Supplementary Files

The impact of intravesical instillations on Quality of Life (QoL) in patients with Non-Muscle Invasive Bladder Cancer (NMIBC): A systematic review

Authors

Ingolf Griebsch[1], Sunil Shrestha[2,3], Yair Lotan[4], Chi-Fai Ng[5] Andrew Bottomley[6]

[1]Global Health Economics & Outcome Research, Ferring Pharmaceuticals A/S, Ferring International PharmaScience Center, 2770 Copenhagen, Denmark

[2]Department of Research and Academics, Kathmandu Cancer Center, Tathali, Nala Road, Bhaktapur, Nepal

[3]Person-Centred Research, Eastern Health Clinical School, Monash University, Melbourne, VIC, Australia

[4]Department of Urology, UT Southwestern Medical Center at Dallas, Dallas, Texas, United States of America.

[5]Department of Surgery, SH Ho Urology Centre, The Chinese University of Hong Kong, Hong Kong, China

[6]Bottomley Consulting Group, Brussels, Belgium

Corresponding Author

Andrew Bottomley, PhD.

Director, Bottomley Consulting Group, Brussels, Belgium

Email ID: Andrew.bottomley@icloud.com

Supplementary Files

Contents

Risk of Bias assessment using the (ROB 2.0) for randomized controlled trials

Koga et al. (2010)

Domain	Judgment	Rationale
1. Bias arising from the randomization process	Low Risk	Randomization was described, and groups were comparable in terms of baseline characteristics (e.g., gender, age, tumor stage). Allocation concealment was not explicitly mentioned but no major imbalances were observed.
2. Bias due to deviations from intended interventions	Low Risk	The study design ensured that participants received treatment as planned and had high adherence to protocols (75% for maintenance therapy). Deviations due to patient withdrawal were not related to treatment efficacy or outcomes.
3. Bias due to missing outcome data	Low Risk	The majority of enrolled patients were evaluated (84 of 90). Only 6 patients were excluded for specific reasons (e.g., protocol violations, no available data), unlikely to introduce significant bias.
4. Bias in measurement of outcomes	Low Risk	Outcomes were objectively measured using validated tools (e.g., recurrence-free survival via cystoscopy, QOL via EORTC QLQ-C30). Investigators assessing recurrence were likely blinded to group assignment.
5. Bias in selection of the reported result	Low Risk	All outcomes specified in the methods (e.g., recurrence-free survival, adverse effects, QOL) were reported comprehensively. No evidence of selective reporting was found.
Overall Risk of Bias	Low Risk	The study was conducted rigorously, with detailed methods and consistent outcome reporting. Minor risks from allocation concealment were unlikely to influence conclusions.

Straten 2023

Domain	Risk of Bias Judgment	Reasoning
Bias Arising from the Randomization Process	Low Risk	Randomization was conducted using a validated program with appropriate stratification factors. Baseline characteristics were balanced.
Bias Due to Deviations from Intended Interventions	Some Concerns	Patients were not blinded to their treatment arm, which could influence self-reported QoL outcomes. However, adherence rates were monitored and reported.
Bias Due to Missing Outcome Data	Low Risk	QoL data collection was robust. While fewer participants completed late assessments (e.g., at T7), statistical adjustments accounted for baseline scores.
Bias in Measurement of the Outcome	Some Concerns	QoL outcomes were self-reported and could have been affected by the lack of blinding. However, a validated tool (EORTC QLQ-C30) was used to mitigate this risk.
Bias in Selection of the Reported Results	Low Risk	The study reported all pre-specified outcomes, and statistical analyses were appropriate for the study design and data available.

Overall Risk of Bias	Some Concerns	The main concern is the lack of blinding, which could affect subjective outcomes like QoL. However, other domains were well-addressed.

