Published in final edited form as:
Urology. 2022 November ; 169: 134–140. doi:10.1016/j.urology.2022.07.057.

An Evaluation of Peer-Rated Surgical Skill and its Relationship With Detrusor Muscle Sampling in Transurethral Resection of Bladder Tumor

Minh N. Pham[1], **Oliver S. Ko**[1], **Reiping Huang**[2], **Amanda X. Vo**[1], **Kyle P. Tsai**[1], **Jeremy D. Lai**[1], **Matthew T. Hudnall**[1], **Joshua A. Halpern**[1,2], **Joshua J. Meeks**[1], **Jonas Benson**[3,4,5], **Ricardo Soares**[6], **Ronald Kim**[7,8], **Karl Y. Bilimoria**[2], **Jonah J. Stulberg**[9], **Gregory B. Auffenberg**[10]

[1]Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, IL

[2]Surgical Outcomes and Quality Improvement Center, Feinberg School of Medicine, Chicago, IL

[3]Northwestern Medicine Central DuPage Hospital, Winfield, IL

[4]Northwestern Medicine Delnor Hospital, Geneva, IL

[5]UroPartners LLC, Winfield, IL

[6]Department of Urology, Northwestern Medicine Kishwaukee Hospital, DeKalb, IL

[7]Northwestern Medicine Lake Forest Hospital, Lake Forest, IL

[8]UroPartners LLC, Lake Forest, IL

[9]Department of Surgery, University of Texas Health Science Center at Houston, Houston, TX

[10]Urology of St. Louis, St. Louis, MO

Abstract

Objective: To assess the reliability of peer-review of TURBT videos as a means to evaluate surgeon skill and its relationship to detrusor sampling.

Methods: Urologists from an academic health system submitted TURBT videos in 2019. Ten blinded peers evaluated each surgeon's performance using a 10-item scoring instrument to quantify surgeon skill. Normalized composite skill scores for each surgeon were calculated using peer ratings. For surgeons submitting videos, we retrospectively reviewed all TURBT pathology results (2018-2019) to assess surgeon-specific detrusor sampling. A hierarchical logistic regression model was fit to evaluate the association between skill and detrusor sampling, adjusting for patient and surgeon factors.

Results: Surgeon skill scores and detrusor sampling rates were determined for 13 surgeons performing 245 TURBTs. Skill scores varied from −6.0 to 5.1 [mean: 0; standard deviation

Corresponding author: Gregory B. Auffenberg, M.D., M.S., Urology of St. Louis, 12855 N. 40 Drive, Suite 375, St. Louis, MO 63141. gbanwu@gmail.com.

SUPPLEMENTARY MATERIALS
Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.urology.2022.07.057.

(SD): 2.40]. Muscle was sampled in 72% of cases, varying considerably across surgeons (mean: 64.5%; SD: 30.7%). Among 8 surgeons performing >5 TURBTs during the study period, adjusted detrusor sampling rate was associated with sending separate deep specimens (odds ratio [OR]: 1.97; 95% confidence interval [CI]: 1.02-3.81, P = .045) but not skill (OR: 0.81; 95% CI: 0.57-1.17, P = .191).

Conclusion: Surgeon skill was not associated with detrusor sampling, suggesting there may be other drivers of variability of detrusor sampling in TURBT.

Transurethral resection of a bladder tumor (TURBT) is the diagnostic and therapeutic backbone for managing non-muscle invasive bladder cancer (NMIBC). Despite being an ubiquitous urologic procedure grounded in widely taught principles,[1,2] substantial variability exists across surgeons in achieving critical endpoints, such as detrusor muscle sampling and complete tumor resection.[3] For instance, rates of detrusor sampling vary from 36% to 90% in published series.[4-9] The absence of muscle in specimen has oncologic ramifications: failure to sample muscle is associated with residual tumor, recurrence, and inaccurate staging.[10-12]

Despite clear recognition of surgeon-level variation in outcomes, less is known about why such broad variation is observed. Previous work, for other surgical procedures, has demonstrated that skill can be reliably quantified through peer-review of surgical videos and is associated with differences in outcomes.[13-15] Surgeon skill is thought to influence TURBT resection quality,[3] but variation in TURBT skill has never been objectively evaluated nor has its link with clinical outcomes, including detrusor sampling, a marker of resection quality that may be verified pathologically.

