.001$ (consolidation).

Secondary stage comparisons (Ta vs. T1) at 12 months used Mann – Whitney U tests with BH correction (family = 8 for SF-36; 6 for BLS24). Note: one T1 stage patient had missing T4 SF-36 data (n = 33 for stage comparisons); BLS24 comparisons also used n = 33 where applicable.

A conservative no-change imputation scenario assigned all 22 excluded patients T4 = T0 completer mean (synthetic zero-change pairs); Wilcoxon and d_z recomputed on the augmented sample. Sample sizes differ by outcome: for RP (paired n = 86), the augmented sample is n = 108 (86 + 22); for HADS and Future Worries (full n = 90), n = 112 (90 + 22). This represents a no-change (BOCF-equivalent) scenario, not a worst-case analysis.

GEE models (exchangeable working correlation; selected as a parsimonious robust choice given variable follow-up intervals across centres; robust SE) were fitted for the primary endpoint (SF-36 RP) and two key psychological outcomes (HADS-A, HADS-D), which were selected a priori before analysis. Three specifications: (1) time only; (2) time + stage + time × stage interaction; (3) additionally adjusting for study centre. GEE was applied to the complete-case sample (n = 90) and addresses within-subject correlation; it does not address the 22 excluded patients. Analyses: IBM SPSS v22.0; R v4.3.0 (geepack for GEE; boot for bootstrap CI).

Results

Patient Characteristics

Final cohort: 90 patients (79 male, 11 female; mean age 67.8 ± 10.0 years). Stages: Ta 52 (57.8%), T1 34 (37.8%), CIS 4 (4.4%) (Table 1). Among Ta patients, re-TURBT was indicated for incomplete resection, high-grade suspicion, or absent detrusor.

Table 1. Patient characteristics (N = 90)

Characteristic	Value	Range / %
Age (years), mean ± SD	67.8 ± 10.0	39 – 87
Sex — Male / Female, n	79 / 11	87.8% / 12.2%
Stage Ta / T1 / CIS, n	52 / 34 / 4	57.8% / 37.8% / 4.4%
Tumour size (mm), mean ± SD	26.2 ± 15.7	—
Systemic comorbidity, n (%)	57 present / 33 absent	63.3% / 36.7%
Index TUR anaesthesia Spinal / General, n	40 / 50	44.4% / 55.6%
Complications ≥ Clavien 1, n (%)	2	2.2%
Prior surgery, n (%)	30	33.3%

Grade, EAU risk group, multifocality, and detrusor status were not uniformly available across centres. Recurrence/progression data and intravesical therapy details were not captured in the retrospective dataset.

Floor and Ceiling Effects

Role-Physical (RP) showed pronounced floor effects at T0: 44.8% scored 0 (maximum limitation) and 25.3% scored 100. Role-Emotional (RE) showed similar bimodality (T0: 40.0% floor, 26.7% ceiling). By 12 months, floor effects markedly decreased (RP: 6.7%; RE: 3.4%) while ceiling effects increased (RP: 53.9%; RE: 29.2%), consistent with normalization. Full floor/ceiling distributions are in Supplementary Table S2. The contribution of regression to the mean to T4 gains in RP and RE cannot be quantified without a concurrent comparator.

Longitudinal Changes in SF-36 Quality of Life Scores

Friedman tests indicated significant temporal variation across all eight SF-36 subscales (Table 2). A triphasic trajectory was observed in most subscales; the pattern was not uniform (Table 3, Fig. 2).

Table 2. Friedman test results for SF-36 subscales

Subscale (abbrev.)	Full name	χ^2	df	p-value	Kendall's W
PF	Physical Functioning	30.6	4	<0.001	0.086
RP	Role-Physical	81.1	4	<0.001	0.228
BP	Bodily Pain	93.8	4	<0.001	0.263
GH	General Health	56.3	4	<0.001	0.158
VT	Vitality	54.2	4	<0.001	0.152
SF	Social Functioning	43.8	4	<0.001	0.123
RE	Role-Emotional	69.5	4	<0.001	0.195
MH	Mental Health	51.6	4	<0.001	0.145

Table 3. SF-36 subscale scores across five time points

Subscale	T0 mean ± SD	T1 mean ± SD p_adj/d_z	T2 mean ± SD p_adj/d_z	T3 mean ± SD p_adj/d_z	T4 mean ± SD p_adj/d_z	d_z T0 → T4 [95% CI]
PF	72.4 ± 21.2	65.8 ± 24.3 <0.001* / −0.36	73.7 ± 15.0 0.956 / +0.05	71.1 ± 16.1 0.323 / −0.10	72.7 ± 16.3 0.976 / −0.01	−0.01 [−0.21, +0.20]
RP	41.4 ± 43.3	37.1 ± 40.1 0.424 / −0.10	42.2 ± 39.3 0.926 / +0.01	48.3 ± 42.3 0.424 / +0.13	71.8 ± 34.0 <0.001* / +0.67	+0.67 [+0.43, +0.96]
BP	72.0 ± 31.1	70.4 ± 29.8 0.895 / −0.11	72.9 ± 27.2 0.895 / +0.06	79.5 ± 24.6 0.001* / +0.36	85.2 ± 20.8 <0.001* / +0.61	+0.61 [+0.46, +0.76]
GH	55.2 ± 20.9	51.1 ± 22.4 0.162 / −0.20	53.0 ± 18.9 0.181 / −0.13	57.4 ± 18.6 0.207 / +0.15	62.3 ± 21.3 0.012* / +0.38	+0.38 [+0.19, +0.57]
VT	59.0 ± 24.7	55.8 ± 25.8 0.020* / −0.23	53.3 ± 25.5 0.001* / −0.40	57.0 ± 23.4 0.085 / −0.17	64.3 ± 24.9 0.028* / +0.29	+0.29 [+0.10, +0.49]
SF	65.6 ± 24.3	57.8 ± 32.3 0.026* / −0.29	65.6 ± 23.1 0.711 / 0.00	71.2 ± 21.9 0.026* / +0.23	77.5 ± 26.5 <0.001* / +0.52	+0.52 [+0.31, +0.76]
RE	43.0 ± 41.6	28.9 ± 33.6 0.009* / −0.38	31.1 ± 31.7 0.059 / −0.27	49.8 ± 34.1 0.171 / +0.11	64.0 ± 28.5 <0.001* / +0.40	+0.40 [+0.19, +0.66]
MH	67.1 ± 21.3	64.8 ± 23.4 0.386 / −0.15	66.1 ± 20.4 0.504 / −0.09	69.6 ± 23.9 0.386 / +0.15	75.3 ± 20.6 <0.001* / +0.50	+0.50 [+0.30, +0.71]

