A multilevel, step-based model to evaluate progress in procedure efficiency for laparoscopic appendicectomy in surgical training: structured evaluation using ‘ebb-and-flow’ and ‘string-of-pearls’ concepts

Kjetil Søreide1,2,* and Benedicte Skjold-Ødegaard1,3,4

1Department of Gastrointestinal Surgery, Stavanger University Hospital, Stavanger, Norway
2Department of Clinical Medicine, University of Bergen, Bergen, Norway
3Faculty of Health Sciences, University of Stavanger, Stavanger, Norway
4Department of Surgery, Haugesund Hospital, Haugesund, Norway

*Correspondence to: Kjetil Søreide, Department of Gastrointestinal Surgery, P.O. Box 8100, N-4068 Stavanger, Stavanger University Hospital, Stavanger, Norway (e-mail: ksoreide@mac.com)

Abstract

Background: Surgical training is aimed towards entrusted professional activity to obtain operative independence. Laparoscopic appendicectomy is performed early in training but except for simulators, real-life evaluation towards proficiency is scarce. The aim of this study was to model how each consecutive step may impact on the overall proficiency score for surgical trainees performing laparoscopic appendicectomy.

Methods: This was an observational cohort study of laparoscopic appendicectomy performed by junior trainees (PGY1–4) under supervision and evaluated for each of eight steps. Each step was scored on a validated six-point performance scale and classified as ‘fail’, ‘pass’, or ‘proficient’. Modelling was conducted with a multivariable regression model and artificial neural network model with a multilayer perceptron for the relationship between steps and overall performance.

Results: Of 157 procedures, 97 (61.8 per cent) procedures were evaluated as ‘proficient’, 46 (29.3 per cent) were ‘pass’, and 14 (8.9 per cent) were ‘fail’. In regression analyses, handling the mesoappendix was significantly associated with procedure proficiency, as were division of appendix, access to abdomen, and ability to handle the small bowel. The widest variation in operative efficiency for laparoscopic appendicectomy

Conclusions: Consistent stumbling blocks for junior trainees performing laparoscopic appendicectomies can be conceptualized through novel models that identify steps deemed to be the most difficult to trainees with variable experience.

Introduction

Concerns about surgical trainees’ ability to obtain independence during surgical training has become an increasingly debated topic. Trainees need to practice independently as they progress through training, with increasing responsibility as they master a wider procedure spectrum in general surgery. Rather than the century-old adage of ‘see one, do one, teach one’ the current competency-based practice includes so-called ‘entrusted professional activity’3, whereby residents are allowed to take on tasks that they are entrusted to master.

Acute appendicitis is one of the most common surgical conditions worldwide2. Consequently, appendicectomy is among the most frequently performed procedures in general surgery. In healthcare systems in the western world, most appendectomies are now performed minimally invasively3,4; hence, laparoscopic appendicectomy is often encountered early on in surgical training and is a procedure for which residents are expected to have a very high degree of autonomy at the end of training5.

Several studies have demonstrated the safety of residents performing appendicectomy independently6–8; however, less is known about which objective criteria to use to certify a resident’s capability of performing a procedure proficiently and, eventually, independently. Also, little data exist on what parts of procedures are considered stumbling blocks to performance and which are steppingstones to mastery. In a training programme for trainees and trainers in general surgery evaluating real-life surgery, we have previously demonstrated that there are no differences in trainee performance according to sex9. Previous study data have also demonstrated that the cut-off for mastering any given step of a laparoscopic appendicectomy is widely different10, suggesting that a fixed procedure number is insufficient to demonstrate ‘experience’. Notably, a minimum number of appendectomies to complete training in general surgery is often proposed by endorsing...
bodies. Data on numbers needed to achieve proficiency are scarce, particularly for real-life procedures. Learning curves for simulator models have been reported11–13, and data for single surgeons have been presented14,15; however, the focus on a single number neglects the progress and evaluation during training, with possible need for focused training tasks to improve performance. Indeed, we found a highly variable degree of proficiency obtained when evaluating the various procedure steps of laparoscopic appendicectomy16.

The aim of this study was to investigate the procedure flow and its relation to the mastery of a procedure using well described techniques from other industries, yet hitherto not explored for evaluation during surgical training to the best of our knowledge. We believe the specific findings of this study will have potential wide generic applications across training of procedures, as it allows us to define stumbling blocks that hamper flow and proficiency of a procedure’s overall performance.