In this context, we investigate whether peer-review of TURBT videos can be used to quantify TURBT surgeon skill and whether skill is associated with observed differences in detrusor sampling across surgeons. To achieve this, we asked surgeons from a large academic health system to submit a video of a TURBT they performed for peer-review. Peer surgeons reviewed submitted videos in a blinded process, rating performance to derive a composite skill score for each surgeon. Next, we used retrospective data to assess the relationship between skill scores and detrusor sampling. We hypothesized that peer raters would be able to determine variation in skill and that higher skill ratings would be associated with increased rates of detrusor sampling.

MATERIALS AND METHODS

Study Environment

The study was conducted in an academic health system comprised of 10 hospitals with 37 urologists who performed TURBTs in a health system hospital during 2019. This study was deemed exempt by the Institutional Review Board at our institution.

Surgical Video Peer-Review

We invited all attending urologists practicing within our health system to submit a full-length, unedited video of a TURBT they personally performed without trainee involvement.

Surgeons were asked to submit a video of a resection representative of their technique. Participating surgeons recorded a video using recording equipment available on the operating room video tower. Submissions were edited to remove only elements where the scope was not within the bladder and uploaded to an online, institutional platform created for surgical video peer-review. Using this platform, we invited attending urologists within our health system, blinded to the operating surgeon, to review submitted videos. Each video was anonymously evaluated by 10 reviewers.

Reviewers completed a 2-component online scoring instrument. In the absence of procedure-specific precedents, a 10-item questionnaire (Supplemental Table 1) was derived in part from the Objective Structured Assessment of Technical Skill (OSATS) survey, a validated instrument used in video-based assessments in other surgical disciplines.[13-18] Using a 5-point Likert scale, reviewers evaluated the surgeon's (1) respect for tissue, (2) time and motion, (3) instrument handling, (4) flow of operation, and (5) overall performance. A score of 3 represented the performance expected of an average attending surgeon with higher scores representing greater skill. In addition to the OSATS domains assessed, reviewers completed a 5-item checklist (answering yes or no), evaluating whether 5 TURBT-specific elements were performed including: (1) a full cystoscopic evaluation before resection, (2) consistent resection toward the lens, (3) adequate hemostasis for visualization/resection, (4) complete tumor resection, and (5) muscle visualization with appropriate depth.

Clinical Data Source

To study the relationship between detrusor sampling and technical skill, a retrospective review of all TURBTs performed by participating surgeons between September 2018 and September 2019 was completed. Using our institution's electronic medical record-based data repository and manual chart review of operative and pathology reports, we created a database containing all patient, procedural, and pathologic characteristics for TURBTs performed during the measurement period. Pathologic review was completed by the surgeon's institutional pathologist. Included cases were classified as either primary TURBT or re-TURBT, defined as a case performed within 8 weeks after a primary TURBT at the same institution. To mitigate variable reporting of muscle sampling in benign specimens, we excluded primary TURBTs with benign pathology (n = 70) and cases with malignant, non-urothelial pathology (n = 15). The presence or absence of muscularis propria was reported in all reviewed cases of TURBT.

Outcomes and Covariates

The results of peer-review were used to assign each participating surgeon a composite technical skill score (elaborated below). Our primary outcome was detrusor sampling, defined as the presence of detrusor muscle in pathologic specimen. Several surgeon-specific covariates were included in our regression models including: years in attending practice at the time of the study, case volume (the number of TURBTs performed during our study period), any fellowship training, and formal training in a Society of Urologic Oncology-accredited (SUO) fellowship. We accounted for patient age, gender, and Charlson comorbidity index. Procedural and pathologic factors included TURBT type (primary vs re-TURBT), the collection of a separate deep specimen intraoperatively for analysis (determined by review of the case's pathology report), and difficult tumor location (tumor at the dome or anterior bladder wall).