*BH FDR per subscale (family=4). *p_adj<0.05. Higher=better. d_z = mean(Ti − T0) / SD(Ti − T0); positive d_z = improvement (increase). 95% CI: 5,000 bootstrap iterations. PF: significant transient T1 decline (p_adj<0.001) with recovery to baseline; 95% CI includes zero, providing no evidence of sustained T0-to-T4 difference. VT: paradoxical T2 worsening (p_adj = 0.001) that resolved by T4.*

Triphasic and Non-Triphasic Subscale Patterns

Subscales with a statistically significant T1 decline (p_adj < 0.05): PF (p_adj < 0.001; transient, full recovery by T4), Vitality (p_adj = 0.020), Social Functioning (p_adj = 0.026), and Role-Emotional (p_adj = 0.009). Subscales showing T4 improvement without a statistically significant T1 decline: RP (p_adj = 0.424), BP (p_adj = 0.895), GH (p_adj = 0.162), and MH (p_adj = 0.386). The "triphasic" label is a descriptive pattern; time-point-level significance is in Table 3. Notable exceptions:

Physical Functioning: Showed a significant transient T1 decline (65.8 vs. T0: 72.4; p_adj < 0.001) with complete recovery to baseline by T2, and no evidence of a sustained T0-to-T4 difference (d_z =− 0.01; 95% CI: − 0.21 to +0.20). Consecutive comparisons confirmed T1-to-T2 recovery (p < 0.001).

Vitality: Showed a T2 worsening (53.3 vs. T0: 59.0; p_adj = 0.001) that resolved by T4 (64.3, d_z = +0.29), consistent with intravesical therapy-period fatigue, though the absence of therapy-level data prevents formal attribution.

Role-Physical and Role-Emotional: Bimodal baseline distributions (floor effects 44.8% and 40.0%) mean a portion of the T4 gains may reflect regression to the mean; the contribution cannot be quantified without a concurrent comparator.

Responder Analysis (SF-36 RP)

At 12 months, 58.1% of patients (n = 50 of 86 evaluable) exceeded the 10-point improvement threshold, 34.9% (n = 30) remained within ±10 points, and 7.0% (n = 6) showed >10-point worsening. The denominator for this analysis is n = 86 (not 90): three patients had all RP items

coded as 3 at baseline and one patient (stage T1) had missing T4 SF-36 data, precluding paired difference computation for these four patients. These proportions reflect the same baseline distributional constraints, including pronounced RP floor effects; regression to the mean may contribute to observed gains.

Quality of Life Scores by Pathological Stage (Exploratory)

As a pre-defined exploratory secondary analysis, SF-36 scores at 12 months were compared by stage (Ta n = 52 vs. T1 n = 33; one T1 patient had missing T4 SF-36 data; CIS n = 4 excluded). No subscale reached statistical significance after BH correction (all p_adj > 0.06; Table 4). The secondary hypothesis that pathological stage would be associated with 12-month QoL outcomes was not supported. All analyses are unadjusted for confounders.

Table 4. SF-36 scores at 12 months by pathological stage (exploratory)

SF-36	Ta (n=52) mean ± SD	T1 (n=33[a]) mean ± SD	p-value[b]	p_adj[b]
PF	70.1 ± 15.4	77.1 ± 16.6	0.033	0.088
RP	66.3 ± 35.0	84.1 ± 28.5	0.016	0.064
BP	80.8 ± 22.1	91.3 ± 18.0	0.010	0.064
GH	60.5 ± 22.0	64.8 ± 19.7	0.265	0.265
VT	59.1 ± 27.1	70.5 ± 19.7	0.065	0.126
SF	73.6 ± 28.8	84.1 ± 21.3	0.106	0.130
RE	59.6 ± 29.8	70.7 ± 27.3	0.079	0.126
MH	71.8 ± 22.0	79.3 ± 17.8	0.114	0.130

[a] One T1 patient had missing T4 SF-36 data; n=33 for stage comparisons. [b] Mann – Whitney U test; BH FDR correction (family=8). No subscale reached significance. Exploratory analysis; unadjusted for confounders. BLS24 stage analysis: Supplementary Table S1b (no significant differences).

Anxiety and Depression Outcomes (HADS)

Both HADS subscales showed the triphasic pattern (Table 5, Fig. 5). The proportion meeting the probable case criterion (HADS ≥ 11) increased from 20.0% at T0 to 35.6% at T1, then fell to 6.7% at T4 for anxiety; and from 12.2% to 32.2% at T1, then 10.1% at T4 for depression. Effect sizes at 12 months were modest (d_z ≈ 0.37 – 0.39), consistent with gradual psychological adaptation; no published HADS MID exists.

Table 5. HADS scores across five time points

Subscale	T0 mean ± SD	T1 mean ± SD / p_adj / d_z	T2 mean ± SD / p_adj / d_z	T3 mean ± SD / p_adj / d_z	T4 mean ± SD / p_adj / d_z	d_z T0 → T4 [95% CI]
HADS-A	6.2 ± 4.5	7.7 ± 5.1 / 0.006* / +0.36	7.1 ± 3.8 / 0.017* / +0.27	5.5 ± 3.6 / 0.133 / −0.15	4.7 ± 4.2 / 0.002* / −0.37	−0.37 [−0.64, −0.16]
HADS-D	5.7 ± 3.7	7.6 ± 4.8 / <0.001* / +0.48	7.2 ± 4.2 / <0.001* / +0.47	5.6 ± 4.0 / 0.818 / −0.02	4.5 ± 4.0 / 0.001* / −0.39	−0.39 [−0.63, −0.18]

*BH FDR per subscale (family=4). *p_adj<0.05. Lower=better. d_z = mean(Ti − T0) / SD; negative d_z = improvement (decrease). 95% CI: 5,000 bootstrap iterations.*

EORTC QLQ-NMIBC24 (BLS24) Disease-Specific Outcomes

Table 6 summarises BLS24 trajectories per official EORTC scoring. Future Worries showed the largest improvement among the six reported subscales ($d_z = -0.84$; 95% CI: −1.13 to −0.60), significant from T2 onwards. Urinary Symptoms, Malaise, and Male Sexual Problems improved significantly at T4. Sexual Functioning showed no significant change. Female Sexual Problems (Q54; n = 10) is presented descriptively only. Q51 (Sexual Intimacy) and Q52 (Risk of Contaminating Partner) are in Supplementary Table S1c, where exploratory (unadjusted) results for all remaining module areas are also reported.