Risk of Bias assessment using the Newcastle-Ottawa Quality Assessment Scale for cohort studies

Li Wei et al. (2014)

Domain	Criteria	Assessment	Score
Selection			
1. Representativeness of the cohort	a) Truly representative of the average NMIBC patients in the community	Patients were enrolled from a single hospital, but representative of intermediate/high-risk NMIBC	1
2. Selection of the non-exposed cohort	No non-exposed cohort as this was a single-arm study	Not applicable	N/A
3. Ascertainment of exposure	a) Secure record (e.g., surgical records)	Exposure (intravesical treatment) verified via medical procedures in a clinical setting	1
4. Outcome not present at start	a) Yes	QoL outcomes were measured before starting treatment, ensuring outcomes were not present initially	1
Comparability			
1. Comparability of cohorts	a) Study controls for confounding factors (age, gender, etc.)	No explicit control for confounders via multivariable analysis	0
Outcome			
1. Assessment of outcome	c) Self-report	Outcomes were assessed using self-reported questionnaires (EORTC QLQ-C30, CLSS)	1
2. Follow-up duration	a) Yes (6 weeks follow-up for short-term outcomes)	Follow-up was adequate for assessing short-term QoL outcomes	1
3. Adequacy of follow-up	a) Complete follow-up - all subjects accounted for	All 106 patients completed the 6-week follow-up	1

Total Score: 6/8

Risk of Bias:

- Selection: Low risk (representative cohort, secure ascertainment of exposure).
- Comparability: High risk (no adjustment for confounding factors).
- Outcome: Moderate risk (self-reported outcomes, but complete follow-up).

The risk of bias is moderate, mainly due to the lack of adjustment for confounding factors and reliance on self-reported outcomes.

Wang et al. (2024)

Category	Assessment Criteria	Rating
Selection		

1) Representativeness of the exposed cohort	a) Truly representative of the average NMIBC patients in the community	1
2) Selection of the non-exposed cohort	a) Drawn from the same community as the exposed cohort	1
3) Ascertainment of exposure	a) Secure record (e.g., medical/surgical records)	1
4) Outcome of interest not present at start	a) Yes, outcomes like complications and QOL are measured post-intervention	1
Comparability		
1) Comparability of cohorts based on design or analysis	a) Study controls for baseline characteristics (age, gender, cancer stage)	1
	b) Study controls for additional factors (psychological interventions, compliance)	1
Outcome		
1) Assessment of outcome	a) Independent blind assessment	1
2) Was follow-up long enough for outcomes	a) Yes, follow-up within 1 month post-discharge	1
3) Adequacy of follow-up of cohorts	a) Complete follow-up; all subjects accounted for	1

Total Score: 9/9 score

Interpretation: This study demonstrates a high quality based on the NOS for cohort studies, showing strong selection methods, appropriate comparability between groups, and robust outcome assessment.

Blichert-Refsgaard 2024

Domain	**Item**	**Assessment**	**Stars**
Selection	**1. Representativeness of the exposed cohort**	Truly representative of the average NMIBC patients in Denmark.	1
	2. Selection of the non-exposed cohort	It is not applicable; the study does not compare exposed vs. non-exposed cohorts.	-
	3. Ascertainment of exposure	Secure record (Danish National Patient Registry and Prescription Registry).	1
	4. Demonstration that the outcome of interest was not present at the start of the study	Yes, outcomes were assessed after treatment exposure.	1
Comparability	**1. Comparability of cohorts based on design or analysis**	Controlled for key factors like treatment frequency and instillation exposure in multivariate models.	2
Outcome	**1. Assessment of outcome**	Record linkage from validated national registries.	1

	2. Follow-up long enough for outcomes to occur	Yes, 5 years for treatment impacts and medication use.	1
	3. Adequacy of follow-up of cohorts	Complete follow-up with minimal loss due to comprehensive Danish registries.	1

Total Stars: 8/9

This study demonstrates a low risk of bias based on robust population representation, validated data sources, sufficient follow-up duration, and adequate adjustments for confounding variables. However, it does not include a true non-exposed cohort for direct comparison.

