Journal of Robotic Surgery (2022) 16:677–685
https://doi.org/10.1007/s11701-021-01293-6

ORIGINAL ARTICLE

Development and validation of the metric-based assessment of a robotic vessel dissection, vessel loop positioning, clip applying and bipolar coagulation task on an avian model

Stefano Puliatti[1,2,3] · Marco Amato[1,2,3] · Elio Mazzone[4,5] · Giuseppe Rosiello[4,5] · Ruben De Groote[1,2] · Pietro Piazza[1,2,6] · Luca Sarchi[1,2,3] · Rui Farinha[1,2,7,8] · Alexandre Mottrie[1,2] · Anthony G. Gallagher[1,9,10]

Received: 14 May 2021 / Accepted: 31 July 2021 / Published online: 12 August 2021

Abstract

The evolution of robotic technology and its diffusion does not seem to have been adequately accompanied by the development and implementation of surgeon training programs that ensure skilled and safe device use at the start of the learning curve. The objective of the study is to develop and validate performance metrics for vessel dissection, vessel loop positioning, clip applying and bipolar coagulation using an avian model. Three robotic surgeons and a behavioral scientist characterized the performance metrics of the task according to the proficiency-based progression methodology. Fourteen experienced robotic surgeons from different European countries participated in a modified online Delphi consensus. Eight experienced surgeons and eight novices performed the robotic task twice. In the Delphi meeting, 100% consensus was reached on the performance metrics. Novice surgeons took 26 min to complete the entire task on trial 1 and 20 min on trial 2. Experts took 10.1 min and 9.5 min. On average the Expert Group completed the task 137% faster than the Novice Group. The amount of time to reach the vessel part of the task was also calculated. Novice surgeons took 26 min on trial 1 and 20 min on trial 2. Experts took 5.5 min and 4.8 min. On average the experts reached the vessel 200% faster than the novices. The Expert Group made 155% fewer performance errors than the Novice Group. The mean IRR of video-recorded performance assessments for all metrics was 0.96 (95% confidence intervals (CI) lower = 0.94–upper = 0.98). We report the development and validation for a standard and replicable basic robotic vessel dissection, vessel loop positioning, clip applying and bipolar coagulation task on an avian model. The development of objective performance metrics, based on a transparent and fair methodology (i.e., PBP), is the first fundamental step toward quality assured training. This task developed on the avian model proved to have good results in the validation study.

Keywords Robotic surgical training · Proficiency-based metrics · Dissection skills · Training task validation

✉ Marco Amato
marcohz92@gmail.com

1 ORSI Academy, Melle, Belgium

2 Department of Urology, OLV, Aalst, Belgium

3 Department of Urology, University of Modena and Reggio Emilia, Modena, Italy

4 Division of Oncology/Unit of Urology, URI, IRCCS Ospedale San Raffaele, Milan, Italy

5 Vita-Salute San Raffaele University, Milan, Italy

6 Division of Urology, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy

7 Urology Department, NOVA Medical School, Lisbon, Portugal

8 Urology Department, Centro Hospitalar Universitário de Lisboa Central, Lisboa, Portugal

9 Faculty of Life and Health Sciences, Ulster University, Derry, Northern Ireland, UK

10 Faculty of Medicine, KU Leuven, Leuven, Belgium

Introduction

The da Vinci Robot (Intuitive Surgical Inc., Sunnyvale, CA) was the first robot to receive FDA approval for general laparoscopic surgery in 2000 [1]. Since then, the spread of robotic surgery has increased worldwide. Seven degrees of freedom, motion scaling, 3D visualization, precision, and instinctive movements are some of the robotic platform's advantages. Moreover, robotic surgery appears to offer better results in specific procedures (i.e., robotic-assisted radical prostatectomy) compared to the open and pure laparoscopic approach, although this is still debatable [2, 3]. They are however unsafe if the surgeon is not trained properly [4]. There are a number of new robotic-assist devices due to be introduced into the surgical market imminently. The training of surgeons to use these devices effectively, efficiently, and safely is imperative.

There is evidence in the literature of adverse events occurring in robotic surgery [5] which are directly related to the skill of the operating surgeon to use the robot and not the technology [6]. The evolution of robotic technology and its diffusion does not seem to have been adequately accompanied by the development and implementation of surgeon training programs that ensure skilled and safe device use at the start of the learning curve. The development of structured, efficient training curricula is still sub-optimal and there are still no objective parameters that trainees must reach before operating on real patients in the operating room [7–9]. To mitigate adverse events during robotic surgery, the scientific community currently agrees that it should be mandatory for the trainee, before performing even parts of an operation on the patient in the operating room [10]. There should be a requirement for trainees to demonstrate a high level of performance in the use of surgical technology (basic device training) and in the acquisition of basic surgical skills and these should be verified [11–13].