Table 6. EORTC QLQ-NMIBC24 subscale scores across five time points

Subscale (items; n[a])	T0 mean ± SD	T1 / p_adj / d_z	T2 / p_adj / d_z	T3 / p_adj / d_z	T4 mean ± SD / p_adj	d_z T0 → T4 [95% CI]
Urinary Symptoms (Q31 – Q37; n=90)	42.1 ± 23.4	41.9 ± 26.0 / 0.864 / −0.01	39.3 ± 24.1 / 0.408 / −0.14	38.8 ± 27.0 / 0.186 / −0.17	29.5 ± 19.5 / <0.001*	−0.77 [−1.00, −0.57]
Malaise (Q38 – Q39; n=90)	24.1 ± 27.3	25.6 ± 24.9 / 1.000 / +0.06	26.2 ± 24.3 / 1.000 / +0.10	21.5 ± 25.8 / 0.346 / −0.11	14.2 ± 20.7 / <0.001*	−0.61 [−0.80, −0.43]
Future Worries (Q41 – Q44; n=90)	39.5 ± 27.6	41.0 ± 28.1 / 0.278 / +0.08	34.7 ± 23.5 / 0.011* / −0.24	31.0 ± 26.5 / <0.001* / −0.45	24.3 ± 22.9 / <0.001*	−0.84 [−1.13, −0.60]
Sexual Functioning[b] (Q47 – Q48; n=87)	80.7 ± 18.7	85.1 ± 19.4 / 0.134 / +0.18	83.7 ± 18.4 / 0.134 / +0.21	82.9 ± 23.7 / 0.215 / +0.08	77.7 ± 25.0 / 0.465	−0.10 [−0.30, +0.11]
Male Sexual Problems (Q49 – Q50; n=68)	35.8 ± 34.6	41.4 ± 32.1 / 0.069 / +0.26	37.6 ± 32.5 / 0.267 / +0.06	38.8 ± 29.8 / 0.365 / +0.09	23.6 ± 27.5 / 0.016*	−0.40 [−0.57, −0.21]
Female Sexual Problems[c] (Q54; n=10)	6.7 ± 14.1	13.3 ± 28.1	20.0 ± 17.2	20.0 ± 17.2	13.3 ± 17.2 / descriptive	descriptive only

*BH FDR per subscale (family=4). *p_adj<0.05. Symptom scales: higher=worse; [b] Sexual Functioning: higher=better (functional). [a] n varies by valid data. [c] n=10 female patients; underpowered; descriptive only. Q51 (Sexual Intimacy) and Q52 (Risk of Contaminating Partner) are single items presented in Supplementary Table S1c. 95% CI: 5,000 bootstrap iterations. d_z: negative = symptom improvement.*

BLS24 Stage-Based Exploratory Analysis

No significant stage-related differences were found in BLS24 subscales at 12 months (all $p_adj > 0.37$ after BH correction; Supplementary Table S1b). The secondary hypothesis was not supported.

GEE Sensitivity Analysis

GEE confirmed significant time effects for SF-36 RP ($p<0.001$), HADS-A ($p<0.001$), and HADS-D ($p<0.001$) across all three model specifications. The time-by-stage interaction was not statistically significant for RP ($p = 0.63$; no statistical evidence of a differential RP trajectory by stage). Fully adjusted T4-vs-T0 contrasts (Model 3): SF-36 RP +29.3 points

(95% CI: 20.1 – 38.5); HADS-A −1.52 (−2.21 to −0.83); HADS-D −1.14 (−1.79 to −0.49). GEE was applied to the complete-case sample and does not address the 22 excluded patients.

Conservative No-Change Sensitivity Analysis

Under the conservative no-change scenario (22 excluded patients assigned T4 = T0 completer mean; for RP: T0 anchor = 41.9), the SF-36 RP T4 mean reduced from 71.8 to 65.7 (augmented n = 108; $d_z = 0.58$, $p < 0.001$). HADS-A: $d_z = -0.33$ (n = 112, $p < 0.001$); Future Worries: $d_z = -0.71$ (n = 112, $p < 0.001$). All primary directional findings were preserved.

Discussion

This multicenter cohort characterises 12-month patient-reported QoL trajectories across the NMIBC treatment pathway including re-TURBT. In most subscales, trajectories were consistent with a transient early burden followed by progressive recovery. The pattern was not universal: PF showed transient T1 decline with baseline recovery and no sustained T4 change; VT worsened at T2 before recovering; Future Worries improved progressively from T2. Concordant patterns across three complementary instruments support cross-instrument consistency.[12,13,22,23,24]

SF-36 RP improved by +30.4 points ($d_z = +0.67$; 95% CI: +0.43 to +0.96) at 12 months. This exceeds the 10-point threshold[21] and 0.5 SD (~21.7 points); responder analysis showed 58.1% improved >10 points. The extent to which regression to the mean contributed to RP and RE gains (which showed pronounced floor effects at baseline) cannot be quantified in the absence of a concurrent comparator.

PF showed a statistically significant T1 decline (65.8 vs. T0: 72.4; $p_{adj} < 0.001$) with complete recovery to baseline by T2 and no T4 difference ($d_z = -0.01$; 95% CI: −0.21 to +0.20). This is consistent with a transient post-procedural impact on physical functioning without sustained impairment, and strengthens the triphasic narrative for PF.

VT worsened significantly at T2 (53.3 vs. T0: 59.0; $p_{adj} = 0.001$), a finding that is compatible with the fatigue burden of concurrent intravesical therapy during this period. One possible explanation is intravesical therapy-related side effects; the present design cannot test this mechanism as therapy data were not captured.