Subiela 2021

Domain	Criteria	Assessment	Score
Selection			
1) Representativeness of the cohort	a) Truly representative of average NMIBC patients in the community	Likely "somewhat representative" as it is a hospital-based cohort, but not a community-based sample	1
2) Selection of non-exposed cohort	a) Drawn from the same community as the exposed cohort	Not applicable (no non-exposed comparison group in this study)	0
3) Ascertainment of exposure	a) Secure record (e.g., surgical/medical records)	Exposure was based on medical records and structured protocols	1
4) Outcome absent at start of study	a) Yes	Patients with muscle-invasive disease or prior radical cystectomy were excluded	1
Comparability			
1) Comparability of cohorts	a) Study controls for important factor (e.g., BCG responsiveness)	Comparability addressed using inverse probability weighting (IPW) methods	2
Outcome			
1) Assessment of outcome	a) Independent blind assessment or record linkage	Outcomes (e.g., PFS, CSS) were derived from medical records and IPW analysis	1
2) Follow-up period	a) Long enough for outcomes to occur	Median follow-up of 66.5 months is adequate for assessing long-term outcomes	1
3) Adequacy of follow-up	a) Complete follow-up, all subjects accounted for	Follow-up details provided; no substantial losses reported that could introduce bias	1

Total Score:7/9 score, indicating a moderate risk of bias.

Interpretation: The study provides robust exposure and outcome data but lacks a non-exposed comparison group, limiting some generalizability. The use of IPW methods strengthens comparability, reducing confounding bias. Detailed follow-up mitigates concerns about incomplete outcome data.

Serretta et al., 2016

Domain	Subdomain	Assessment	Score
Selection	1. Representativeness of the Exposed Cohort	Patients were from two tertiary referral centers, likely representing the target population.	1
	2. Selection of the Non-Exposed Cohort	Not applicable (no comparison group).	
	3. Ascertainment of Exposure	Exposure (BCG treatment) was well-documented in clinical records.	1

	4. Demonstration That Outcome Was Not Present at Start of Study	Outcomes (e.g., dropout rates) were measured after treatment began.	1
Comparability	1. Comparability of Cohorts	Adjustments for potential confounders were not explicitly reported.	
Outcome	1. Assessment of Outcome	Outcomes (dropout, toxicity) were clearly defined and assessed using medical records.	1
	2. Was Follow-Up Long Enough for Outcomes?	Follow-up duration (1 year) was adequate to capture maintenance outcomes.	1
	3. Adequacy of Follow-Up of Cohorts	A dropout rate of ~22% during maintenance was acknowledged, which limits follow-up adequacy.	

Total Score: 5/9 score

- **Risk of Bias: Moderate**
    - Strengths: Good documentation of exposure and outcomes, and a representative patient population.
    - Weaknesses: No explicit control group or confounder adjustments, and incomplete follow-up in some cases.

This moderate risk of bias reflects the limitations of retrospective designs but highlights the study's strengths in outcome assessment and representativeness.

Risk of bias assessment using the Newcastle-Ottawa Quality Assessment Scale (adapted for cross-sectional studies)

Kuperus et al. 2021

Domain	Criteria	Score	Comments
Selection			
Representativeness of the cases	a) Truly representative of the HCC patients (consecutive or random sampling of cases)	1	Patients were consecutively surveyed for IVT, indicating representativeness.
Sample size	b) Not justified (<400 HCC patients included)	0	The sample size is 113 patients, below the threshold of 400 for justification.
Non-Response rate	a) The response rate is satisfactory (≥95%)	1	Response rate was nearly 100% as per the report.
Ascertainment of the screening tool	b) Non-validated screening/surveillance tool, but the tool is available or described	1	The survey was designed for this study and described, but is not a validated tool.
Comparability			
Control for confounding factors	b) The study does not investigate potential confounders	0	No mention of controlling for confounders through subgroup or multivariable analysis.
Outcome			
Assessment of the outcome	c) Self-report	1	The outcomes were based on self-reported surveys from patients.
Statistical test	a) The statistical test used to analyze the data is clearly described and appropriate	1	Statistical analysis was performed using SAS JMP 13, and appropriate tests were applied.

Total Score: 5/9

This assessment highlights areas where the study performed well, particularly in representativeness, response rate, and use of appropriate statistical methods, but it also indicates limitations such as sample size and lack of confounder control.

