




Editorial Perspective: Robot-Assisted Evaluation of Robotic Surgical Skills

Shih-Chun Cheng, MD¹, and Yin-Kai Chao, MD² 

¹Taiwan AI Labs, Taipei, Taiwan; ²Division of Thoracic Surgery, Chang Gung Memorial Hospital-Linkou, Chang Gung University, Taoyuan, Taiwan

Robot-assisted surgery is an advanced minimally invasive technique that provides more precision, control, and flexibility compared with conventional approaches.^{1–3} While robotic technology allows surgeons to perform complex interventions in limited anatomical spaces, the operators' technical skills remain a key determinant of successful clinical outcomes. In this scenario, there is growing interest in the assessment of proficiency among trainees practicing robotic surgery. This would allow estimating the position of an individual on a learning curve.⁴

The da Vinci Skills Simulator (dVSS) has emerged as an interesting platform for objective evaluation of the abilities required during robot-assisted surgery.⁵ On the simulator, proficiency can be assessed through various virtual exercises (e.g., “ring and rail”) by means of built-in assessment criteria. While simulation-based training allows trainees to practice a procedure in a safe and controlled environment, it does not invariably reflect real surgical situations. One potential solution to address this issue is the use of the Global Evaluative Assessment of Robotic Skills (GEARS), a standardized and validated qualitative assessment tool for robotic surgical skills.⁶ However, GEARS is a Likert-scale measure susceptible to response biases. Another traditional possibility is to investigate the improvement in surgical performance over time, described as the learning curve. A learning curve can be defined as the number of cases

required and/or the time taken by a surgeon to become proficient in key indicators (e.g., operating time or occurrence of certain index complications).⁷

In this issue of the *Annals of Surgical Oncology*, Takeuchi et al.⁸ describe a novel automated surgical step recognition system for robot-assisted minimally invasive esophagectomy (RAMIE). The tool was developed by applying deep learning algorithms to video analysis of standardized procedures. Specifically, the system was designed to quantitatively analyze the relationships between the duration of each step and the surgeon's learning curve. While there have been several previous attempts to automatically identify surgical phases through artificial intelligence (AI), the application of this technique in the assessment of surgical proficiency is certainly innovative. By taking each surgical step into account in an automated manner, this tool holds great promise for robot-assisted evaluation of robotic surgery. In addition, longitudinal investigation of surgical indicators using AI tools may provide valuable information for improving surgical training.

However, certain hurdles that will likely impede the rapid and effective translation of the automated system developed by Takeuchi et al. should be briefly discussed. First and foremost, the vast majority of previous learning studies have relied on operative videos obtained from either a single or a limited number of surgeons.^{9–13} In this regard, the work by Takeuchi and coworkers is no exception. Specifically, all of the RAMIE video recordings derived from a single surgeon without previous experience in this procedure but with an extensive record in the field of minimally invasive esophagectomy (> 300 procedures). The fact that the surgical steps utilized in the study were not standardized and the lack of a consensus definition for surgical proficiency are other significant caveats. For example, the definition for proficiency used by Takeuchi et al. (i.e., a minimum of 20 procedures) was based on a

limited number of previous studies, which limits the robustness of the analysis. These shortcomings call for international working groups to make recommendations vetted by the participating organizations. Second, marked discrepancies in the number of frames across different surgical steps may introduce bias in terms of recognition accuracy. For example, the number of frames was as low as 104 in the video depicting azygos vein division but as high as 4230 in the video showing left recurrent laryngeal nerve (RLN) lymph node dissection (LND). The F1 scores for these two videos were 76% and 89%, respectively, further suggesting the existence of imbalanced sampling. Data augmentation techniques, including image flipping, zooming, shifting, and generative adversarial networks (GANS) synthesis,¹⁴ should also be implemented cautiously. Accordingly, the resulting data are not always reflective of real anatomy. Finally, object delineation and motion tracking may be incorporated into the step recognition task for confirmation purposes. While step recognition algorithms are entirely based on automated feature extraction from whole pictures or selected image portions, the extracted features generally pose interpretation problems. Only the incorporation of more contextualized tasks (e.g., object delineation¹⁵ and motion tracking¹⁶) will facilitate proper assessment of the intrinsic discriminative ability of an algorithmic system, with the ultimate goals of improving its accuracy and addressing potential pitfalls. By comparing the instrument-handling abilities of trainees versus those of experienced surgeons, information from object delineation and motion tracking will also be paramount in the field of surgical training. Finally, establishing ground-truth labels, viz. data that accurately represent real-world situations, will be a crucial milestone for surgical training, although this will surely be more demanding than the relatively simple recognition step. That might sound ambitious, but transformative advances are still expected in the field of AI applied to surgical learning. There is much we can learn from robots, very much like they are currently learning from our own experiences.

