BJU Int 2025 doi:10.1111/bju.16789

Check for updates

Trainees' Corner

BJUI
BJU International

Research Letter

Residents surgical training using metrics for transurethral resection of bladder tumours

Transurethral resection of bladder tumours (TURBT) is one of the most common endoscopic procedures performed by urologists around the world. Moreover, the quality of TURBT can play a relevant role in the correct staging of non-muscle-invasive vs muscle-invasive bladder cancer. Therefore, TURBT must be considered as one of the most important procedures that residents should perform confidently [1].

Several studies have raised concerns regarding the ability of European residents in urology to perform a TURBT safely and effectively [2]. In this context, the introduction of performance metrics specific to TURBT represents a crucial step forward in improving surgical training. These performance metrics, which assess errors and successes systematically and objectively during the procedure, offer a more reliable alternative to traditional educational tools. While virtual reality simulators and structured proctoring have proven effective in improving technical skills, evidence indicates that these tools lack universal validation, particularly for more complex procedures [3]. Performance metrics, on the other hand, have been designed to address these gaps, providing a more accurate and objective assessment of residents' skills. Evidence suggests that the adoption of quantitative metrics based on critical errors and measurable performance can significantly reduce execution errors, improving not only resident training but also long-term clinical outcomes [4]. In order to improve the surgical training programme in Europe, an international consensus panel of experts, developed a performance metrics that could be a reference for the approach to a TURBT and could be used for the training curriculum of residents [5]. According to the performance metrics, TURBT has been divided into six phases and 63 steps: Phase 1: patient positioning and preparation (Steps 1–5); Phase 2: preparation of instrumentation and operative field (Steps 6–14); Phase 3: insertion of the instrument into the bladder (Steps 15–26); Phase 4: diagnostic evaluation of the bladder (Steps 27–35); Phase 5: tumour resection, haemostasis and evacuation (Steps 36–53); and Phase 6: instrument removal (Steps 54–63) [3]. Furthermore, 27 errors and 19 critical errors have been described.

Errors include failure of patient positioning, such as the anus not being positioned on the edge of the table or both legs not being correctly placed in the stirrups, failure to attach the diathermy pad, incorrect handling of the instrument, violation of aseptic technique, and failure to confirm irrigation fluid correctness. Errors also arise from failure to confirm the proper setup of the camera, energy cables, and resectoscope, or when the foot pedal is not correctly positioned.

Critical errors include more serious failures that directly impact the safety and outcome of the procedure. For example, not completing the WHO checklist or the time-out procedure, failure to fix the patient on the stirrups, or mishandling anatomical abnormalities (such as the meatus) are critical errors. Additionally, critical errors include not performing essential tasks like inserting the instrument correctly into the bladder, failure to ensure the irrigation flow is functioning, and causing damage to the urethra or prostate/bladder neck. Other critical errors involve failure to identify lesions or characteristics of the bladder, not achieving complete haemostasis, and incomplete resection of tumours. These critical errors can jeopardise the success of the procedure and patient safety.

Metrics and errors are summarised in Supplement Tables S1 and S2.

The aim of this feasibility study was to present the preliminary experience of an Italian Academic Hospital using these performance metrics during surgical training of residents. Moreover, we tested the ability of metrics to distinguish an experienced surgeon from a resident in training.

Nine consecutive TURBT procedures that met the selection criteria were selected for this study. In detail, single, papillary lesions of <2 cm, were selected for the present study. Moreover, patients with active haematuria, UTIs, urethral stenosis, and prostates >50 mL were excluded.

Every procedure was performed under general or loco-regional anaesthesia with a 26-F bipolar resectoscope. This preliminary series involved two surgeons: an experienced surgeon with a fully accomplished learning curve on bladder resection, and a urologist in training in his first year of residency. The study included two steps:

- First step: both surgeons were unaware of the metrics and performed the procedure based on their usual patterns or routines.
- Second step: only the Resident was involved after studying the TURBT performance metrics.

Data about the steps reached and errors made were collected by a third experienced surgeon, present in the operating room during the surgical procedures; the two surgeons undergoing the test were unaware that they would be evaluated, thus preventing the so-called 'Hawthorne effect'; however, the third surgeon scorer was unblinded to who the operator was. A total score was calculated as the summary of the single steps reached. We analysed the differences between the total scores and errors made by the Expert Surgeon and the Resident during the first and the second step of the study.