Statistical Analysis

The composite technical skill score was created using items from the scoring instrument. For each item, we first assessed inter-rater reliability (IRR) using Krippendorff's alpha for binary items and the G(q,k) reliability estimator for ordinal items.[19,20] An IRR coefficient of 0.8 to 1.0 was considered strong; 0.6 to 0.8 was very good; 0.4 to 0.6 was moderate; and 0 to 0.4 was weak. Each item with moderate IRR or better (IRR ≥0.4) was included in the composite technical skill score. The composite technical skill score was calculated as the normalized average of 6 items with IRR ≥0.4 (resection toward the lens + 5 technical performance elements).

Surgeon-specific rates of detrusor sampling were calculated for all TURBTs within our data. Univariate analyses of the association between detrusor sampling and each surgeon, patient, and procedural characteristic were performed using Kruskal-Wallis tests for continuous variables or χ^2 testing for categorical variables with expected counts >5. In cases of categorical variables with expected counts ≤5, Fisher's exact tests were used.

To evaluate the association between composite technical skill scores and detrusor sampling, we fit a 2-level hierarchical logistic regression model with the provider as the random intercept, using surgeon and non-surgeon covariates as fixed effects. We restricted this analysis to surgeons with >5 TURBTs to increase the stability of surgeon-specific estimates. The model included the following covariates: surgeon variables included composite technical skill score, surgeon experience, and case volume. Non-surgeon variables included patient age, gender, Charlson comorbidity index, TURBT type, sending a separate deep specimen intraoperatively, and difficult tumor location. Statistical analyses were performed using SAS 9.4 software (Cary, NC). Statistical significance was considered at a 2-tailed *P* value <.05.

RESULTS

Participant Surgeons

Of 37 surgeons performing TURBT at our institution, 13 surgeons submitted videos for peer-review. Surgeons submitting a video accounted for 57% of TURBTs within our system (330 of 575) between September 2018 and September 2019. After excluding cases with benign or non-urothelial histology, these surgeons completed 245 TURBTs in 191 patients. Surgeon and non-surgeon characteristics are detailed in Supplemental Table 2. All surgeons were men with attending experience ranging from 0 to 31 years (median: 8, inter-quartile range [IQR]: 2-17). Six surgeons practiced in academic settings (46%). Seven participants (54%) had fellowship training; 3 trained in SUO-accredited fellowships (23%).

Peer-Review and Inter-Rater Reliability

Thirteen videos were submitted for scoring, each reviewed by a different set of 10 raters. A total of 130 reviews were completed by 17 raters. The median number of videos reviewed by a given rater was 8 videos (range 1-12). Individualized IRR estimates for each scoring item are in Table 1, with IRR ranging from 0.060 to 0.808. In 4 of 5 TURBT-specific objectives, IRR estimates were weak (0.060 to 0.332). The item with the strongest IRR was determining consistent resection toward the lens (IRR: 0.808). Moderate IRR from 0.407 to 0.667 was reached on all 5 technical performance elements.

Surgeon-Specific Skill

We constructed composite technical skill scores using 6 items with IRR ≥0.4 (Table 1). The technical score distribution among 13 surgeons ranged between −5.99 and 5.05 (mean: 0; standard deviation [SD]: 2.40) (Fig. 1). Most surgeons (85%) were clustered within approximately 2 points of the mean. Among participants with >5 TURBTs (n = 8), the mean score was −0.16 (SD: 3.00) with 75% of surgeons scoring within approximately 2 SDs of the mean.

Procedural Characteristics and Detrusor Muscle Sampling

A total of 191 patients underwent 245 TURBTs with participating surgeons during the study period. Our patient cohort consisted of elderly patients with multiple comorbidities who were predominantly male and White (Supplemental Table 2). Of 245 cases, 27% were re-resections. Separate deep specimens were sent during 37% of cases.