Simulation-based training is recognized as a fundamental and imperative part of training (for all levels of practitioners in training), across all surgical and medical disciplines [14, 15]. Research has shown that simulation works best when it is integrated into a proficiency-based progression (PBP) curriculum [16, 17]. The PBP approach to training is based on the development of detailed metrics that can serve as an objective performance feedback and as an understanding of what needs to be done and what needs to be avoided to carry out the procedure correctly and safely. Moreover, the same metrics are also used to establish performance benchmarks (i.e., a level of proficiency) which trainees should demonstrate before progressing in the training pathway [18–21]. A systematic review and meta-analysis reported that PBP in comparison to conventional training and other quality assured training pathways was associated with an improvement in terms of number of errors (i.e., 60% reduction), procedural time (i.e., 15% reduction) and steps completed (i.e., 47% increase) [22]. Based on these premises, our working group is developing and validating different PBP-based robotic surgical basic skills tasks. The goal is to develop tasks which are based on widely training materials, are inexpensive, preferably based on an animal tissue model. Furthermore, training tasks should have as an explicit learning outcome of the training of psychomotor performance units that are the fundamental building blocks of robotic surgical procedure performance. To date, we have identified and implemented metric-based training and performance benchmarks for suturing, knotting, dissection, and coagulation tasks using a chicken model.

In this study, we seek to expand this basic surgical skill training package and report on the development and validation of performance metrics for the training and assessment of vessel dissection, vessel loop positioning, clip applying and bipolar coagulation using the avian model. We hypothesize that the explicitly defined performance metrics which characterize optimal and sub-optimal performance metrics will distinguish between the performance of experienced and novice robotic surgeons. We also hypothesize that the performance metrics can be scored reliably by trained surgeon raters.

Materials and methods

This study involved two phases. The first phase is the identification of the task, development of the metrics and the subsequent metrics discussion, modification and approval during a Delphi consensus (Study 1). The second phase is the metrics validation study (Study 2).

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (Institutional Review Board, Onze Lieve Vrouw Hospital, Aalst, Belgium) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all participants for being included in the study. All institutional and national guidelines for the care and use of laboratory animals were followed.

Metrics development, task and Delphi consensus characteristics (Study 1)

Task development and characteristics of the model

Two consultant urologist robotic surgeons (SP, RDG) identified the task to be performed in the chicken leg. A blood vessel, with a diameter of about 2–5 mm, always located in the same position above and posterior to the tibiotarsus bone was identified. The defrosting process of the model, the robot positioning and setup have been reported elsewhere [12]. The 0 degrees camera, the monopolar scissors, the forceps bipolar, the Prograsp™ and the needle driver are the robotic instruments available to perform the task. (Fig. 1d).

The chicken is placed prone with the breast on the tray, with the monopolar plate attached to it. The right tibia (left for left-handers) is adducted to the right femur which is adducted to the body (Fig. 1a, b). A line of about 10 cm is drawn on the leg along its entire length. Two brackets drawn at the ends of the line delimit the external borders of the surgical field. Two supplementary brackets drawn in the center of the leg and 1 cm apart delimit the initial cutting line (Fig. 1c). Once the position and characteristics of the model had been defined, three surgeons (AM, RDG, SP) and a behavioral scientist (AGG) characterized the optimal and sub-optimal performance of the task according to the PBP methodology.

Seven steps, 18 general errors and 3 critical errors were identified and described in detail. The final metrics scoresheet is shown in Fig. 2. The exercise begins with the dissection of the skin along the drawn line. The trainee must navigate between the muscle bellies, delicately, until the vessel, located above and behind the bone, is identified. The fourth arm should be used to create space and facilitate dissection. The blood vessel is then delicately isolated for a length of 3 cm. After creating the space, the trainee can request the change of an instrument with the needle driver and must then place a vessel loop around the vessel without damaging it. Two clips are applied with the help of an assistant on the blood vessel. Finally, bipolar energy is applied medially to the clips and the vessel is cut. Different phases of the task execution are shown in Fig. 3a–d.

Delphi assessment

An online modified Delphi consensus was then organized to present and discuss the draft metrics, via the “Zoom” web platform. Fourteen participants from different countries participated in the Delphi meeting; their characteristics are reported in Table 1. The procedural steps, errors and critical errors were presented, discussed and edited live during the