Alcorn 2019

Domain	Criteria	Assessment	Score
Selection			
1. Representativeness	Truly representative of NMIBC patients (consecutive sampling used in a well-defined population).	Consecutive sampling of NMIBC patients in a defined NHS Cancer Unit.	1
2. Sample Size	Justified and satisfactory (≥400 patients).	Sample size is 234, not justified.	0
3. Non-Response Rate	Response rate is satisfactory (≥95%).	All cases meeting the inclusion criteria were used (response rate not relevant).	1

4. Ascertainment of Tool	Validated tools used for analysis and measurement of withdrawals (e.g., Kaplan-Meier, Cox regression).	Statistical tools are validated, but specific patient-reported outcome tools were not validated or mentioned.	1
Comparability			
5. Confounders Investigated	Study investigates potential confounders using subgroup analysis or regression modeling.	Multivariable Cox regression was used to explore age, side effects, and information factors as confounders.	1
Outcome			
6. Outcome Assessment	Outcome assessed using objective data (retrospective case notes) and statistical linkage.	Retrospective case analysis; no independent blind assessment, but robust statistical analysis.	2
7. Statistical Test	Statistical tests are clearly described and appropriate for the study design.	Kaplan-Meier and Cox regression modeling were used effectively to analyze withdrawals over time.	1

Total Score: 7/9

Risk of Bias Level: Moderate

This study demonstrates robust data analysis techniques and well-defined sampling but falls short in sample size and does not employ independent blind outcome assessment.

Myers 2024

Domain	**Criteria**	**Assessment**	**Score**
Selection (Maximum: 5 points)			
1. Representativeness of the cases	a) Truly representative (consecutive/random sampling): 1 point	Somewhat representative (surveyed online community BCAN PSN; likely non-random sampling)	**1**
2. Sample size	a) Justified and satisfactory (≥400 cases): 1 point	Not justified (233 cases, below threshold)	**0**
3. Non-response rate	a) Satisfactory (≥95% response rate): 1 point	Not satisfactory (response rate not explicitly reported)	**0**
4. Ascertainment of the screening tool	a) Validated tool: 2 points	Described but non-validated tool	**1**
Comparability (Maximum: 1 point)			
1. Confounding factors	a) Investigated via subgroup/multivariable analysis: 1 point	Potential confounders not investigated	**0**
Outcome (Maximum: 3 points)			

1. Assessment of outcome	a) Independent blind assessment or record linkage: 2 points; self-report: 1 point	Self-reported outcomes	**1**
2. Statistical test	a) Appropriate and described: 1 point	Appropriate and described (e.g., REDCap, IQR)	**1**

Total Score: 4/9

Assessment explanation

- Selection: The study had a somewhat representative sample (BCAN PSN participants), but it didn't justify the sample size, and the response rate was unclear.
- Comparability: There was no analysis to control for confounding factors.
- Outcome: The study relied on self-reported data but used statistical methods to analyze outcomes appropriately.

This indicates moderate risk of bias, mainly due to potential sampling issues and a lack of analysis of confounders.

Wildeman 2021

Domain	**Points**	**Reason**
Selection (Maximum 5 points):		
1. Representativeness of the cases	1	Truly representative of NMIBC patients undergoing BCG/MMC treatment (random sampling).
2. Sample size	0	The sample size 80 was not justified (<400 patients included).
3. Non-response rate	0	Response rate was not explicitly mentioned, and withdrawals were noted.
4. Ascertainment of the screening/surveillance tool	2	Validated tools (EORTC BLS-24, Psychosocial Distress Screening Tool) were used.
Selection Subtotal:	**3/5**	
Comparability (Maximum 1 point):		
1. Potential confounders investigated	0	No subgroup or multivariable analyses to account for confounders.
Comparability Subtotal:	**0/1**	
Outcome (Maximum 3 points):		
1. Assessment of the outcome	1	Outcomes were based on self-reported questionnaires by participants.
2. Statistical test	1	Statistical methods were clearly described and appropriate.
Outcome Subtotal:	**2/3**	

Total Score: 5/9 Moderate risk of bias due to small sample size, response rate issues, and lack of confounder adjustments.