HADS threshold analysis revealed a clinically informative trajectory: the proportion with probable anxiety (≥11) increased from 20.0% at T0 to 35.6% at T1, then fell to 6.7% at T4. Effect sizes were modest ($d_z \approx 0.37 – 0.39$), consistent with gradual adaptation; no published HADS MID exists for this context.

Future Worries (EORTC QLQ-NMIBC24 items Q41 – Q44) showed the largest improvement among the six reported subscales ($d_z = -0.84$; 95% CI: −1.13 to −0.60), significant from T2 onwards. Notably, the Intravesical Treatment Issues item (Q40) showed a T0 → T1 increase (40.0 → 47.4; Supplementary Table S1c), providing a patient-reported signal of intravesical treatment burden during that period, consistent with the T1 – T2 Vitality worsening. This partially corroborates the "intravesical therapy data not captured" limitation. Bloating and Flatulence (Q45 – Q46) showed a large improvement by T4 ($d_z = -0.71$; S1c), close in magnitude to Future Worries; the latter remains the largest across the six pre-defined subscales.

Stage-based comparisons showed no significant differences in any SF-36 or BLS24 subscale after BH correction, and the secondary hypothesis was not supported.

The observed trajectories provide no evidence of persistent QoL deterioration over 12 months among patients who underwent clinically indicated re-TURBT. The T1 – T2 window may represent a transient burden period amenable to targeted support. This is a hypothesis-generating observation that cannot be verified without a controlled trial. Furthermore, the absence of recurrence data and intravesical therapy records means the trajectories may reflect a mix of treatment effects and natural adaptation.

The Spearman correlation matrix (Fig. 4) confirmed the expected SF-36 multidimensional structure at T0. The strongest correlation was between General Health and Vitality ($\rho = 0.82$).

Limitations

The absence of a concurrent control group precludes causal inference. Recurrence and progression data were not available; 12-month recurrence rates of 30 – 50% are reported in NMIBC, and patients experiencing recurrence or additional procedures may have had systematically different trajectories, representing a potential survivorship selection bias that could inflate observed T4 improvements. Intravesical therapy type, schedule, and instillation count were not captured; BCG and intravesical chemotherapy have well-known side effects that may directly explain T1 – T2 changes. Grade, EAU risk group, multifocality, and detrusor data were not uniformly available; re-TURBT indication for Ta patients could therefore not be fully documented. RP and RE showed pronounced baseline floor effects (44.8% and 40.0% at score = 0); regression to the mean may contribute to observed T4 gains. Baseline assessments preceded the index TUR, meaning diagnosis shock may have depressed T0 values. Informative dropout cannot be excluded; the conservative sensitivity analysis shows directional robustness but baseline PRO data were unavailable for the 22 excluded patients. Face-to-face administration may introduce social desirability bias. Extended enrolment

(2016 – 2023, spanning COVID-19) introduces temporal confounding. The CIS subgroup (n = 4) and female subsample (n = 11) preclude meaningful subgroup analyses. Prospective controlled designs with pre-registration, oncological outcome capture, intravesical therapy documentation, and formal missing-data modelling are needed.

Conclusion

This multicenter longitudinal cohort documents patient-reported QoL trajectories consistent with transient early burden followed by recovery across most assessed domains at 12 months. Subscales with significant T1 decline (PF, VT, SF, RE) all recovered to or above baseline by T4. Future Worries showed the largest improvement among the six pre-defined subscales ($d_z = -0.84$), significant from T2 onwards. Responder analysis confirmed the ≥10-point improvement threshold was exceeded by 58.1% of patients (n = 50/86) for Role-Physical. Conservative sensitivity confirmed directional robustness. When re-TURBT is medically indicated, available observational data provide no signal of persistent HRQOL impairment over 12 months, although the observational design without a comparator limits causal inference. Phase-targeted psychosocial support during the T1 – T2 window is a hypothesis-generating recommendation that prospective trials should evaluate.

Figure Captions

Fig. 1. STROBE flow diagram. Records screened (n=127); met clinical eligibility (n=112); enrolled in PRO follow-up (n=112); not included in complete-case analysis (n=22): unable to contact (n=10), incomplete questionnaires (n=7), withdrew consent (n=5); completed all 5 time points and final analysis (n=90).

Fig. 2. SF-36 subscale trajectories (T0 – T4) with 95% CI. *$p_{adj} < 0.05$ vs. T0 (BH correction, per-subscale family of four). Higher scores = better health. Scored per official RAND Health algorithm.

Fig. 3. SF-36 subscale scores at 12 months by pathological stage (exploratory). Mean ± 95% CI. Mann – Whitney U test with BH FDR correction (family=8). No subscale reached significance after correction (all $p_{adj} > 0.06$). Ta n=52; T1 n=33 (one T1 patient had missing T4 data). CIS (n=4) excluded. Analysis is exploratory and unadjusted.

Fig. 4. Spearman correlation matrix of SF-36 subscales at baseline (T0). Values represent Spearman's ρ. **Significant after Bonferroni correction (28 comparisons, α=0.0018). Colour scale: 0 – 1 (all pairwise correlations are positive).

Fig. 5. HADS subscale trajectories with 95% CI. Green dashed line: normal range boundary (≤7); red dashed line: probable case threshold (≥11). *$p_{adj} < 0.05$ vs. T0 (BH correction). HADS is a screening instrument; threshold crossings are indicative only.

Fig. 6. EORTC QLQ-NMIBC24 trajectories for four primary subscales (Urinary Symptoms, Malaise, Future Worries, Sexual Functioning) with 95% CI. Male and Female Sexual Problems are in Table 6 and Supplementary Table S1b. Symptom scales: lower = better. Sexual Functioning (functional): higher = better. Scored per official EORTC manual. Significance markers (asterisks) are not shown; for time-point-level significance, see Table 6.