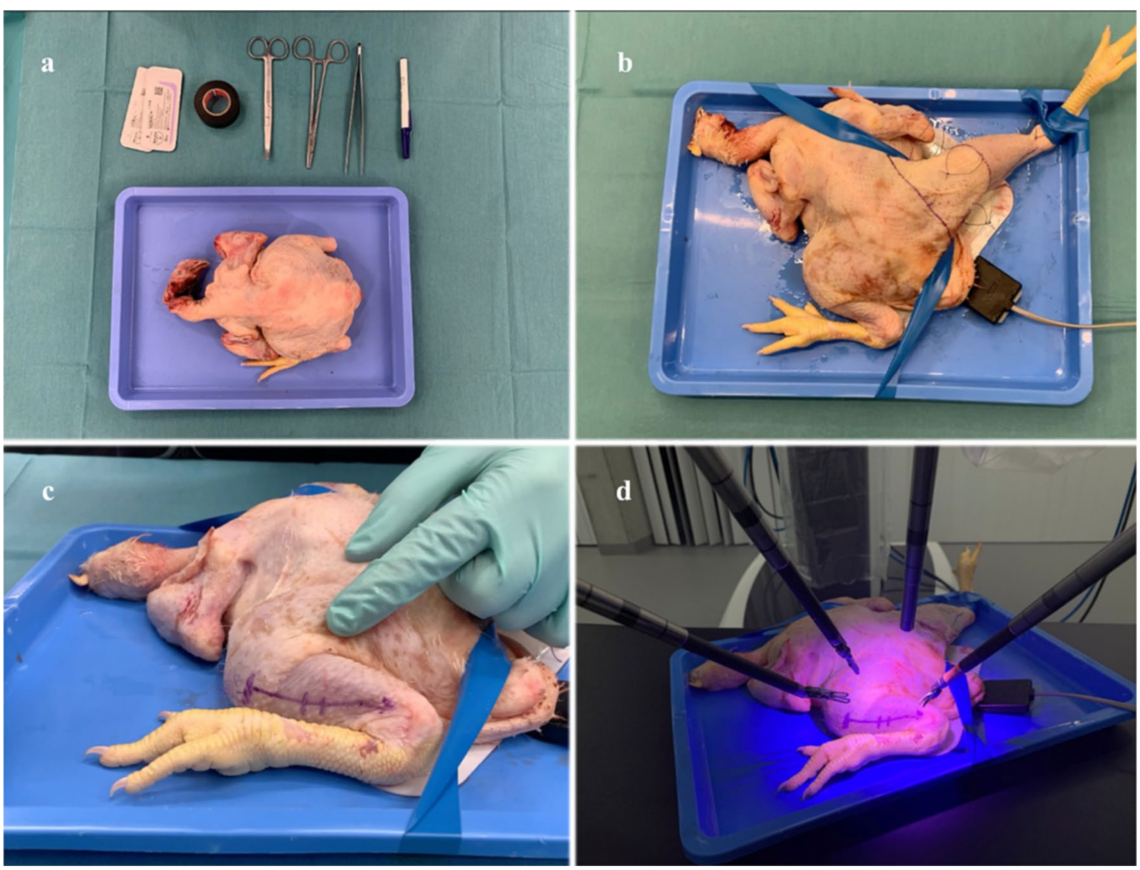

Fig. 1 Chicken model setup. **a** Surgical instruments needed for the preparation of the chicken. **b** Fixation of the chicken. **c** Drawn lines and model ready for the beginning of the exercise. **d** Chicken positioned and robot docked to start the exercise