Methods
Study design and ethical approval
This was a cross-sectional, observational cohort study of junior trainees (1–4 years of surgical experience) performing real-life laparoscopic appendicectomy in a structured training programme from 1 January to 31 December 2018. The structure, theory, and implementation of the training and the train-the-trainer programme is described and presented in detail elsewhere16. As an observational study, the STROBE guidelines17 were adhered to, as applicable.

Ethical approval
The study was approved by the Regional Ethics Committee (REK no. 2018/811 Health Region West) and Data Protection Officer at Stavanger University Hospital. Consent was obtained from all participating residents and consultant surgeons for the study.

Inclusion and exclusion criteria
Eligible for inclusion were all consecutive daytime laparoscopic appendectomies performed by junior surgical trainees, as described in detail elsewhere16. Any laparoscopic procedures performed by junior trainees (4 years training or less of the surgery) under supervision by a senior trainee or consultant surgeon were eligible for inclusion. Procedures performed in children aged 16 years or under were not included.

Only procedures performed by the junior trainee in its entirety were included in analyses. Hence, evaluation forms that did not have procedure evaluation scores for any one of the eight steps (either indicating steps performed by instructors/supervisors, or missing info) were not included.

Laparoscopic appendicectomy
The procedure was a standardized, three-port laparoscopic appendicectomy defined through eight steps, as depicted in Fig. 1. The steps have been described in detail previously10,16, and follow the same lines as key steps identified by a Delphi approach18.

Definitions
The operating time for each procedure was defined as ‘knife-in’ to ‘knife-out’ and recorded in total minutes.

Junior trainees were defined as any surgical trainee with 4 years or less of training performing a laparoscopic appendicectomy under supervision by a senior trainee or consultant surgeon.

The score obtained for each of eight steps and the overall evaluation of performance was scored on a previously validated six-point scale19. For the present study (Table 1), a designation of ‘fail’, ‘pass’, or ‘proficient’ was given for each step and the complete procedure overall. A score of 1–2 was deemed a ‘fail’, a score of 3–4 as a ‘pass’, and scores of 5–6 were needed to score as ‘proficient’ for a given step or the overall procedure. The score was based on the score given by the supervisor on a formal score sheet to the trainee on each procedure step and the overall assessment of the procedure immediately after the procedure was performed.

Statistical analyses
Statistical analyses were conducted using SPSS® version 26 (IBM, Armonk, New York, USA). Descriptive statistics for continuous or categorical data were reported as medians with interquartile range (i.q.r.) or rates with percentages respectively. For analyses, a non-parametric test or chi-squared test was used as appropriate.

To explore the relationship between the performance scores for each step of the procedure to the next step (using the three-tier outcomes), we performed crosstabs that were plotted as a Sankey diagram to relate each procedure step to the evaluated performance (as ‘fail’, ‘pass’, or ‘proficient’, Table 1). A Sankey diagram was built using the SankeyMATIC software (https://sankeymatic.com) built on the open-source tool D3.js (code available at github.com/nowthis/sankeymatic).

A binary logistic regression was performed with ‘proficient’ as the dichotomous outcome variable for the procedure, exploring the independent role of each step on the final procedure outcome. The model fit was evaluated using Hosmer and Lemeshow’s goodness of fit and Nagelkerke’s $R^2$.

To further explore the weight of each step to the final procedure evaluation, an artificial neural network model was created using a multilayer perceptron. As we have previously demonstrated equal outcome on overall evaluation of the junior residents. To control for experience, a sensitivity analysis was performed for 20 or fewer, 30 or fewer, or more than 30 laparoscopic appendicectomies. Normalized importance for each step was analysed, as well as model fit using predicted pseudoprobability to actual outcomes, receiver operating characteristic (ROC) curves with area under the curve, gain graphs, and lift graphs. All statistical tests were two-tailed and statistical significance was set at $P < 0.050$.

Results
Among all procedures performed during the study interval (Fig. 2), there were 157 laparoscopic procedures that were eligible for the training of junior trainees ($n = 19$) in general surgery, for which all parts of the forms were completed by both the trainee and the supervisor ($n = 26$). The number of laparoscopic appendicectomies performed by each junior trainee at the beginning of the study was median 20 (i.q.r. 8–33) for a median procedure duration of 60 min (i.q.r. 48–74 min). Six trainees had no previous experience with laparoscopic appendicectomy before the study interval. There were no differences in the number or rate of cumulative procedures performed between trainees across sexes in the present cohort.
For overall evaluation of the junior surgeons’ performance, 97 procedures were evaluated as ‘proficient’ (61.8 per cent), 46 (29.3 per cent) as a ‘pass’, and 14 (8.9 per cent) as a ‘fail’ in the supervising surgeon’s evaluation.

