



OPEN

Validation of a high-fidelity training model for fetoscopic spina bifida surgery

Luc Joyeux^{1,2,3,4,12}✉, Allan Javaux^{5,12}, Mary P. Eastwood^{1,2,6}, Felix R. De Bie^{1,2,7}, Gert Van den Bergh², Rebecca S. Degliuomini¹, Simen Vergote^{1,2,3}, Talita Micheletti^{1,8}, Geertje Callewaert^{1,2,3}, Sebastien Ourselin⁹, Paolo De Coppi^{1,2,3,4}, Frank Van Calenbergh¹⁰, Emmanuel Vander Poorten⁵ & Jan Deprest^{1,2,3,11}✉

Open fetal surgery for spina bifida (SB) is safe and effective yet invasive. The growing interest in fetoscopic SB repair (fSB-repair) prompts the need for appropriate training. We aimed to develop and validate a high-fidelity training model for fSB-repair. fSB-repair was simulated in the abdominal cavity and on the stomach of adult rabbits. Laparoscopic fetal surgeons served either as novices ($n = 2$) or experts ($n = 3$) based on their experience. Technical performance was evaluated using competency Cumulative Sum (CUSUM) analysis and the group splitting method. Main outcome measure for CUSUM competency was a composite binary outcome for surgical success, i.e. watertight repair, operation time ≤ 180 min and Objective-Structured-Assessment-of-Technical-Skills (OSATS) score $\geq 18/25$. Construct validity was first confirmed since competency levels of novices and experts during their six first cases using both methods were significantly different. Criterion validity was also established as 33 consecutive procedures were needed for novices to reach competency using learning curve CUSUM, which is a number comparable to that of clinical fSB-repair. Finally, we surveyed expert fetal surgeons worldwide to assess face and content validity. Respondents (26/49; 53%) confirmed it with $\geq 71\%$ of scores for overall realism $\geq 4/7$ and usefulness $\geq 3/5$. We propose to use our high-fidelity model to determine and shorten the learning curve of laparoscopic fetal surgeons and retain operative skills.

Open fetal repair for spina bifida aperta (SBA) effectively reduces postnatal morbidity^{1,2}. Fetoscopic SBA repair (fSBA-repair) may minimize maternal risks and preterm delivery with similar neonatal neuroprotective effects^{3–6}. However, this approach is challenging with a long learning curve (LC) of at least 56 cases⁷. A valid training model would reduce the LC and avoid or limit training on clinical subjects^{8–11}. Such simulator would also accelerate the safe and ethically acceptable transition from the open approach to fetoscopy or its implementation in a center without experience based on the available case volumes. It should be paired with off- and on-site clinical training and guidance from established fetal centers¹⁰. Subsequently, it would enable skill retention and dissemination of the procedure¹².

The IDEAL recommendations for surgical innovation state that preclinical studies, including simulators and valid animal models, are essential prior to first-in-human trials^{12,13}. Simulators may be low-fidelity (e.g. computer simulators and inanimate bench-top trainers) or high-fidelity (e.g. in vivo animal models and human

¹MyFetUZ Fetal Research Center, Department of Development and Regeneration, Cluster Woman and Child, Biomedical Sciences, Catholic University KU Leuven, Leuven, Belgium. ²Center for Surgical Technologies, Faculty of Medicine, KU Leuven, Leuven, Belgium. ³Department of Obstetrics and Gynecology, Division Woman and Child, Fetal Medicine Unit, University Hospitals Leuven, Leuven, Belgium. ⁴Specialist Neonatal and Paediatric Surgery Unit, Great Ormond Street Hospital, University College London Hospitals, NHS Trust, London, UK. ⁵Department of Mechanical Engineering, KU Leuven, Leuven, Belgium. ⁶Department of Pediatric Surgery, Belfast, Northern Ireland, UK. ⁷Center for Fetal Diagnosis and Treatment, & the Perelman School of Medicine, the Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA, USA. ⁸BCNatal | Fetal Medicine Research Center, Hospital Clínic and Hospital Sant Joan de Déu, University of Barcelona, Barcelona, Spain. ⁹School of Biomedical Engineering and Imaging Sciences, King's College University, London, UK. ¹⁰Department of Neurosurgery, University Hospitals Leuven, Leuven, Belgium. ¹¹Institute of Women's Health, University College London Hospitals, London, UK. ¹²These authors contributed equally: Luc Joyeux and Allan Javaux. ✉email: luc.joyeux@kuleuven.be; jan.deprest@uzleuven.be

cadavers) training models. For fSBA-repair, two low-fidelity inanimate models have been proposed yet without validation^{14,15}. However, they may be useful in initial training and help reducing the numbers of animals needed in further training. Two in vivo SBA models have previously been proposed for feasibility studies, i.e. fetal rhesus monkey and fetal lamb. While only the latter was used for fetoscopic surgery¹⁶, neither were validated¹⁷ or specifically designed for training. Among *smaller* animal models, mice and rats cannot be used for surgical training purposes, due to their size.