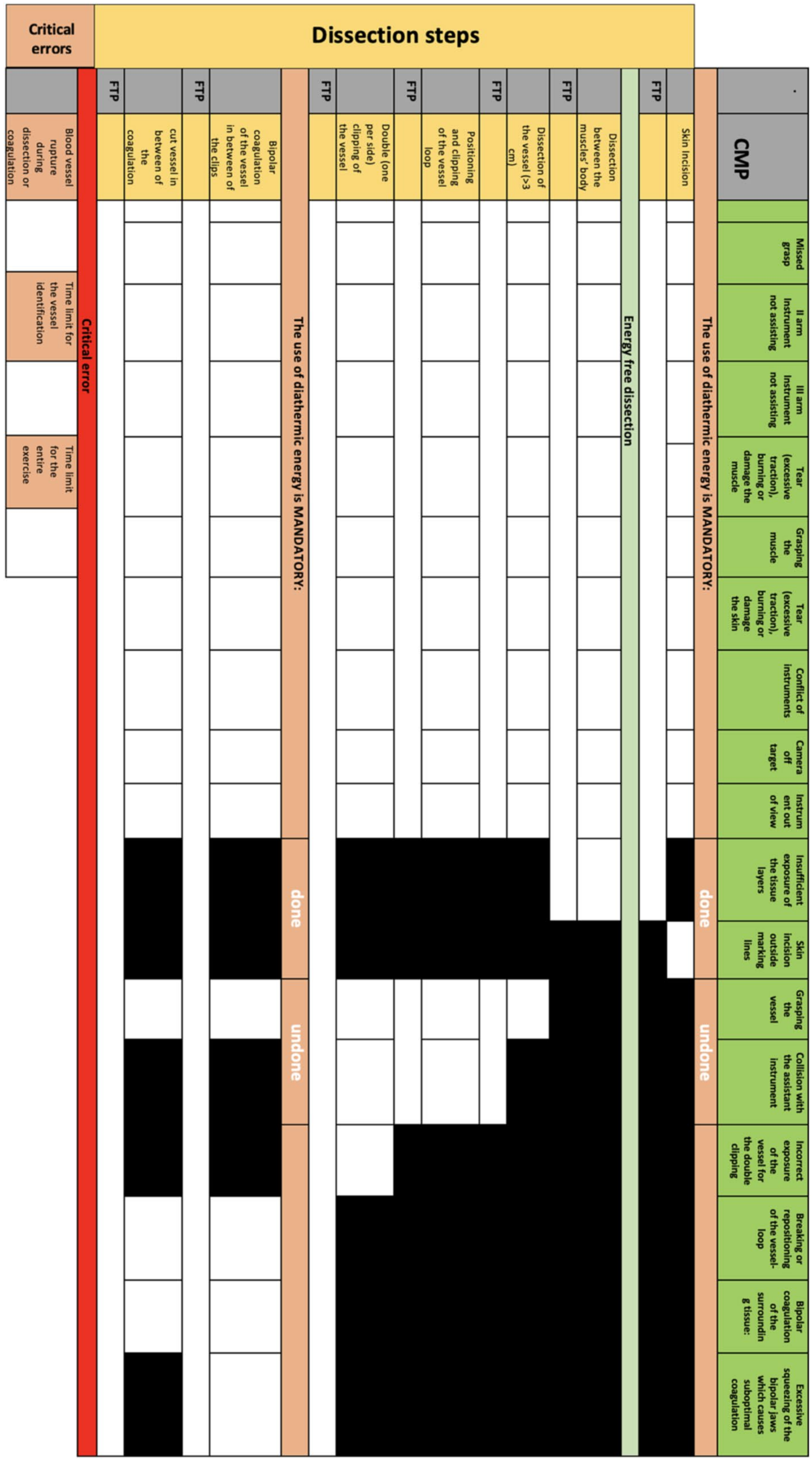

Fig. 2 Metrics final scoresheet

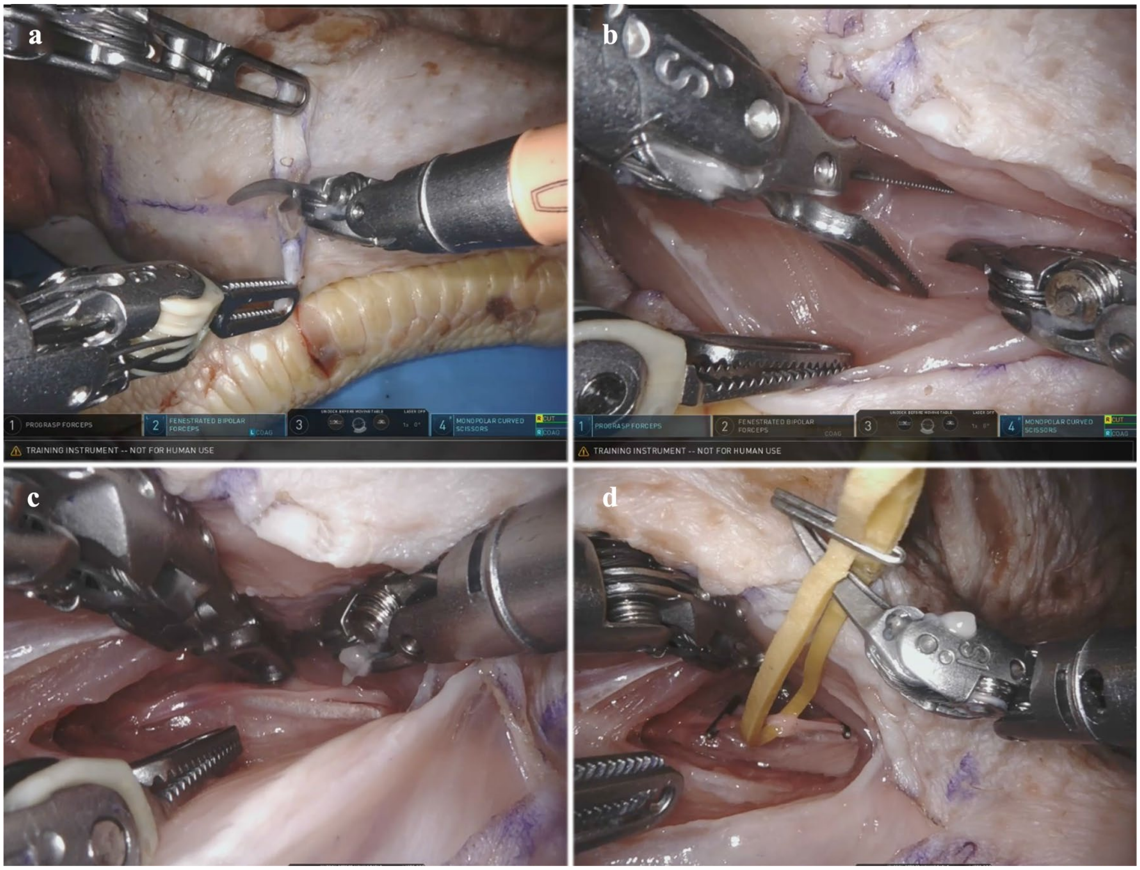

Fig. 3 Different phases of the chicken dissection task execution. **a** Skin incision and development of the skin incision. **b** Blunt dissection between the muscle bellies. **c** Identification of the millimetric blood vessel. **d** Vessel loop positioned around the blood vessel