By conventional regression analysis, significant associations between each of the steps and the outcome were found. The model provided a good fit (Hosmer–Lemeshow’s test, $P = 0.753$ and Nagelkerke’s $R^2$, $P = 0.735$) The variables that were retained in the final step in multivariable analyses for the model prediction are shown in Table 2 (full data are available in Table S1).

The ‘ebb-and-flow model’ is depicted in Fig. 3 and displays the fail–pass–proficiency distribution and the corresponding flow between procedure steps. From visual perception alone, the most considerable variation between steps seems to occur in the transition from step 4 to steps 5 and 6 (division of mesoappendix and the appendix).

**Fig. 1 Depiction of each procedure step and details evaluated for eight steps**

A standard three-port surgical technique was used. The steps are described as follows:

1. Abdominal cavity entry was performed by the open (Hasson’s) technique.
2. Ports were placed in a 12-mm port in the umbilicus, a 12-mm port in the left iliac fossa, and a 5-mm port over the symphysis pubis. A camera with 30° optics was used with 12–14 mmHg capno-peritoneum.
3. Appendix was identified per the location and inflammatory status.
4. Atraumatic graspers were used for manipulation and handling of intestines.
5. Mesoappendix was divided with coagulation and securing proper haemostasis of the appendicular artery through sequentially use of bipolar diathermy and cutting with the use of cold scissors.
6. The appendix base was secured using two endo-loop sutures and cut with cold scissors between loops.
7. The specimen was placed in an endo-bag and extracted through the 12-mm left iliac fossa port.
8. Fascia was closed with 1-0 Vicryl (polyglactin 910; Ethicon, Cincinnati, Ohio, USA) sutures and intracutaneous sutures and strips were used for skin closure.
A further model by a ‘string-of-pearls’ depicted in Fig. 4 demonstrates how the stumbling blocks varies between the experienced groups. The relative contribution to each step is shown in Table 3 and depicted in Fig. 4a. For the beginners (Fig. 4b), the entry of the abdomen is the major stumbling block to overcome and is hence much larger in size relative to the last
**Fig. 3** The ‘ebb-and-flow’ model of efficiency through the procedure steps
Sankey diagram showing the flow between each procedure step and the relation between score for any step, the subsequent step, and eventually to the overall evaluation of the procedure. The top third of each step represents a ‘fail’, the middle indicates a ‘pass’, and the bottom part indicates ‘proficient’ scores.

**Fig. 4** The ‘string-of-pearls’ model depicting the weighted importance of each step of the procedure
The node size represents the relative importance of the step in predicting ‘proficiency’ for the completed, overall procedure within its specific class of experience. a The entire cohort. Notably, the size of the node is only important within class and is not comparable between experience classes. As depicted, the step involving division of the mesoappendix has a relative importance independent of experience class. b,c The ‘beginners’ and ‘novices’ groups respectively. The initial steps of abdominal entry, placing ports, and identification of the appendix have a much larger relative contribution to overall proficiency—these are the steps in which they fail to perform relative to the last steps. d In the ‘expert’ group, the initial steps have a relatively smaller contribution, with a higher relative contribution of the last steps for reaching an overall proficient score.
steps, whereas for the more experienced trainees this represents a smaller obstacle than dividing the mesoappendix or subsequent steps (Fig. 4c and d).

The artificial neural network analyses produced similar results, with the highest normalized importance given to step five and step one for the entire cohort, with some notable differences according to experience of the trainee (Fig. S1) and a very good model fit (Fig. 5).

**Discussion**

In the present study we have demonstrated that the chance of completing a laparoscopic appendicectomy with a proficient score is related to the success or failure of the flow of each step through the procedure. This can be depicted in what we call an ‘ebb-and-flow’ model that depicts the flow of performance from one step to another, and a ‘string-of-pearls’ that identifies the relative size of each steps contribution to overall proficiency. These are two novel ways of depicting the relevance related to experience and performance for each predefined step in a laparoscopic appendicectomy. In the present study it was evident that for most junior trainees, the handling and division of the mesoappendix and appendix were the most essential stumbling blocks, with the largest variation between the steps and the lowest overall proficiency scored for these steps. The visual perception in the diagrams was corroborated by both