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74: 229–263.
2. Antoni S, Ferlay J, Soerjomataram I, et al. Bladder cancer incidence and mortality: A global overview and recent trends. *Eur Urol* 2017; 71: 96–108.
3. European Association of Urology. Guidelines on non-muscle-invasive bladder cancer (TaT1 and CIS) — 2025 update. https://uroweb.org/guidelines/non-muscle-invasive-bladder-cancer (2025, accessed 14 September 2026).
4. Naselli A, Hurle R, Paparella S, et al. Role of restaging transurethral resection for T1 non-muscle invasive bladder cancer: A systematic review and meta-analysis. *Eur Urol Focus* 2018; 4: 558–567.
5. Cumberbatch MGK, Foerster B, Catto JWF, et al. Repeat transurethral resection in non-muscle-invasive bladder cancer: A systematic review. *Eur Urol* 2018; 73: 925–933.
6. Chang SS, Boorjian SA, Chou R, et al. Diagnosis and treatment of non-muscle invasive bladder cancer: AUA/SUO guideline. *J Urol* 2016; 196: 1021–1029.
7. Gontero P, Sylvester R, Pisano F, et al. The impact of re-transurethral resection on clinical outcomes in a large multicentre cohort with T1 high-grade/Grade 3 bladder cancer treated with bacille Calmette-Guérin. *BJU Int* 2016; 118: 44–52.
8. Yanagisawa T, Mori K, Katayama S, et al. Repeat transurethral resection for non-muscle-invasive bladder cancer: An updated systematic review and meta-analysis. *Eur Urol Focus* 2024; 10: 93–105.
9. Divrik RT, Şahin AF, Yildirim Ü, et al. Impact of routine second transurethral resection on the long-term outcome of patients with newly diagnosed pT1 urothelial carcinoma: A prospective randomized clinical trial. *Eur Urol* 2010; 58: 185–190.
10. Fayers PM, Machin D. *Quality of Life: The Assessment, Analysis and Reporting of Patient-Reported Outcomes*. 3rd ed. Chichester: Wiley, 2016.
11. Catto JWF, Downing A, Mason S, et al. Quality of life after bladder cancer: A cross-sectional survey of patient-reported outcomes. *Eur Urol* 2021; 79: 621–632.
12. Beeren I, Voskuilen CS, van Rhijn BWG, et al. Patient-reported outcomes after transurethral resection of non-muscle-invasive bladder cancer: Results from a prospective cohort study. *World J Urol* 2024; 42: 123.
13. Rafiei S, Yazdani A, Ghasemi SR, et al. Quality of life in bladder cancer patients: A systematic review and meta-analysis. *BMJ Support Palliat Care* 2023; 13: e1089–e1099.

14. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–483.

15. 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. *Eur Urol* 2014; 66: 1148–1156.

16. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; 67: 361–370.

17. Aydemir Ö, Güvenir T, Küey L, et al. Hastane Anksiyete ve Depresyon Ölçeği Türkçe formunun geçerlilik ve güvenilirlik çalışması. *Türk Psikiyatri Derg* 1997; 8: 280–287.

18. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* 2007; 370: 1453–1457.

19. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: A new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; 240: 205–213.

20. Koçyiğit H, Aydemir Ö, Fişek G, et al. Kısa Form-36 (KF-36)'nın Türkçe versiyonunun güvenilirliği ve geçerliliği. *İlaç ve Tedavi Derg* 1999; 12: 102–106.

21. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; 41: 582–592.

22. Singer S, Ziegler C, Schwalenberg T, et al. Quality of life in patients with muscle invasive and non-muscle invasive bladder cancer. *Support Care Cancer* 2013; 21: 1383–1393.

23. Böhle A, Jocham D, Bock PR. Intravesical bacillus Calmette-Guérin versus mitomycin C for superficial bladder cancer: A formal meta-analysis. *J Urol* 2003; 169: 90–95.

24. Patel HD, Srivastava A, Patel ND, et al. A systematic review and meta-analysis of quality of life outcomes in patients with bladder cancer undergoing radical cystectomy. *BJU Int* 2021; 128: 512–524.