Table 1 Delphi panel participants' characteristics

Experts Delphi panel		
Country	Years of experience	Expertise
Belgium	12	RALP
Belgium	4	RALP
Belgium	30	RALP, RAPN
Belgium	20	RALP, RAPN
Germany	10	Upper GI
Italy	11	RALP
Italy	6	RALP
Italy	9	RALP
Italy	5	RALP
Poland	13	RALP
Portugal	15	RAPN
United Kingdom	15	RA colorectal
USA	30	Upper GI
USA	30	Arthroscopy

meeting. The final edited draft of the metrics was then voted on and the consensus level was established.

Validation study (Study 2)

Eight experienced surgeons and eight novices participated in the construct validity study. The eight experienced surgeons were all Belgians, five urologists, two gynecologists and one general surgeon. They all performed more than 300 robotic procedures. The novices were all Belgian residents in urology, with minimum or no experience in robotic surgery. The characteristics of the participants are reported in Table 2. All the participants completed two consecutive trials (16 trials × 2 in total). The full videos of the trials performed were recorded and collected. Two robotic surgeons (MA, SP) objectively assessed, accordingly to the final agreed metrics scoresheet, the videos of the subjects involved in the study. Before having access to the videos, the two reviewers were trained to score videos with an inter-rater reliability (IRR) above 0.8. The reviewer detailed training methodology is described

Table 2 Baseline characteristics of participants in the study

Characteristics				
Attribute	Experts	Mean %	Novices	Mean %
Total Number	8	–	8	–
Nationality				
Belgian	8	100	8	100
First language				
Flemish	8	100	8	100
Gender				
Male	8	100	7	87.5
Female	–		1	12.5
Handedness				
Right	7	87.5	7	87.5
Left	1	12.5	1	12.5
Specialty				
Gynecology	2	25	–	
Urology	5	62.5	8	100
General surgery	1	12.5	–	
PG year				
1	–	–	5	62.5
2	–	–	3	37.5

elsewhere [23]. The reviewers underwent a training process during which they analyzed multiple video examples of the task under the supervision of a third part (AG). The training last until an IRR (agreements/agreements + disagreements) greater than 0.8 was reached. Then, the two raters started the participants' video assessment. All videos were fully evaluated by each individual reviewer blindly. The reviewers remained blinded to the identity of the operator, their group (i.e., experienced or novice surgeon) and trial number (i.e., trial 1 o 2).

Statistical analysis

Statistical analysis was performed with SPSS 26 (Armonk, New York). A 2 × 2 mixed model analysis of variance (ANOVA) was used to determine if there was a statistical difference for primary end points (time to complete the task and the total number of errors made) between the Expert and the Novice groups Trial 1 and Trial 2 (i.e., repeated measures). Strength of association between variables was assessed with Pearson product–moment correlation.

Results

Baseline characteristics with respect to age, gender, handedness, sight-corrected status, surgical discipline, nationality, first language, etc., of the participants in each group are shown in Table 2.

The mean IRR of video-recorded performance assessments for all metrics was 0.96 (95% Confidence intervals (CI) lower = 0.94–upper = 0.98). None of the video-recorded assessments were below the 0.8 IRR level. Both groups completed all the procedure steps on both assessment trials.

Multivariate statistics

The overall mean and 95% CI for the mean number of minutes taken to complete the task by the Expert and Novice groups are shown in Fig. 4a. Also shown are the mean scores for both groups on Trial 1 and Trial 2. Both groups showed improvement from Trial 1 to Trial 2. The performance improvement for the Expert Group was marginal (10.1 to 9.5 min). The improvement was more pronounced for the Novice Group (i.e., 23%, 26 to 20 min), but they did demonstrate considerable performance variability as indicated