---

**Table 3 Independent variable importance based neural network analyses**

| Step variable                   | Importance | Normalized importance |
|---------------------------------|------------|-----------------------|
| Step 1: abdominal access        | 0.155      | 45.1%                 |
| Step 2: placing trocar          | 0.063      | 18.4%                 |
| Step 3: appendix identification | 0.080      | 23.3%                 |
| Step 4: handle of small bowel   | 0.081      | 23.5%                 |
| Step 5: mesoappendix division   | 0.345      | 100.0%                |
| Step 6: appendix division       | 0.132      | 38.2%                 |
| Step 7: appendix extraction     | 0.048      | 13.9%                 |
| Step 8: closing the fascia      | 0.096      | 27.9%                 |

Data are based on artificial neural network analyses for the entire cohort.

---

**Fig. 5 Prediction model evaluation by neural network analyses**

- **a** An artificial neural network model was created using a multilayer perceptron and output presented by pseudoprobability estimates to proficiency.
- **b** Receiver operating characteristics curves with area under the curve indicating model accuracy.
- **c** Gain plots.
- **d** Lift plots.
regression analyses and outcomes from artificial neural network modelling. As one would expect, slight differences were noted with lower or higher numbers of procedures performed, but the mesoappendix was consistently noted as a stumbling block across model analyses.

We believe that this model of evaluation in surgical training is generalizable to other procedures, given that pre-specified and agreed steps have been defined for evaluation. As such, this could further be pursued for laparoscopic cholecystectomy, laparoscopic totally extraperitoneal (TEP) or transabdominal preperitoneal (TAPP) hernia repair, laparoscopic distal pancreatectomy, and laparoscopic liver surgery, to mention a few. Further, it should also have appeal to robotic surgery training.

Surgeons continue to debate the number needed to obtain certification when discussing procedures, but little consensus exists to establish a universal procedure cut-off to define proficiency based on the existing literature, maybe for obvious reasons; there may be none that suits all. Where some trainees may be fast learners and obtain rapid understanding and technical handling of a specific task, others may require a longer time before reaching the level of comfort and proficiency.

In a structured training programme described previously, and in which both trainee and trainer have specific obligations and preparations to consider for enhanced learning during laparoscopic appendicectomy, we have previously shown that there was no difference between male or female trainees’ performance during surgical training. A previous study also demonstrated that most trainees became proficient in laparoscopic appendicectomy when having completed around 30 procedures; however, the same investigation also found that the mastery of each of the procedure steps varied considerably, which led to the current in-depth analyses of the connectivity between each of the steps.

Some limitations should be noted. The mesoappendix was identified as consistently the most difficult procedure step to overcome. This may be related to the institutional standard of securing the mesoappendix by means of bipolar diathermy and cold scissors, placing two endo-loops to secure the appendix, before division of the appendix with scissors. Notably, in a surgical residency programme where the institutional standard is the use of a stapler rather than placing two endo-loops, this step may be viewed as much easier; hence, scoring and evaluating must be viewed in the context of the institutional standard operating procedure for each step. However, while using endo-loops may be a more complex task for trainees, we believe that eventual mastery to the level of proficiency may have a spill-over effect on other, subsequent, and more-complex laparoscopic tasks. Thus, we believe that this is instrumental to skills training in addition to being cost-effective in our healthcare system. One should also note that, as a prerequisite for evaluation, all steps had to be performed or scored by the trainee and trainer; hence, procedures not performed or only partly performed by trainees were not part of the evaluation. There may be reasons pertained to training or competence that excludes evaluation of the trainee on these grounds, but this was not within the scope of this evaluation. The inter-observer agreement for the scoring by the trainer and trainee has been reported elsewhere.

The overall score given by the trainer was based on an overall assessment, whereas the individual steps were scored on their own. As the same trainer scored both the steps and the overall assessment, a correlation may be expected; however, the evaluation of each step still allows an evaluation of the individual steps contribution to the overall assessment, as failure or lower performance of any given procedure step stands out. This will give room for more detailed feedback in future evaluation of a procedure, rather than just an overall assessment. It also allows the identification of steps that commonly receive a lower score early on in training, and hence need focused practice to overcome the technical difficulty for such steps.

Reaching the level for which entrusted professional activity can be assured is the overall goal of surgical training. Focused training to achieve this must be the aim of modern surgical training rather than a completed number of procedures. Here we have demonstrated training models evaluating the ‘ebb-and-flow’ and ‘string-of-pears’ for procedure proficiency. We believe that identification of procedure-specific steps that are particularly cumbersome for many if not most trainees to overcome may be more beneficial than the single focus on a particular number for certification. This will allow for specific procedure-step training and preparation, which should be built into simulation tasks and dry-lab programmes. We believe the models can be used in other procedures for evaluating progress and targets for efficient training.