Miyake 2021

Domain	Description	Score
Selection (Max: 5)		
Representativeness of cases	The study included 10 NMIBC patients undergoing BCG therapy. The sampling was not random and was only from a single center, limiting generalizability.	0
Sample size	The study had 10 participants, which is below the threshold of 400 for adequate sample size in this context.	0
Non-response rate	The response rate is not explicitly mentioned but appears complete given the small sample size and focused recruitment.	1
Ascertainment of tool	The study used validated tools such as IPSS, EORTC QLQ-C30, FACT-BL, and actigraphy for sleep quality.	2
Comparability (Max: 1)		
Adjustment for confounders	The study did not investigate confounders (e.g., pre-existing comorbidities were excluded, but no subgroup or multivariable analysis was conducted).	0
Outcome (Max: 3)		
Assessment of outcomes	Sleep quality and urinary symptoms were self-reported and monitored via actigraphy, but no independent blind assessment was performed.	1
Statistical test	The statistical methods were appropriate and clearly described, including using the Wilcoxon signed-rank and Mann-Whitney U-tests.	1

Total Score 5/10

The study has a moderate risk of bias due to small sample size, lack of random sampling, and limited adjustment for confounders. However, its use of validated measurement tools and robust statistical analysis adds credibility to its findings.

Risk of Bias Assessment using CASP checklist for qualitative studies and mixed-method studies

Vo 2021

CASP Criteria	Yes	No	Can't tell	Comments
1. Was there a clear statement of the aims of the research?	✓			The study aimed to explore patient-defined priorities and gaps in care for female bladder cancer patients, focusing on physical impacts, emotional well-being, and patient-provider interaction.
2. Is a qualitative methodology appropriate?	✓			Qualitative methodology was appropriate as the study sought to understand the subjective experiences and priorities of the patients.

3. Was the research design appropriate to address the aims of the research?	✓			Focus groups and semi-structured interviews were appropriate for exploring patient experiences and priorities.
4. Was the recruitment strategy appropriate to the aims of the research?	✓			Recruitment was clearly explained, with appropriate inclusion/exclusion criteria. Reasons for non-participation were also addressed.
5. Was the data collected in a way that addressed the research issue?	✓			Data collection methods (focus groups and interviews) were explicitly described, with justification for the methods and discussion of data saturation.
6. Has the relationship between researcher and participants been adequately considered?			✓	There was limited discussion of the researchers' critical reflections on their roles or biases during data collection and analysis.
7. Have ethical issues been taken into consideration?	✓			Ethical approval was obtained (IRB), and informed consent was clearly described. Participant confidentiality and autonomy were maintained.
8. Was the data analysis sufficiently rigorous?	✓			Thematic analysis was rigorous, with independent coding by two raters and sufficient data presentation, supported by quotes.
9. Is there a clear statement of findings?	✓			The findings were explicit, organized into clear themes, and supported by participant quotes. The discussion linked findings to the research questions.
10. How valuable is the research?	✓			The study offers valuable insights for improving care for female bladder cancer patients and addressing gender disparities, with potential applications in clinical practice.

Summary

Category	Details
Positive/Methodologically Sound	- Well-defined objective to explore patient-defined priorities and gaps in care for female bladder cancer patients.
	- Use of qualitative methods (focus groups and semi-structured interviews) to capture subjective experiences.
	- Clearly outlined, with relevant inclusion/exclusion criteria and explanation of reasons for non-participation.
	- Well-documented methods, justified settings, detailed descriptions, and inclusion of data saturation.
	- Ethical approval obtained, with clear procedures for informed consent.
	- Rigorous thematic analysis, independent coding, and data supported by participant quotes.
	- Explicit themes with sufficient discussion linking results to research questions.
	- Provides insights to inform clinical practice and address gender disparities in female bladder cancer care.

Negative/Relatively Poor Methodology	- Limited discussion on researchers' reflection on roles or biases, potentially affecting data interpretation.
Unknowns	- Lack of explicit focus on procedural aspects of BCG/chemotherapy installations and their impact on adherence or quality of life.
	- Predominantly white sample, limiting generalizability to more diverse populations.