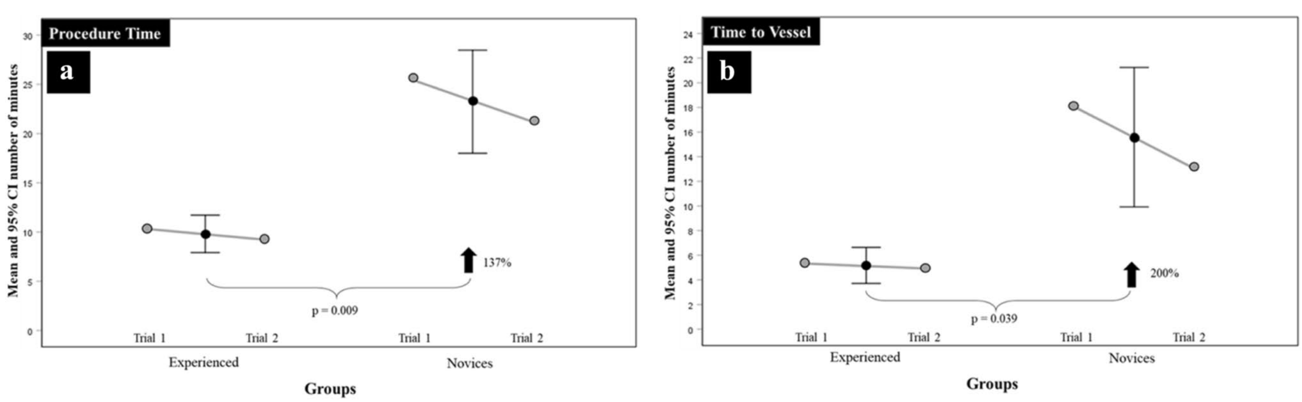

Fig. 4 The mean and 95% CI of **a** the time taken to complete the procedure and **b** the time take to get to the vessel dissection portion of the task by experienced and novice surgeons on the robot-assisted task as well as performances on Trials 1 and 2

by the larger confidence intervals. Neither of these differences were statistically significant. In terms of overall time to perform the task, the Expert Group on average completed the task 137% faster than the Novice Group and this difference was statistically significant (95% CI of the difference, lower = − 19.19–upper = − 3.24, df = 14, t = − 3.02, p = 0.009).

The same analysis was completed for the amount of time it took subjects from the start of the task to reach the vessel part of the task. Figure 4b shows the mean 95% CI of the performance of the two groups as well as the amount of time they took on trials 1 and 2. As with procedure time, the improvement in performance time was marginal (5.5–4.8 min). The improvement from Trial 1 to Trial 2 was more pronounced for the Novice Group (18.2 to 12.9 min–29%). This difference was not statistically significant. On average, the Expert Group reached the vessel portion of the task 200% faster than the Novice Group. Despite the large variability in the times observed for the Novice Group, this difference was found to be statistically significant (95% CI of the difference, Lower = − 15.73–Upper = − 0.46, df = 14, t = − 2.28, p = 0.039).

Figure 5 shows the mean and 95% CI number of errors that both groups made as well as the mean number of errors they made during Trials 1 and 2. The Experienced Group, on average made 155% fewer errors than the Novice Group and showed lower performance variability as indicated by standard deviation (SD) scores (Experienced SD = 2.6 and Novice SD = 3.8). The difference between the two groups was statistically significant (95% CI of the difference, lower = − 9.97–upper = − 2.91, df = 14, t = 3.91, p < 0.002). A significant change in performance from Trial 1 to Trial 2 was observed (95% CI of the difference, lower = 0.23–upper = 6.37, df = 26.84, t = 2.2, p < 0.036) and the change depended on group membership. In the Experienced Group they made 48% more errors in Trial 2 (T1 = 4.8 errors and in T2 = 7.1 errors). The reverse was true for the Novice Group, as they made 20% fewer errors in Trial 2 than they did in Trial 1 2 (i.e., T1 = 16.8 errors and in T2 = 13.5 errors). This interaction effect was statistically significant (95% CI of the difference, lower = − 9.9–upper = − 1.2, df = 26.84, t = 2.62, p < 0.014).

The error scores for the Novice Group were moderately and statistically significantly positively correlated with the time it took for them to perform the task (i.e., r = 0.619, p = 0.014). This meant that the longer they took to perform the task, the more likely they were to have a higher error score. The same was not observed for the Experienced Group (r = 0.291, p = 0.275).

Discussion

In this study, we found that the objectively assessed performance metrics which we developed to characterize optimal/sub-optimal dissection, clipping and coagulation skills distinguished between experienced robotic surgeons and novice's completion of the task. The experienced surgeons completed the task significantly faster and made fewer objectively assessed performance errors. Furthermore, the experienced surgeons demonstrated considerably better performance homogeneity in comparison to the novice group. The performance metrics also demonstrated good inter-rater reliability and all performances were scored with an inter-rater reliability > 0.8.

Performance time and number of errors made are fundamental performance units that characterize important, related, but different aspects of surgical performance. The