Muscle was present in 177 specimens (72%). To assess for heterogeneity in detrusor sampling, we calculated provider-specific rates: substantial variability in sampling existed across 13 study participants ranging from 0 to 100% (mean: 64.5%; SD: 30.7%). After restricting our analysis to providers with >5 TURBTs, detrusor sampling rates ranged from 50% to 90% among 8 surgeons (mean: 70.4%; SD: 11.3%) (Fig. 2).

Association Between Detrusor Muscle Sampling and Technical Skill

On univariable analysis (Supplemental Table 3), both lower technical skill ($P = .006$) and difficult tumor location ($P = .014$) were significantly associated with detrusor sampling. Surgeon skill was inversely related to detrusor sampling. Other factors evaluated were not associated with detrusor sampling, including surgeon experience ($P = .853$), case volume ($P = .324$), fellowship status ($P = .417$), female patient gender ($P = .116$), and sending a separate deep specimen ($P = .059$). After adjusting for patient, procedural, and surgeon factors, however, we found that surgeon skill score was no longer associated with detrusor sampling (odds ratio [OR]: 0.81; 95% confidence interval [CI]: 0.57-1.17; $P = .191$) (Table 2). There was a significant positive association between sending a separate deep sample and detrusor sampling (OR: 1.97; 95% CI: 1.02-3.81; $P = .045$). The model did not demonstrate a relationship between the other included factors and detrusor sampling (Table 2).

COMMENT

In this study, we evaluated surgeon skill in TURBT through peer-review of surgical videos from 13 urologists and investigated the link between skill and detrusor muscle sampling. Using a multi-component scoring instrument, blinded peer-reviewers assessed skill and demonstrated modest reliability in differentiating skill amongst participating surgeons, especially for TURBT-specific items. After developing a composite skill score, we demonstrated a moderate degree of variation in skill across surgeons. After accounting for covariates using regression, we did not detect a relationship between composite skill scores and detrusor sampling. A positive relationship between sending a separate deep sample and detrusor sampling was demonstrated.

A review process of surgical technique and outcomes provides objective information important to understanding and improving variation in skill and outcomes. Prior work from bariatric and colorectal surgery has shown that peer-review of surgical videos can reliably assess variation in skill and that surgeons with higher skill ratings achieve better outcomes.[13,14] These works have established a foundation for efforts to inform video-based interventions and surgical coaching to improve surgeon skill in an effort to improve outcomes.[21-24] This study represents the first attempt to quantify TURBT skill. We demonstrate modest reliability in assessing skill and show that skill varies across surgeons.

The relationship between surgeon skill and detrusor sampling during TURBT has never been objectively assessed. Given knowledge that lack of detrusor sampling has been associated differences in residual tumor, recurrence, and inaccurate staging,[10-12] a need exists to better understand how to reduce this variation. We hypothesized that a relationship between surgeon skill and detrusor sampling existed; however, this link was not found in the current study. Documenting such a connection would have been an important step in establishing a rationale for interventions to improve TURBT skill.

The lack of association between skill and detrusor sampling may have several possible explanations. First, our composite skill score was determined primarily using technical performance elements from our scoring instrument that measured operative fluidity and finesse rather than successful achievement of TURBT endpoints. It is possible these 2 areas are not related and that our instrument may lack the specificity to reliably discriminate TURBT skills important for explaining outcome variation. Second, we demonstrate poor to modest IRR on many items used in our assessment tool. This raises the possibility that the definition of a “skillful” TURBT may not be shared between reviewers. It is also possible that the definition of a skillful TURBT may not be shared across surgeons due to the relatively unstructured nature of the procedure. This lack of a shared definition of success may explain some of the known variation in detrusor sampling and other procedural outcomes. This speaks to a need to better describe the objective goals of TURBT and to clarify the definition of procedural success.