Garg 2018

CASP Questions	**Yes**	**No**	**Can't Tell**	**Comments**
Section A: Are the results valid?				
1. Was there a clear statement of the aims?	✓			The study aimed to understand patient experiences, define care priorities, and identify targets for improving NMIBC care.
2. Is a qualitative methodology appropriate?	✓			Qualitative methods were suitable for exploring subjective patient experiences and illuminating their challenges with NMIBC treatments.
3. Was the research design appropriate?	✓			The focus group design was justified to gather in-depth perspectives on the research topic.
4. Was the recruitment strategy appropriate?	✓			Recruitment of NMIBC patients via the Cancer Registry and clinics was explained, but details on non-participants or dropouts were not provided.
5. Was the data collected in a way that addressed the issue?	✓			Semi-structured focus groups with a guide were described. Data collection methods were explicitly stated and justified.
6. Has the relationship between researcher and participants been adequately considered?			✓	While not leading focus groups mitigated the role of the lead investigator (a urologist), potential bias from other researchers was not discussed.
Section B: What are the results?				
7. Have ethical issues been taken into consideration?	✓			Ethical approval was obtained, and confidentiality was maintained. Participant incentives (gift cards, meals) were disclosed.
8. Was the data analysis sufficiently rigorous?	✓			Thematic analysis was used, with iterative coding and concordance among coders. However, saturation was not explicitly mentioned, though consensus was discussed.
9. Is there a clear statement of findings?	✓			Findings were clearly presented, with themes supported by quotes. Contradictory data or outliers were not specifically addressed.
Section C: Will the results help locally?				

10. How valuable is the research?	✓			Findings highlight areas for improving care for NMIBC patients, especially in rural settings, and suggest actionable interventions (e.g., better communication).

Summary

Category	Details
Positive/Methodologically Sound	- Clear aims and justification of qualitative methodology.
	- Transparent description of focus group data collection and thematic analysis.
	- Relevant patient population recruited from diverse clinical settings.
	- Ethical considerations were appropriately addressed.
Negative/Relatively Poor Methodology	- Potential bias due to the urologist's affiliation with some participants was inadequately explored.
	- No explicit mention of data saturation, though coding consensus was achieved.
Unknowns	- Details about reasons for non-participation or dropout were missing.
	- Contradictory data were not discussed in depth.

The study is generally robust, with some minor limitations in reporting. Its findings can be considered trustworthy but should be interpreted cautiously when applied to diverse populations due to the homogeneity of the sample (all white participants from rural areas).

Alcorn et al. (2020)

Question	Yes	No	Can't Tell	Comments
Section A: Are the results valid?				
1. Was there a clear statement of the aims of the research?	✓			The aim of exploring patient experiences and factors influencing withdrawal from BCG treatment is clearly stated.
2. Is a qualitative methodology appropriate?	✓			A qualitative methodology is appropriate as the study focuses on subjective experiences of patients, which aligns with the aim.
3. Was the research design appropriate to address the aims of the research?	✓			Semi-structured interviews and thematic analysis are appropriate for exploring subjective patient experiences.
4. Was the recruitment strategy appropriate to the aims of the research?	✓			The purposive sampling of participants who withdrew from BCG treatment ensures that the data is relevant to the research question.
5. Was the data collected in a way that addressed the research issue?	✓			Semi-structured interviews were conducted, recorded, and transcribed verbatim, allowing in-depth exploration of participant experiences.
6. Has the relationship between researcher and participants been adequately considered?	✓			The researchers were not involved in administering BCG treatment, which minimizes potential bias.

Section B: What are the results?				
7. Have ethical issues been taken into consideration?	✓			Ethical approval was obtained, and participants' consent was clearly outlined.
8. Was the data analysis sufficiently rigorous?	✓			Thematic analysis using NVivo was conducted. Member checking was used to enhance credibility.
9. Is there a clear statement of findings?	✓			Key themes are clearly presented with supporting quotes from participants.
Section C: Will the results help locally?				
10. How valuable is the research?	✓			The study provides insights into patient withdrawal from BCG treatment, identifies unmet needs, and proposes areas for healthcare improvements, such as better communication and support.

Summary of Appraisal:

Positive/Methodologically Sound	Clear research aim, appropriate methodology, and robust data collection and analysis.
	Ethical considerations and member checking enhance credibility.
Negative/Relatively Poor Methodology	Small sample size (6 participants) may limit generalizability.
Unknowns	The long-term impact of the findings is not explored.