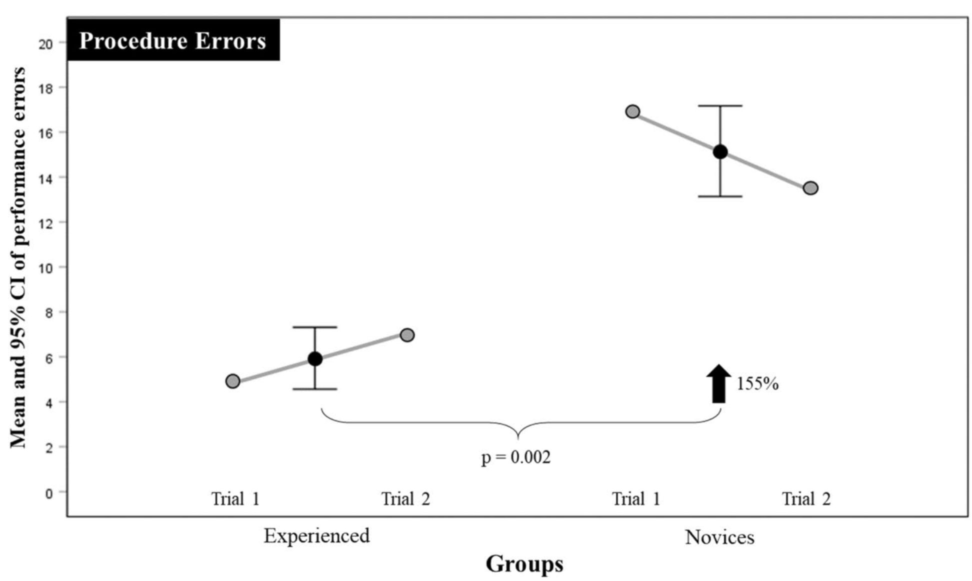

Fig. 5 The mean and 95% CI of the number of procedure errors made by the experienced and novice surgeons on the robot-assisted task as well as the number of errors made during Trials 1 and 2

amount of time it takes to perform a task is an important performance yardstick which historically has been used as a metric to measure skill. As in this study, the amount of time taken to perform the task reliably discriminated between the two groups. This contrasts with the Steps of the task. Both groups completed all the steps required to complete the task. Therefore, the procedure Steps did not discriminate between the two groups. Procedure Steps and time taken to perform the task are however both measures of ‘process’, i.e., a series of actions or steps taken to achieve a particular end. Process measures are imperative to completion of the task; however, they give no indication of the quality of task performance.

Take for example procedure steps; based on this measure, it might be concluded that there was no performance difference between the groups when in fact there was. Time to complete the task reliably discriminated between the two groups, but this measure gave no indication of how well the tasks was performed. In contrast, the Error metrics reliably discriminated between the two groups performances and give a very accurate and detailed portrayal of how well the task was completed. A task may be completed quickly but this might be because certain steps were omitted or not performed at all. Ideally, performance measurements should consist of both process and quality measures.

Over the last two decades, performance error has emerged as an important measure of surgical performance [20, 21]. Process measures such as procedure steps will always form the foundation of skilled performance. Error metrics, however, will quality assure the performances. In the skills laboratory, both performance metrics form the basis effective and efficient training.

The trainee is educated (usually in an e-learning package) how to perform the procedure, with which instruments and the order in which they should be used. The trainee is then required to integrate this knowledge with the technical completion of the past or procedure in the skills laboratory. This means that the trainee has the opportunity to hone their skills in a safe environment with no risk to the patient. In a proficiency-based progression training paradigm, the performance metrics (similar to those reported on here) are used the help the trainee hone their skills in a deliberate practice [24] training regime. They are given specific and detailed formative feedback on their performances and shown how to improve. This contrasts with the traditional approach to training of repeated practice where the trainees practice again and again with the hope that they eventually will get it right. In this regard, prospective randomized studies comparing PBP training versus traditional training on robotic tasks are underway.

In this context, the performance metrics, derived from and benchmark on very experienced and proficient surgeons forms the core of the training curriculum [18, 20, 21]. The simulation model, e.g., the chicken leg used in this study, is then simply a tool for the delivery of metric-based training curriculum [25]. Obviously, the training model must be appropriate and afford the trainee the opportunity to perform the technical aspects of task completion that emulate performances they will be required to complete on real patients. Virtual reality simulations will always trump animal tissue models because of their capacity to automatically reliably measure the performance [26]. These do not, however, currently exist with sufficient fidelity to train robotic surgical skills. The main advantage of the model we have reported here is that it is widely available and inexpensive. The disadvantage is that there is no automated assessment of performance. Proponents of artificial intelligence (AI) point out that automated performance assessment is easily doable, it is not currently widely available and where it is available only assesses process measures of performance, i.e., instrument tracking [27]. In the short term, skilled performance can be acquired using relatively simple animal tissue models that have been metricized. This will however require the training of faculty to use the performance metrics for training and assessment.

Conclusions

The acquisition of essential basic surgical skills is one of the first steps to be taken within a training curriculum. In this study, we devised an exercise for the acquisition of dissection, clipping and coagulation skills. We developed metrics that have been discussed and agreed by senior international minimally invasive surgeons within an online modified Delphi consensus meeting. In a second study, we demonstrated that the metrics reliably discriminated between the objectively assessed performance of experienced and novice robotic surgeons. The goal is to use this, together with other exercises, to ensure the acquisition of basic skills such as suturing, knotting, dissection, and coagulation. The achievement of an adequate level of proficiency in these basic exercises will be mandatory for access to the advanced training phases.

Declarations

Conflict of interest Stefano Puliatti, Marco Amato, Elio Mazzone, Giuseppe Rosiello, Ruben De Groote, Pietro Piazza, Luca Sarchi, Rui Farinha, Alexandre Mottrie and Anthony G. Gallagher declare that they have no conflicts of interest or financial ties to disclose.