Our work must be interpreted in the context of several limitations. First, our scoring instrument identified challenges in reliably discriminating whether various procedure-specific endpoints were met. This may be unmasking the high degree of subjectivity involved in determining what truly defines an optimal TURBT or could indicate a need for a better validated scoring instrument for TURBT. Second, our limited sample size of both surgeons and procedures may lead to unstable estimates of performance. We sought to minimize this possibility by excluding low-volume providers from our analysis, but longer-term follow-up with larger surgeon-specific cohorts may also lead to more robust results. A trainee confounder may be present in the retrospective analysis: while our measurement of skill was determined using peer-review of procedures performed by attending surgeons themselves, detrusor sampling rates among academic surgeons may be impacted by trainee involvement in procedures within the dataset. Finally, in the context of our goal to measure system-wide outcomes, we did not have a sufficient sample with longitudinal follow-up data from newer system sites to allow for measurement of more distant outcomes such intravesical therapy use, disease-free or progression-free survival. This is another long-term goal of this work.

Despite these limitations, there are important implications to this work. This represents the first published attempt to objectively quantify TURBT skill using peer-review and to evaluate the link between skill and detrusor sampling. The demonstrated variation in surgeon skill is an important finding. Quantifying variation in skill may represent a first step in identifying reproducible technical performance elements that can be disseminated to improve surgeon skill and subsequent outcomes. Despite demonstrating variation in skill, we did not demonstrate a clear association between skill and detrusor sampling. As stated, we suspect a multi-factorial explanation for this observation. However, the limitations of our scoring instrument speaks to a need for shared definitions of success among surgeons in our study and more broadly. As a next step, we plan to work on development of a validated instrument to evaluate TURBT skill, in order to study skill more effectively and develop a reproducible definition of procedural success. With that better established, it will allow for more robust study of variation in skill and assessment of the link between skill and outcomes, including longer term outcomes such as survival. If we remain unable to demonstrate that skill is associated with detrusor sampling or other TURBT outcomes, we will have to search for other mechanisms to improve TURBT. However, if the link between skill and outcomes can be demonstrated, then improvement initiatives tailored to help surgeons improve skill may represent an important strategy to improve the lives of patients with bladder cancer. This study is an early step toward better understanding TURBT and its relationship with the intermediate outcome of detrusor sampling. It also indicates many opportunities remain to improve procedural understanding, quality of this procedure, and outcomes for patients with bladder cancer.

CONCLUSION

Peer-review of surgical TURBT videos can discriminate skill with modest reliability. There was variation in surgeon-specific skill and considerable variation in detrusor sampling. Technical skill, as measured using our instrument, was not associated with detrusor sampling.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. American Urologic Association & Society of Urologic Oncology. Diagnosis and treatment of non-muscle invasive bladder cancer. 2020. https://www.auanet.org/documents/education/clinical-guidance/non-muscle-invasive-bladder-cancer.pdf.