Figures

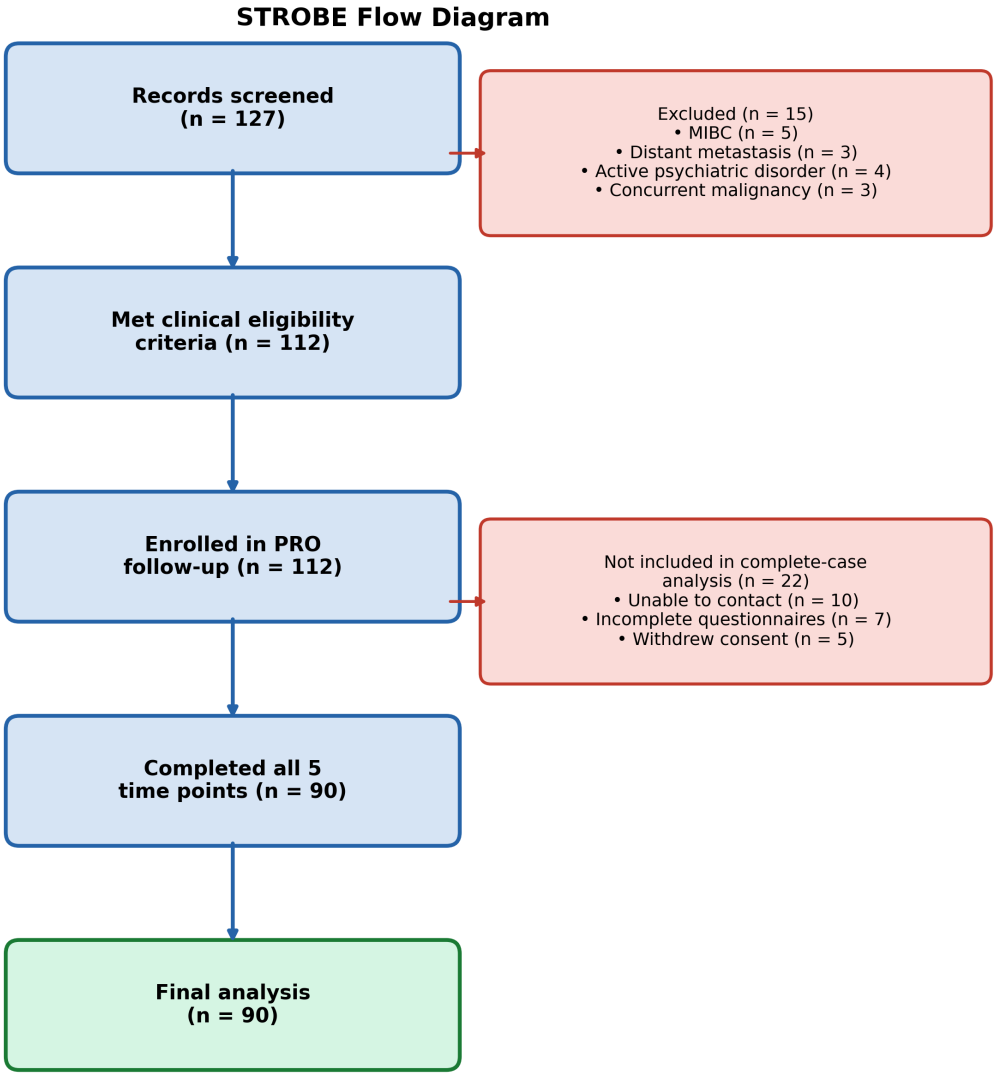


Figure 1

STROBE flow diagram. Records screened (n=127); met clinical eligibility (n=112); enrolled in PRO follow-up (n=112); not included in complete-case analysis (n=22): unable to contact (n=10), incomplete questionnaires (n=7), withdrew consent (n=5); completed all 5 time points and final analysis (n=90).

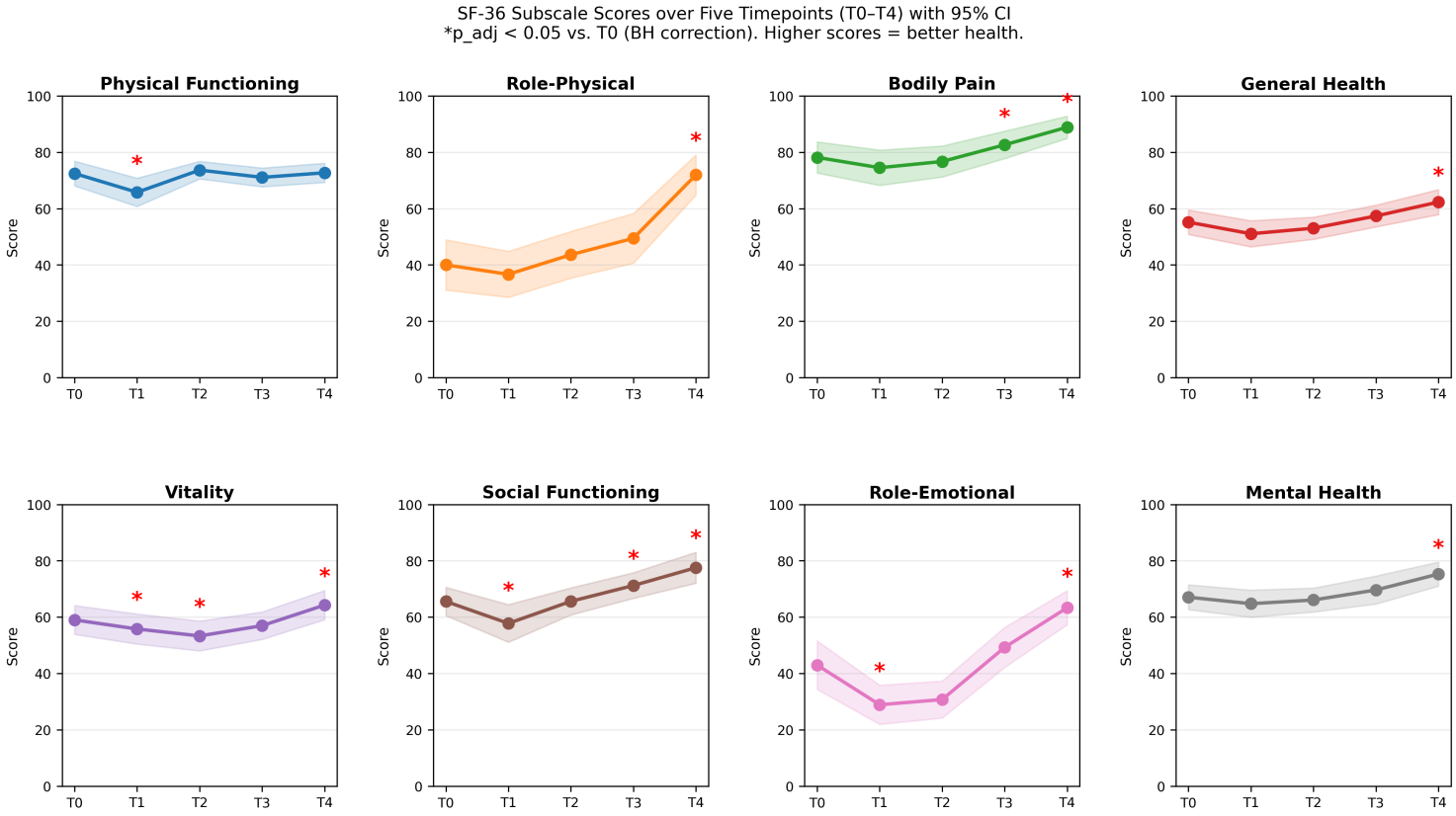


Figure 2

SF-36 subscale trajectories (T0 – T4) with 95% CI. *p_adj < 0.05 vs. T0 (BH correction, per-subscale family of four). Higher scores = better health. Scored per official RAND Health algorithm.