Kopenhafer 2024

Section	Yes	No	Can't Tell	Comments
A. Are the results valid?				
1. Was there a clear statement of the aims of the research?	✓			The study evaluates patient experience, unmet needs, and burden among HR-NMIBC patients receiving BCG treatment.
2. Is a qualitative methodology appropriate?	✓			The qualitative interviews are suitable for exploring patient experiences and perceptions.
3. Was the research design appropriate to address the aims?	✓			The mixed-methods approach (qualitative and quantitative) aligns to comprehensively assess patient perspectives and burdens.
4. Was the recruitment strategy appropriate to the aims?	✓			Recruitment via physician referrals ensured participant eligibility but may limit diversity; however, it aligned with study goals.
5. Was data collected in a way that addressed the research issue?	✓			Semi-structured interviews and quantitative surveys were used with clear descriptions of methods, though patient-reported data carries some bias risks.
6. Has the relationship between researcher and participants been adequately considered?			✓	Limited information is provided on researcher reflexivity, potential biases, or influence on participants during interviews or analysis.
B. What are the results?				

7. Have ethical issues been taken into consideration?	✓			Ethical approval was sought, informed consent was obtained, and confidentiality was maintained.
8. Was the data analysis sufficiently rigorous?	✓			Thematic analysis and descriptive statistics were appropriately applied, though the absence of psychometric validation for some tools is a limitation.
9. Is there a clear statement of findings?	✓			Results are explicitly presented with supporting data, though some areas (e.g., reasons for stopping BCG) could benefit from more detailed exploration.
C. Will the results help locally?				
10. How valuable is the research?	✓			The findings contribute to understanding unmet needs and burdens in HR-NMIBC care, with implications for improving patient-centric treatments and policies.

Summary

Category	**Details**
Positive/Methodologically Sound	- Clear aims and relevance to patient care.
	- Appropriate methods (qualitative interviews and quantitative surveys).
	- Ethical standards were upheld, and findings were well presented.
Negative/Relatively Poor Methodology	- Limited discussion of researcher reflexivity and potential bias.
	- Lack of validation for some patient-reported measures.
Unknowns	- Degree to which data reflect diverse patient demographics and experiences.
	- Extent of influence from physician referral recruitment strategy.

Gontero 2013

Domain	Risk of Bias	Reasoning
Bias arising from the randomization process	Low Risk	Randomization Process: The study employed simple 1:1 randomization, and groups were well-balanced regarding baseline characteristics. However, details on allocation concealment or sequence generation were not explicitly reported.
Bias due to deviations from intended interventions	Low Risk	Adherence to Protocol: Both groups followed the treatment protocols as described. One patient in the BCG group discontinued due to severe adverse effects, but overall adherence was high. No evidence of performance bias.
Bias due to missing outcome data	Low Risk	Outcome Data: At 1-year follow-up, 72% of participants were evaluable. Reasons for dropouts were documented (e.g., disease progression, multiple recurrences).

Bias in measurement of the outcome	Some Concerns	Outcome Measurement: QoL outcomes were measured using validated tools (EORTC QLQ-C30, QLQ-BLS24). However, no blinding of outcome assessors was mentioned, which could introduce detection bias.
Bias in selection of the reported result	Low Risk	Selective Reporting: The study reported all pre-specified outcomes, including both QoL and clinical outcomes like recurrence and progression rates. There is no evidence of selective reporting.
Overall Risk of Bias	Some Concerns	While most domains showed a low risk of bias, the absence of blinding in outcome assessment introduces some concerns in the measurement of the outcomes domain.

Abbreviations

BCAN PSN – Bladder Cancer Advocacy Network Patient Survey Network; CASP – Critical Appraisal Skills Programme; CLSS – Core Lower Urinary Tract Symptom Score; CSS – Cancer-Specific Survival; EORTC QLQ-C30 – European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30; FACT-BL – Functional Assessment of Cancer Therapy – Bladder; HR-NMIBC – High-Risk Non-Muscle Invasive Bladder Cancer; IPSS – International Prostate Symptom Score; IPW – Inverse Probability Weighting; IQR – Interquartile Range; IRB – Institutional Review Board; NMIBC – Non-Muscle Invasive Bladder Cancer; NOS – Newcastle-Ottawa Scale; PFS – Progression-Free Survival; QLQ-BLS24 – EORTC Bladder Cancer–Specific Quality of Life Module; QoL – Quality of Life; REDCap – Research Electronic Data Capture; ROB 2.0 – Risk of Bias 2.0.