2. National Comprehensive Cancer Society. NCCN guidelines for bladder cancer. 2020. https://jnccn.org/view/journals/jnccn/18/3/article-p329.xml.
3. Brausi M, Collette L, Kurth K, et al. Variability in the recurrence rate at first follow-up cystoscopy after TUR in stage Ta T1 transitional cell carcinoma of the bladder: a combined analysis of seven EORTC Studies. European Urology. 2002;41:523–531. 10.1016/S0302-2838(02)00068-4. [PubMed: 12074794]
4. Gendy R, Delprado W, Brenner P, et al. Repeat transurethral resection for non-muscle-invasive bladder cancer: a contemporary series. BJU Int. 2016;117(suppl 4):54–59. 10.1111/bju.13265. [PubMed: 26486968]
5. Capogrosso P, Capitanio U, Ventimiglia E, et al. Detrusor muscle in TUR-derived bladder tumor specimens: can we actually improve the surgical quality? J Endourol. 2016;30:400–405. 10.1089/end.2015.0591. [PubMed: 26597334]
6. Shoshany O, Mano R, Margel D, Baniel J, Yossepowitch O. Presence of detrusor muscle in bladder tumor specimens–predictors and effect on outcome as a measure of resection quality. Urol Oncol. 2014;32:40. 10.1016/j.urolonc.2013.04.009
7. Gordon PC, Thomas F, Noon AP, Rosario DJ, Catto JWF. Long-term outcomes from re-resection for high-risk non-muscle-invasive bladder cancer: a potential to rationalize use. Eur Urol Focus. 2019;5:650–657. 10.1016/j.euf.2017.10.004. [PubMed: 29089252]
8. Mariappan P, Johnston A, Padovani L, et al. Enhanced quality and effectiveness of transurethral resection of bladder tumour in non-muscle-invasive bladder cancer: a multicentre real-world experience from Scotland’s quality performance indicators programme. Eur Urol. 2020;78:520–530. 10.1016/j.eururo.2020.06.051. [PubMed: 32690321]
9. Das A, Cohen JE, Ko OS, et al. Surgeon scorecards improve muscle sampling on transurethral resection of bladder tumor and recurrence outcomes in patients with nonmuscle invasive bladder cancer. J Urol. 2021;205:693–700. 10.1097/JU.0000000000001372. [PubMed: 33021430]
10. Dutta SC, Smith JA, Shappell SB, Coffey CS, Chang SS, Cookson MS. Clinical under staging of high risk nonmuscle invasive urothelial carcinoma treated with radical cystectomy. J Urol. 2001;166:490–493. [PubMed: 11458053]
11. Herr HW. The value of a second transurethral resection in evaluating patients with bladder tumors. J Urol. 1999;162:74–76. 10.1097/00005392-199907000-00018. [PubMed: 10379743]
12. Mariappan P, Zachou A, Grigor KM, Edinburgh Uro-Oncology Group. Detrusor muscle in the first, apparently complete transurethral resection of bladder tumour specimen is a surrogate marker of resection quality, predicts risk of early recurrence, and is dependent on operator experience. Eur Urol. 2010;57:843–849. 10.1016/j.eururo.2009.05.047. [PubMed: 19524354]
13. Birkmeyer JD, Finks JF, O’Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013;369:1434–1442. 10.1056/NEJMsa1300625. [PubMed: 24106936]
14. Stulberg JJ, Huang R, Kreutzer L, et al. Association between surgeon technical skills and patient outcomes. JAMA Surg. 2020. 10.1001/jamasurg.2020.3007.
15. Brajcich BC, Stulberg JJ, Palis BE, et al. Association between surgical technical kkill and long-term survival for colon cancer. JAMA Oncol. 2021;7:127–129. 10.1001/jamaoncol.2020.5462. [PubMed: 33125472]
16. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84:273–278. 10.1046/j.1365-2168.1997.02502.x. [PubMed: 9052454]
17. Scally CP, Varban OA, Carlin AM, Birkmeyer JD, Dimick JB. Michigan bariatric surgery collaborative. Video ratings of surgical skill and late outcomes of bariatric surgery. JAMA Surg. 2016;151: e160428. 10.1001/jamasurg.2016.0428. [PubMed: 27074114]
18. Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current status of technical skills assessment tools in surgery: a systematic review. J Surg Res. 2020;246:342–378. 10.1016/j.jss.2019.09.006. [PubMed: 31690531]
19. Krippendorff K. Computing Krippendorff’s alpha-reliability. Departmental Papers (ASC). Available at: https://repository.upenn.edu/asc_papers/43. Accessed December 18, 2020.