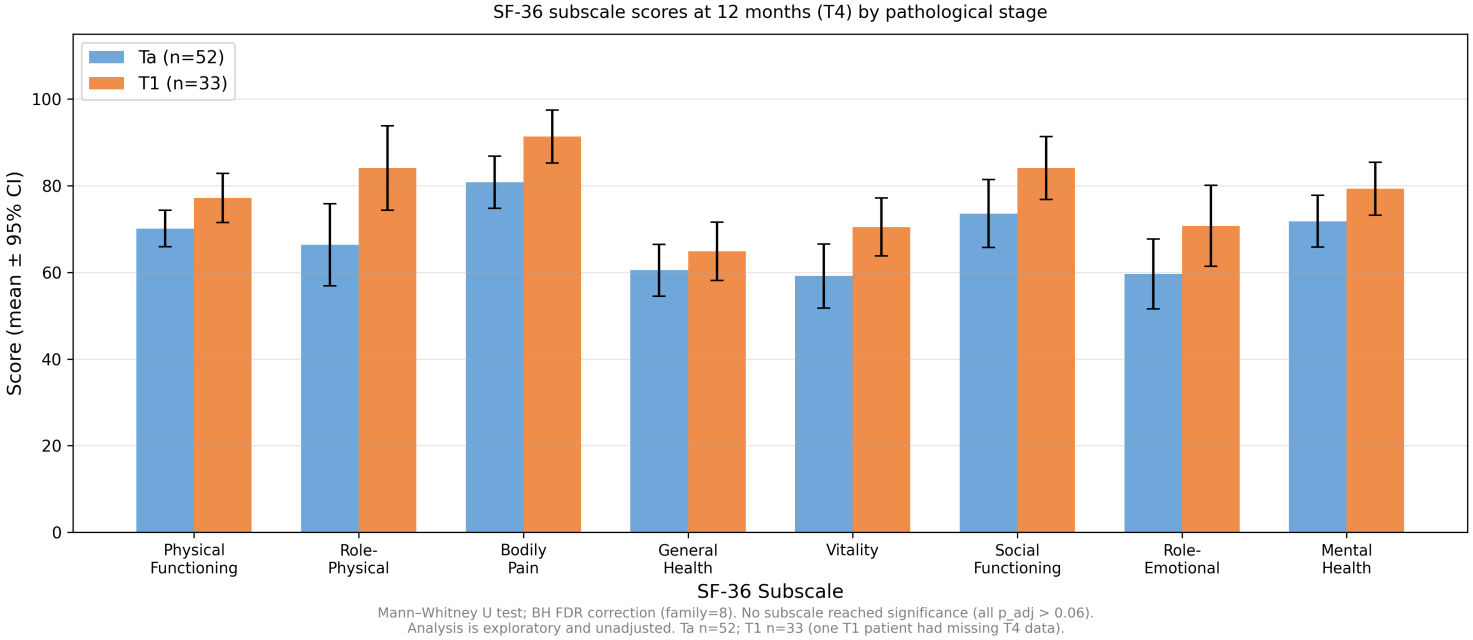


Figure 3

SF-36 subscale scores at 12 months by pathological stage (exploratory). Mean ± 95% CI. Mann – Whitney U test with BH FDR correction (family=8). No subscale reached significance after correction (all p_adj > 0.06). Ta n=52; T1 n=33 (one T1 patient had missing T4 data). CIS (n=4) excluded. Analysis is exploratory and unadjusted.

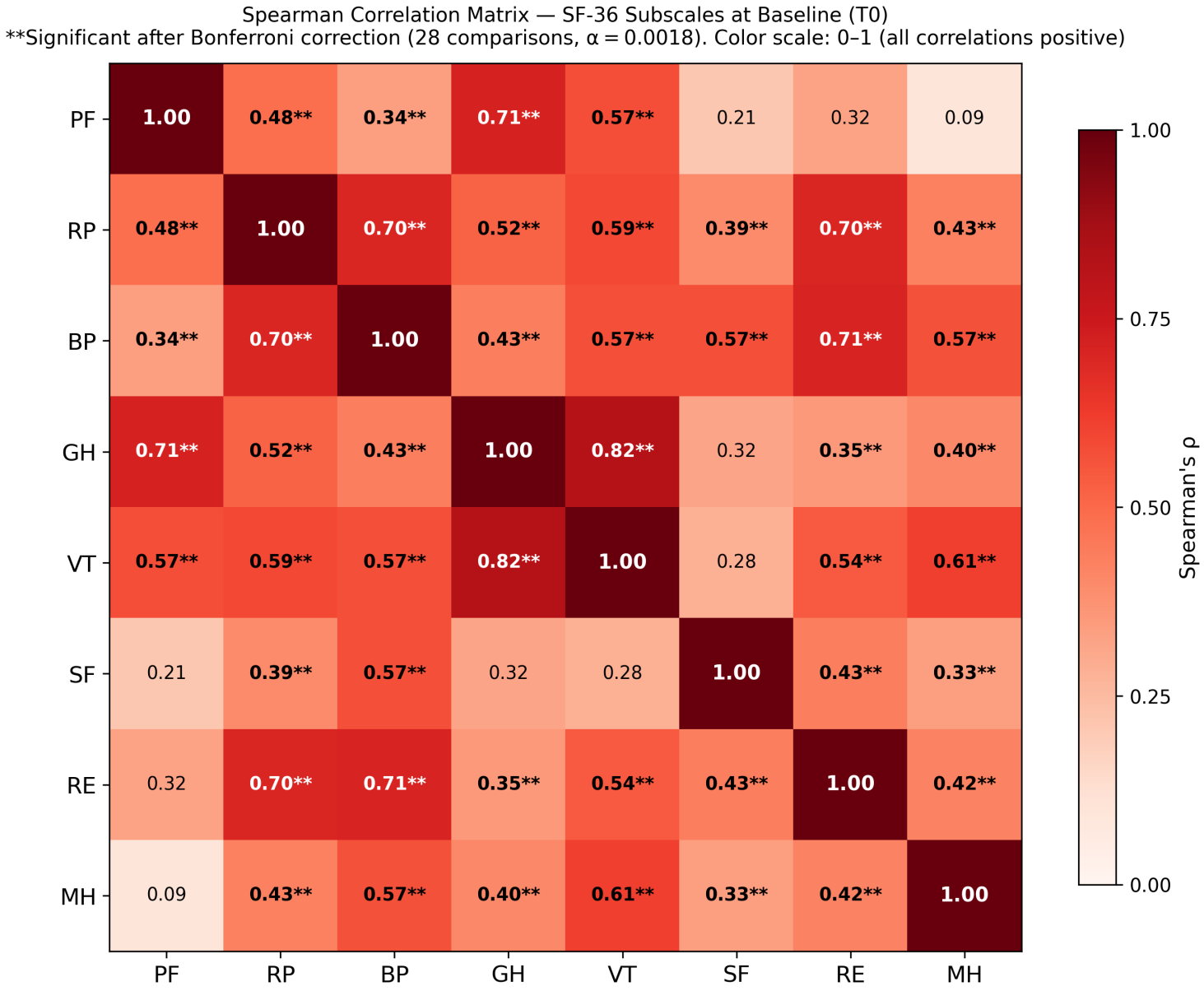


Figure 4

Spearman correlation matrix of SF-36 subscales at baseline (T0). Values represent Spearman's ρ. **Significant after Bonferroni correction (28 comparisons, α=0.0018). Colour scale: 0 – 1 (all pairwise correlations are positive).