Differences between group means were evaluated using a paired t-test to determine if there were statistically significant differences between the groups; differences were considered statistically significant at a $P < 0.05$. Two-tailed P values for each comparison were calculated. Statistical analyses were supported using the Python SciPy library for t-tests and significance calculations [6]. For the total score, the Resident achieved higher post-studying scores compared to pre-studying (mean [SD] 59.3 [1.5] vs 40.3 [7.8]). The Expert Surgeon also outperformed the Resident in the pre-studying phase (mean [SD] total score 61.3 [1.5]), while the difference between the Resident's post-studying performance and that of the Expert Surgeon was neither statistically nor clinically significant ($P = 0.698$).

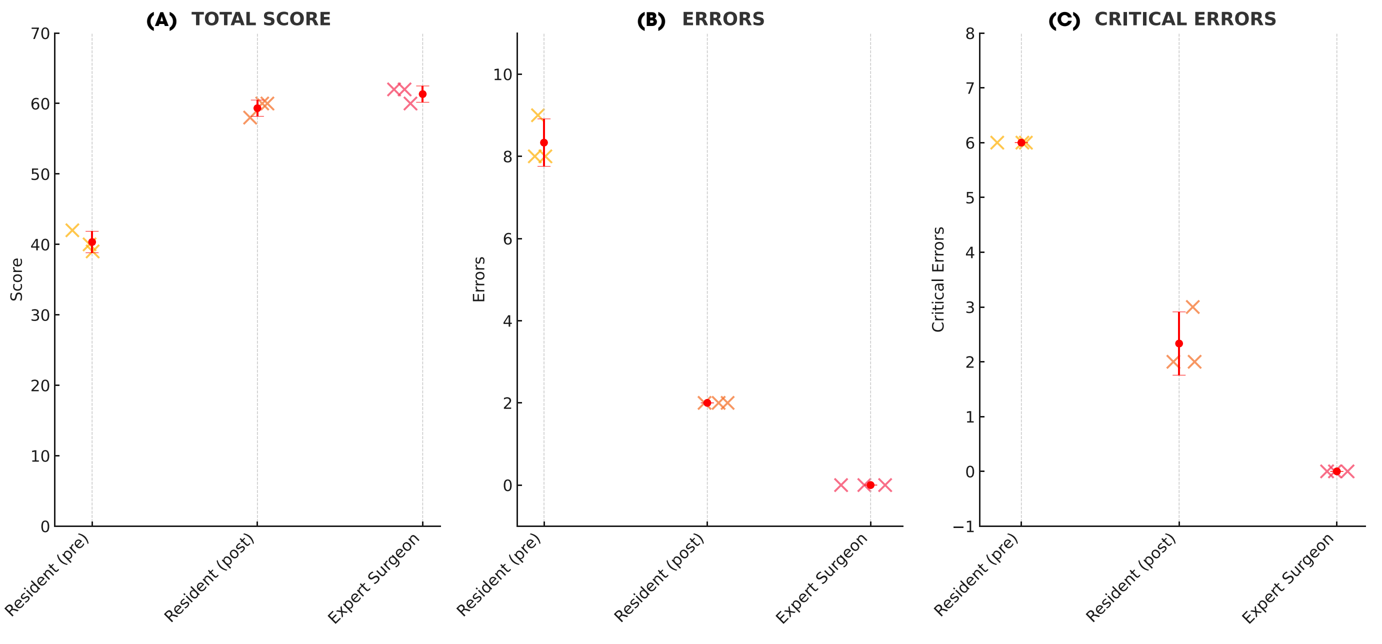
For errors, the Resident made fewer errors after training compared to before (mean [SD] number of errors 2.0 [2.0] vs 8.3 [1.5]), and the Expert Surgeon committed fewer errors than the Resident in the pre-studying phase (mean [SD] number of errors 0.0 [0.0]). No statistically or clinically significant difference was found between the Resident's post-studying performance and that of the Expert Surgeon.