20. Putka DJ, Le H, McCloy RA, Diaz T. Ill-structured measurement designs in organizational research: implications for estimating interrater reliability. J Appl Psychol. 2008;93:959–981. 10.1037/0021-9010.93.5.959. [PubMed: 18808219]
21. Vande Walle KA, Quamme SRP, Beasley HL, et al. Development and assessment of the Wisconsin surgical coaching rubric. JAMA Surg. 2020;155:486–492. 10.1001/jamasurg.2020.0424. [PubMed: 32320026]
22. Pradarelli JC, Yule S, Panda N, et al. Optimizing the implementation of surgical coaching through feedback from practicing surgeons. JAMA Surg. 2021;156:42–49. [PubMed: 33052407]
23. Cathcart P, Sridhara A, Ramachandran N, Briggs T, Nathan S, Kelly J. Achieving quality assurance of prostate cancer surgery during reorganisation of cancer services. Eur Urol. 2015;68:22–29. 10.1016/j.eururo.2015.02.028. [PubMed: 25770482]
24. Prebay ZJ, Peabody JO, Miller DC, Ghani KR. Video review for measuring and improving skill in urological surgery. Nat Rev Urol. 2019;16:261–267. 10.1038/s41585-018-0138-2. [PubMed: 30622365]

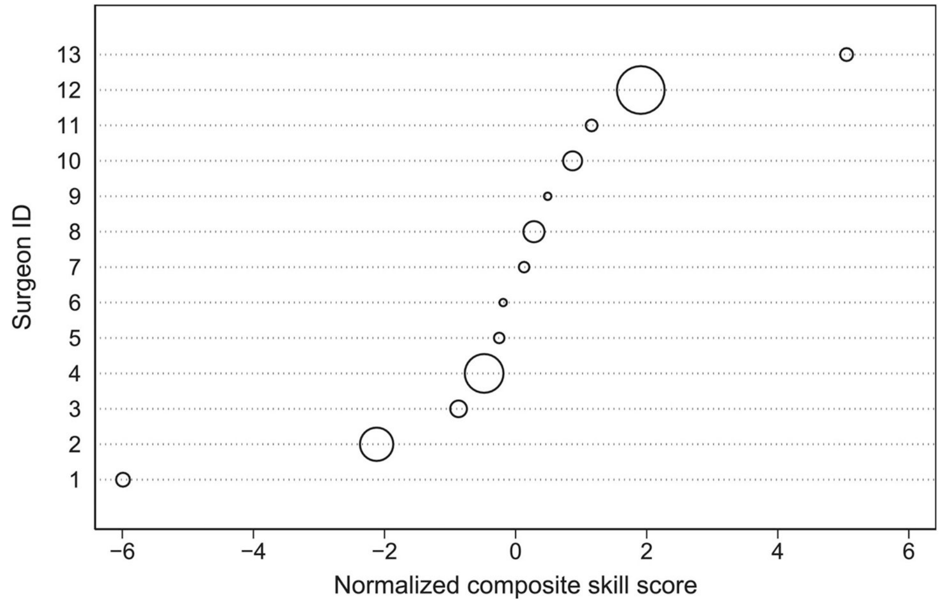

Figure 1.

Surgeon-level variation in TURBT technical skill scores. ID, identifier; TURBT, transurethral resection of bladder tumor. Dots are scaled to represent surgeon TURBT volume in the cohort (range: 2 to 82 cases).