DISCLOSURE The authors declare no conflict of interest in relation to this manuscript.

REFERENCES

1. Tschann P, Szeverinski P, Weigl MP, et al. Short- and long-term outcome of laparoscopic- versus robotic-assisted right colectomy: a systematic review and meta-analysis. *J Clin Med*. 2022;11(9):2387. <https://doi.org/10.3390/jcm11092387>.
2. Arkoncel FR, Lee JW, Rha KH, Han WK, Jeoung HB, Oh CK. Two-port robot-assisted versus standard robot-assisted laparoscopic partial nephrectomy: a matched-pair comparison. *Urology*. 2011;78(3):581–5. <https://doi.org/10.1016/j.urology.2010.10.046>.

3. Cundy TP, Harling L, Hughes-Hallett A, Mayer EK, Najmaldin AS, Athanasiou T, et al. Meta-analysis of robot-assisted versus conventional laparoscopic and open pyeloplasty in children. *BJU Int*. 2014;114:582–94.
4. Khan N, Abboudi H, Khan MS, Dasgupta P, Ahmed K. Measuring the surgical ‘learning curve’: methods, variables and competency. *BJU Int*. 2014;113:504–8.
5. Meier M, Horton K, John H. da Vinci© Skills Simulator™: is an early selection of talented console surgeons possible? *J Robot Surg*. 2016;10(4):289–96. <https://doi.org/10.1007/s11701-016-0616-6>.
6. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol*. 2012;187(1):247–52. <https://doi.org/10.1016/j.juro.2011.09.032>.
7. Abboudi H, Khan MS, Guru KA, Froghi S, de Win G, Van Poppel H, Dasgupta P, Ahmed K. Learning curves for urological procedures: a systematic review. *BJU Int*. 2014;114:617–29.
8. Takeuchi M, Kawakubo H, Saito K, et al. Automated surgical-phase recognition for robot-assisted minimally invasive esophagectomy using artificial intelligence. *Ann Surg Oncol*. 2022. <https://doi.org/10.1245/s10434-022-11996-1>.
9. Chang Y, Qu M, Wang L, Yang B, Chen R, Zhu F, et al. Robotic-assisted laparoscopic radical prostatectomy from a single Chinese center: a learning curve analysis. *Urology*. 2016;93:104–11.
10. Good DW, Stewart GD, Laird A, Stolzenburg JU, Cahill D, McNeill SA. A critical analysis of the learning curve and postlearning curve outcomes of two experience- and volume-matched surgeons for laparoscopic and robot-assisted radical prostatectomy. *J Endourol*. 2015;29:939–47.
11. Kim IK, Kang J, Park YA, Kim NK, Sohn SK, Lee KY. Is prior laparoscopy experience required for adaptation to robotic rectal surgery? Feasibility of one-step transition from open to robotic surgery. *Int J Colorectal Dis*. 2014;29:693–9.
12. Lebeau T, Rouprêt M, Ferhi K, Chartier-Kastler E, Bitker MO, Richard F, et al. The role of a well-trained team on the early learning curve of robot-assisted laparoscopic procedures: the example of radical prostatectomy. *Int J Med Robot*. 2012;8:67–72.
13. Sood A, Ghani KR, Ahlawat R, Modi P, Abaza R, Jeong W, et al. Application of the statistical process control method for prospective patient safety monitoring during the learning phase: robotic kidney transplantation with regional hypothermia (IDEAL phase 2a–b). *Eur Urol*. 2014;66:371–8.
14. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*. 2018;321:321–31. <https://doi.org/10.1016/j.neucom.2018.09.013>.
15. D. Jha et al. Exploring deep learning methods for real-time surgical instrument segmentation in laparoscopy, 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), 2021, pp. 1–4, doi: <https://doi.org/10.1109/BHI50953.2021.9508610>
16. Franco-González IT, Pérez-Escamirosa F, Minor-Martínez A, et al. Development of a 3d motion tracking system for the analysis of skills in microsurgery. *J Med Syst*. 2021;45:106. <https://doi.org/10.1007/s10916-021-01787-8>.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.