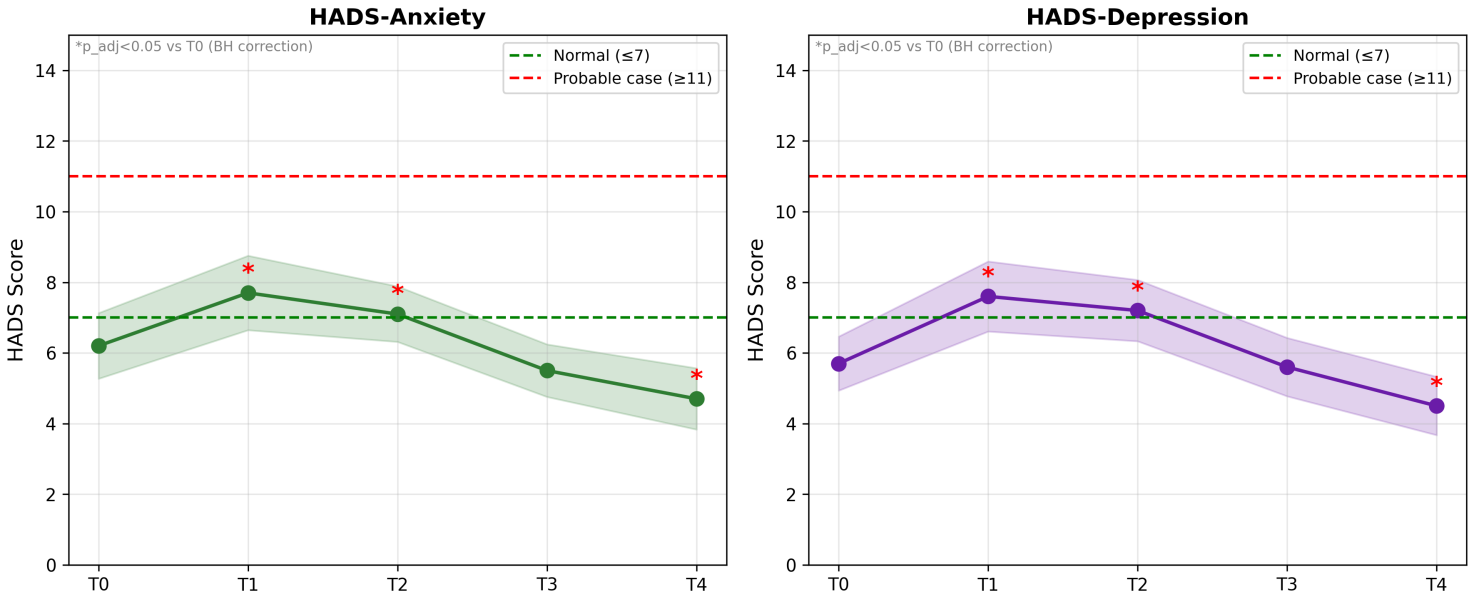


Figure 5

HADS subscale trajectories with 95% CI. Green dashed line: normal range boundary (≤7); red dashed line: probable case threshold (≥11). *p_adj < 0.05 vs. T0 (BH correction). HADS is a screening instrument; threshold crossings are indicative only.

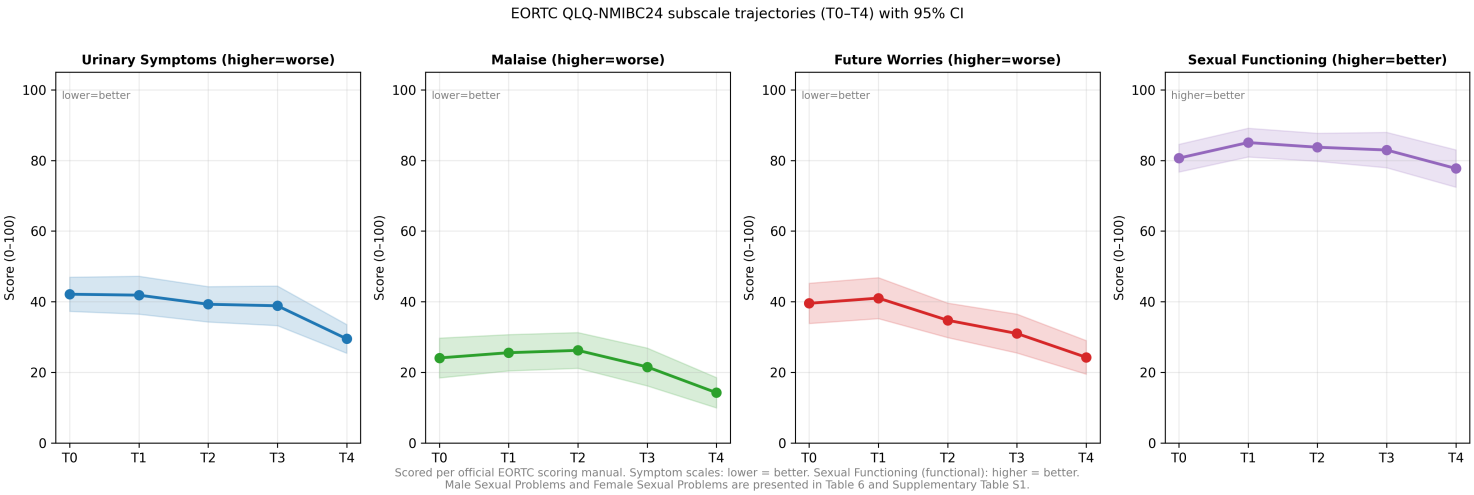


Figure 6

EORTC QLQ-NMIBC24 trajectories for four primary subscales (Urinary Symptoms, Malaise, Future Worries, Sexual Functioning) with 95% CI. Male and Female Sexual Problems are in Table 6 and Supplementary Table S1b. Symptom scales: lower = better. Sexual Functioning (functional): higher = better. Scored per official EORTC manual. Significance markers (asterisks) are not shown; for time-point-level significance, see Table 6.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryS1.docx
- supplementaryS2.docx