Funding
The authors have no funding to declare.

Disclosure. The authors declare no conflict of interest.

Supplementary material
Supplementary material is available at BJ5 Open online.

Data availability
Data from the study may be made available upon request to the investigators and for the purpose of research collaboration.

References
1. Skjold-Ødegaard B, Søreide K. Competency-based surgical training and entrusted professional activities—perfect match or a procrastinian bed? Ann Surg 2021;273:e173–e175
2. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. Lancet 2015;386:1278–1287
3. GlobalSurg Collaborative. Laparoscopy in management of appendicitis in high-, middle-, and low-income countries: a multicenter, prospective, cohort study. Surg Endosc 2018;32:3450–3466
4. Currow C, Patel K, Askari A, Rabie M, Ayl M, Aker M. Current technical surgical practice of emergency appendicectomy: a cross-sectional survey of surgical registrars in the UK. Ann R Coll Surg Engl 2020;102:606–610
5. Bohnen JD, George BC, Zwischenberger JB, Kendrick DE, Meyerson SL, Schuller MC et al. Trainee autonomy in minimally invasive general surgery in the United States: establishing a national benchmark. J Surg Educ 2020;77:e52–e62
6. Canal C, Scherer J, Birrer DL, Vehling MJ, Turina M, Neuhaus V. Appendectomy as teaching operation: no compromise in safety—an audit of 17,106 patients. J Surg Educ 2021;78:570–578
7. Speaks L, Helmer SD, Quinn KR, Lancaster J, Blythe M, Vincent KB. Chief resident indirect supervision in training safety study.
is a chief resident general surgery service safe for patients? J Surg Educ 2021;78:e145–e153
8. Singh P, Turner EJ, Cornish J, Bhangu A. Safety assessment of resident grade and supervision level during emergency appendectomy: analysis of a multicenter, prospective study. Surgery 2014;156:28–38
9. Skjold-Ødegaard B, Ersdal HL, Assmus J, Søreide K. Comparison of performance score for female and male residents in general surgery doing supervised real-life laparoscopic appendectomy: Is there a norse shield-maiden effect? World J Surg 2021;45:997–1005
10. Skjold-Ødegaard B, Hamid S, Lindeman RJ, Ersdal HL, Søreide K. Deciphering the inflection points to achieve proficiency for each procedure step during training in laparoscopic appendicectomy. BJ Open 2021;5:zrab084
11. Sinitsky DM, Fernando B, Potts H, Lykoudis P, Hamilton G, Berlingerio P. Development of a structured virtual reality curriculum for laparoscopic appendicectomy. Am J Surg 2020;219:613–621
12. Brown C, Robinson D, Egan R, Hopkins L, Abdelrahman T, Powell A et al. Prospective cohort study of haptic virtual reality laparoscopic appendicectomy learning curve trajectory. J Laparoendosc Adv Surg Tech A 2019;29:1128–1134
13. Buckley CE, Kavanagh DO, Gallagher TK, Conroy RM, Traynor OJ, Neary PC. Does aptitude influence the rate at which proficiency is achieved for laparoscopic appendectomy? J Am Coll Surg 2013;217:1020–1027
14. Kim SY, Hong SG, Roh HR, Park SB, Kim YH, Chae GB. Learning curve for a laparoscopic appendectomy by a surgical trainee. J Korean Soc Coloproctol 2010;26:324–328
15. Jaffer U, Cameron AE. Laparoscopic appendectomy: a junior trainee’s learning curve. Jsls 2008;12:288–291
16. Skjold-Ødegaard B, Ersdal HL, Assmus J, Nedrebo BSO, Sjo O, Søreide K. Development and clinical implementation of a structured, simulation-based training programme in laparoscopic appendectomy: description, validation and evaluation. BMJ Simul Technol Enhanc Learn 2021;7:517–523
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344–349
18. Bethlehem MS, Kramp KH, van Det MJ, ten Cate Hoedemaker HO, Veeger NJ, Pierie JP. Development of a standardized training course for laparoscopic procedures using Delphi methodology. J Surg Educ 2014;71:810–816
19. Miskovic D, Wyles SM, Carter F, Coleman MG, Hanna GB. Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program. Surg Endosc 2011;25:1136–1142