Regarding critical errors, no statistically significant differences were observed among the three conditions (all $P > 0.05$). However, the reduction in critical errors post-studying (mean [SD] number of critical errors 2.3 [1.5] vs. 6.0 [1.0]) and the complete absence of critical errors in the Expert Surgeon, although not statistically significant, may be considered clinically relevant, given that a single critical error can compromise procedural success or patient safety (Fig. 1).

Our results showed how, after the study of the metrics, the Resident came closer to the results obtained by the Expert Surgeon, while maintaining a statistically significant difference in terms of total scores and clinically significant in terms of critical errors. Furthermore, we did not find a significant difference in terms of critical errors. This evidence underlines the permanence of critical errors after the study of the metrics but, recording an improvement approaching the results of the Expert Surgeon. This gap could hopefully be filled during the learning curve of the surgical procedure.

Our work has some biases, which could be eliminated as the project progresses. This is a preliminary and exploratory

Fig. 1 Individual scores and mean values for a single Resident (pre- and post-training) and one Expert Surgeon, across three outcome domains: Total Score (A), Errors (B), and Critical Errors (C). Each dot represents an individual score (three per phase); red points indicate mean (SD). Minimal horizontal jitter was applied to improve visualisation; no vertical jitter was applied to avoid misrepresentation of discrete outcomes. Statistical comparisons (Student's t-test) yielded the following P values: Total Score: Resident post- vs pre-training, $P = 0.024$; Resident pre-training vs Expert Surgeon, $P < 0.001$; Resident post-training vs Expert Surgeon, $P = 0.698$. Errors: Resident post- vs pre-training, $P = 0.004$; Resident pre-training vs Expert Surgeon, $P = 0.014$; Resident post-training vs Expert Surgeon, $P = 0.310$. Critical Errors: all comparisons non-significant ($P > 0.05$).



study, with a small number of treated patients, which could 'contaminate' the results. The evaluator surgeon was not 'blinded' to who was performing the surgical procedures and therefore may have been influenced in assigning scores and identifying errors. However, we could state that, in this preliminary experience, TURBT metrics seem to be a useful parameter to distinguish an experienced surgeon from a resident in training in terms of total scores; furthermore, the study of these metrics could be useful to significantly improve the intraoperative performance of the residents, in terms of decreasing the number of the errors and/or critical errors.

Disclosure of Interests

No conflicts exist.

Arturo Lo Giudice[1], **Angelo De Stefano**[1], **Gabriele Sorce**[1] **and Ettore Di Trapani**[2]

[1]Azienda Ospedaliera Universitaria G Martino di Messina, Messina, and [2]Department of Urology, European Institute of Oncology, IRCCS, Milan, Italy

References

1 **Babjuk M**, **Burger M**, **Capoun O et al.** European Association of Urology guidelines on non–muscle-invasive bladder cancer (Ta, T1, and carcinoma in situ). *Eur Urol* 2022; 81: 75–94

2 **Carrion DM**, **Rodriguez-Socarrás ME**, **Mantica G et al.** Current status of urology surgical training in Europe: an ESRU–ESU–ESUT collaborative study. *World J Urol* 2020; 38: 239–46

3 **Chahal B**, **Aydin A**, **Ahmed K**. Virtual reality vs. physical models in surgical skills training. An update of the evidence. *Curr Opin Urol* 2024; 34: 32–6

4 **De Groote R**, **Puliatti S**, **Amato M et al.** Discrimination, reliability, sensitivity, and specificity of robotic surgical proficiency assessment with global evaluative assessment of robotic skills and binary scoring metrics: results from a randomized controlled trial. *Ann Surg Open* 2023; 4: e307

5 **Paciotti M**, **Diana P**, **Gallioli A et al.** International consensus panel for transurethral resection of bladder tumours metrics: assessment of face and content validity. *BJU Int* 2024; 134: 932–8

6 **Virtanen P**, **Gommers R**, **Oliphant TE et al.** SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods* 2020; 17: 261–72

Correspondence: Arturo Lo Giudice, Urology Section, Gaetano Barresi Department of Human and Pediatric Pathology, University of Messina, Messina, Italy.
e-mail: arturologiudice@gmail.com

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Summary of different TURBT procedure metric phases and steps.

Table S2. List of unique TURBT procedure metric errors and critical errors.