References

1. Yates DR, Vaessen C, Roupret M (2011) From Leonardo to da Vinci: the history of robot-assisted surgery in urology. BJU Int 108(11):1708–1713 (**discussion 1714**)
2. Puliatti S, Elsherbiny A, Eissa A et al (2019) Effect of pubo-prostatic ligament reconstruction on continence recovery after

robot-assisted laparoscopic prostatectomy: our initial experience. Minerva Urol Nefrol 71(3):230–239
3. Mazzone E, Dell'Oglio P, Rosiello G et al (2020) Technical refinements in superextended robot-assisted radical prostatectomy for locally advanced prostate cancer patients at multiparametric magnetic resonance imaging. Eur Urol 80(1):104–112
4. Nik-Ahd F, Souders CP, Zhao H et al (2019) Robotic urologic surgery: trends in litigation over the last decade. J Robot Surg 13(6):729–734
5. Alemzadeh H, Raman J, Leveson N et al (2016) Adverse events in robotic surgery: a retrospective study of 14 years of FDA data. PLoS One 11(4):e0151470
6. Pradarelli JC, Thornton JP, Dimick JB (2017) Who is responsible for the safe introduction of new surgical technology?: An important legal precedent from the da Vinci Surgical System Trials. JAMA Surg 152(8):717–718
7. Beulens AJW, Hashish YAF, Brinkman WM et al (2020) Training novice robot surgeons: proctoring provides same results as simulator-generated guidance. J Robot Surg 21:S26. https://doi.org/10.1016/S2666-1683(20)35868-7
8. Beulens AJW, Vaartjes L, Tilli S et al (2020) Structured robot-assisted surgery training curriculum for residents in urology and impact on future surgical activity. J Robot Surg 21:S24. https://doi.org/10.1016/S2666-1683(20)35866-3
9. Puliatti S, Mazzone E, Dell'Oglio P (2020) Training in robot-assisted surgery. Curr Opin Urol 30(1):65–72
10. Vanlander AE, Mazzone E, Collins JW et al (2020) Orsi consensus meeting on european robotic training (OCERT): results from the first multispecialty consensus meeting on training in robot-assisted surgery. Eur Urol 78(5):713–716. https://doi.org/10.1016/j.eururo.2020.02.003
11. Pierorazio PM, Allaf ME (2009) Minimally invasive surgical training: challenges and solutions. Urol Oncol 27(2):208–213
12. Puliatti S, Mazzone E, Amato M et al (2020) Development and validation of the objective assessment of robotic suturing and knot tying skills for chicken anastomotic model. Surg Endosc 35(8):4285–4294. https://doi.org/10.1007/s00464-020-07918-5
13. Somani B, Brouwers T, Veneziano D et al (2021) Standardization in Surgical Education (SISE): development and implementation of an innovative training program for urologic surgery residents and trainers by the European school of urology in collaboration with the ESUT and EULIS Sections of the EAU. Eur Urol 79(3):433–434
14. Healy GB (2002) The college should be instrumental in adapting simulators to education. Bull Am Coll Surg 87(11):10–11
15. Gallagher AG, Cates CU (2004) Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. JAMA 292(24):3024–3026
16. Seymour NE, Gallagher AG, Roman SA et al (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 236(4):458–463 (**discussion 463-4**)
17. Ahlberg G, Enochsson L, Gallagher AG et al (2007) Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 193(6):797–804
18. Gallagher A (2012) Metric-based simulation training to proficiency in medical education: what it is and how to do it. Ulster Med J 81(3):107–113
19. Cates CU, Gallagher AG (2012) The future of simulation technologies for complex cardiovascular procedures. Eur Heart J 33(17):2127–2134
20. Gallagher AG, O'Sullivan GC (2011) Fundamentals of surgical simulation principles and practices. Springer Verlag, London
21. Gallagher AG, Ritter EM, Champion H et al (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg 241(2):364–372
22. Mazzone E, Puliatti S, Amato M et al (2020) A systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes. Ann Surg. https://doi.org/10.1097/SLA.0000000000004650
23. Gallagher AG, Ryu RKN, Pedowitz RA et al (2018) Inter-rater reliability for metrics scored in a binary fashion-performance assessment for an arthroscopic bankart repair. Arthroscopy 34(7):2191–2198
24. Ericsson KA, Krampe RT, Tesch-Römer C (1993) The role of deliberate practice in the acquisition of expert performance. Psychol Rev 100(3):363–406
25. Gallagher AG, Henn P (2014) Simulation fidelity: more than experience and mere repetition? Stud Health Technol Inform 196:128–134
26. Satava R (2007) The future of surgical simulation and surgical robotics. Bull Am Coll Surg 92(3):13–19
27. O'Sullivan S, Leonard S, Holzinger A et al (2020) Operational framework and training standard requirements for AI-empowered robotic surgery. Int J Med Robot 30:e2020