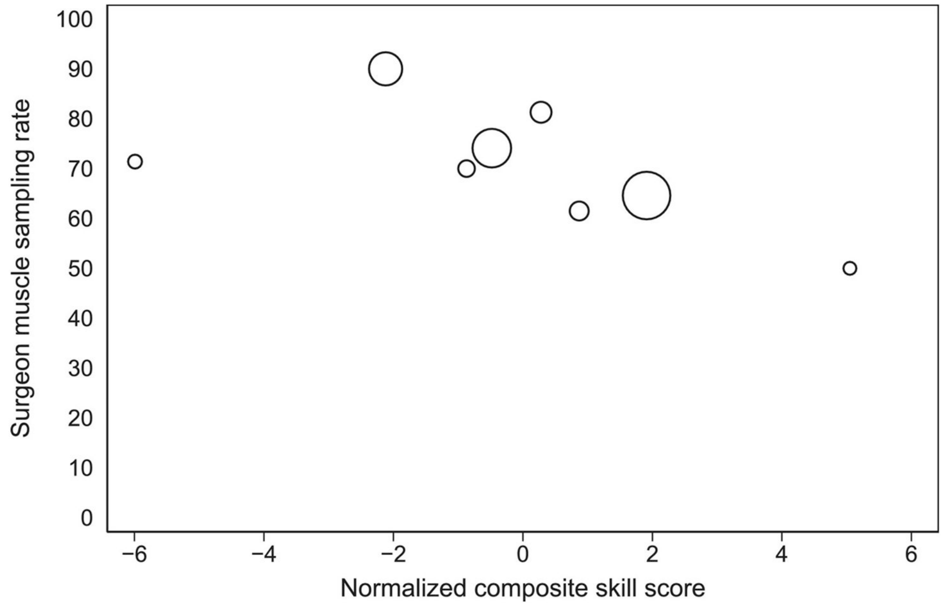

Figure 2.
Surgeon-level detrusor muscle sampling rates in relation to TURBT technical skill scores. Detrusor muscle sampling rates and normalized composite skill scores are shown for surgeons with > 5 cases. Dots are scaled to represent surgeon TURBT volume in the cohort (range: 6 to 82 cases).

Table 1.

TURBT skill assessment instrument and inter-rater reliability of each domain

Procedure-Specific Checklist †	**IRR** §
Full cystoscopy prior to resection	0.3
Consistent resection toward the lens	0.808 *
Hemostasis adequate for resection/visualization	0.06
Complete tumor resection	0.332
Muscle visualization with appropriate depth	0.152
Technical performance elements ‡	**IRR** §
Flow of operation	0.667 *
Instrument handling	0.554 *
Respect for tissue	0.613 *
Time and motion	0.406 *
Overall performance	0.614

IRR, inter-rater reliability.

0.8 - 1.0 Strong.

0.6 - 0.8 Very good.

0.4 - 0.6 Moderate.

0 - 0.4 Weak.

*Item was used to construct a normalized composite skill score.

†IRR for binary procedure-specific skill items was estimated by Krippendorff's alpha.

‡IRR for 5-point items was calculated by G statistics.

§IRR interpretation.

Table 2.

Hierarchical logistic regression model assessing predictors of detrusor muscle sampling*

Surgeon Factors	**OR**	**95% CI**	***P*-value**
Technical skill score †	0.81	0.57 - 1.17	0.191
Years of practice	1.01	0.91 - 1.12	0.768
TURBT case volume	1.01	0.99 - 1.03	0.312
Patient factors	**OR**	**95% CI**	***P*-value**
Age (10 years) ‡	1.18	0.90 - 1.56	0.222
Female gender	0.45	0.19 - 1.05	0.06
Charlson comorbidity index §	0.93	0.84 - 1.03	0.174
Procedural/pathologic factors	**OR**	**95% CI**	***P*-value**
Primary TURBT (reference: re-TURBT)	0.71	0.35 - 1.48	0.301
Separate deep specimen sent by surgeon	1.97	1.02 - 3.81	0.045
Tumor present at the dome / anterior bladder	0.3	0.06 - 1.54	0.11

CI, confidence interval; OR, odds ratio; SE, standard error; TURBT, transurethral resection of bladder tumor.

Random intercept: surgeon.

Variance (SE): 0.42 (0.65).

χ^2 test statistics (*P*-value): .89 (.173).

*Limited to 8 providers who had >5 cases in data (n=228 cases in total after exclusion).

Hierarchical logistic model with a random intercept for provider, using residual pseudo-likelihood estimation method. *P*-values are for 2-sided, Type III F tests with significance level at 0.05.

†Normalized composite score for surgeon technical skill in performing TURBT.

‡Odds ratio estimated for an increase of 10 years from mean age.

§Odds ratio estimated for each increase in 1 point of Charlson comorbidity index.