High-fidelity rather than low-fidelity surgical models enhance training realism and thus minimize potential harms from the LC in animals as well as in humans following translation to clinical practice^{12,13}. Such a model for fSBA-repair requires (1) complex surgical steps like port insertion, dissection and suturing (2) simulated in a realistic environment, i.e. with proper depth perception, live motion and pulsatile blood flow in arteries and veins (3) using a living and breathing animal. Moreover, the use of large animal models should be restricted in accordance with NCR3-guidelines^{18,19}. Therefore we report on the development and validation of a high-fidelity training model for fSBA-repair in a living and breathing rabbit and its use to determine the number of cases needed for a laparoscopic fetal surgeon to achieve competency.

Methods

Ethical statement. This experiment was approved by the Animal Ethics Committee of the Group Biomedical Sciences of the KU Leuven (P093-2016). It followed the NC3Rs and the ARRIVE guidelines for animal research^{18,19}.

Study design. The validation of this animal model followed the consensus guidelines for validation of surgical simulators^{20,21}. It was assessed in two phases: construct-criterion and face-content validity²². For the former, the study was designed to train three laparoscopic fetal surgeons from a single fetal center, i.e. surgeons experienced in open fetal SBA repair as well as multi-port laparoscopy, yet who had never performed fSBA-repair (LJ, PDC, JD).

We categorized our surgeons into novices and experts in our model for simulated fSBA-repair based on surgical experience. Since we aimed to validate a training model and not the port-access approach, we hypothesized that one of them being a single-port and multiple-port laparoscopic neonatal surgeon was an expert (PDC). The multi-port approach is being clinical used for fetoscopic SBA repair^{5,23}. By contrast, the single-port approach currently applied in fetoscopy for twin-to-twin transfusion syndrome (TTTS) or congenital diaphragmatic hernia is the ultimate minimally-invasive technique to achieve^{24,25}. The other surgeons (LJ, JD) having overall less experience in multiple-port laparoscopy and no experience in single-port laparoscopy were considered as novices. They were thus trained in the model until competency in single-port fSBA-repair was reached (Supplementary Methods 1 and 2.1, Supplementary Fig. S1, Video 1). When these two surgeons had completed that training, they were then referred to as experts. In the end, these three experts performed multi-port fSBA-repair to confirm their competency (Supplementary Methods 2.2, Supplementary Fig. 1, Video 1). Overall novices and experts performed the same procedure consisting of 10 surgical steps yet used a different port-access approach (single- or multi-port).

Description of the model. This live model was developed to mimic the operative steps and conditions present for a clinical multilayered fSBA-repair.

Clinical procedure to mimic. The gestational age at fSBA-repair in humans typically is around 24 weeks of gestation^{1,5,26}. At that time the fetal weight is 662 ± 77 g²⁷ and the abdominal circumference 187 ± 10 mm²⁷. The region of interest is lumbar in 95% of cases¹. The current literature on open and fSBA-repair describes several steps in the procedure, which we summarized into 10 consecutive steps to be simulated (Fig. 1)^{1,5,26,28,29}.

Animal model. Rabbits have previously been used for training in pediatric^{30,31} and fetoscopic³² surgery. We used New-Zealand male adult rabbits (weight, 3–4 kg). They were given water and food at libitum under the standard light–dark cycle until the procedure. They were put under general anesthesia without endotracheal intubation (Supplementary Methods 3). In rabbits, the adequately insufflated abdominal cavity mimics the working space or amniotic cavity, which approximately measures $15 \times 10 \times 5$ cm with a pneumoperitoneum of around 3 L with CO₂ at 5 mmHg (Fig. 1A)³². Monogastric herbivores, rabbits have a large single-chamber stomach with a circumference of 165 ± 13 mm mimicking the abdominal circumference of a 22–24 weeks human fetus (Fig. 1A)³³. Overall 10 clinical steps for fSBA-repair are recapitulated by a laparoscopic gastric Nissen fundoplication³⁴ and the suturing of a patch to the gastric wall (Fig. 1B, Video 1). These procedures require the ability to gently manipulate fragile tissue, perform extensive dissection, hemostasis and suture (Supplementary Methods 3).

Technical performance. Clinical outcomes were total operation time (steps 1 to 9)^{8,9}, fetal repair time (steps 4 to 9), CO₂ insufflation volume, Objective Structured Assessment of Technical Skill (OSATS) score and watertightness of the patch repair (step 10, Video 1). The latter was tested post-mortem by fluorescein injection under the patch after completion of the repair (Supplementary Methods 4, Video 1). Operative performance and difficulty were assessed applying an adapted OSATS rating scale on videos of the procedures (Supplementary Methods 4)^{35,36}.

We also used a composite binary outcome for surgical success based on clinically relevant outcomes to measure the LC and competency level of fSBA-repair⁷. Similar to clinical fetal SBA surgery⁷, a successful surgical repair of simulated SBA was defined by a watertight repair, an operation time ≤ 180 min in accordance with

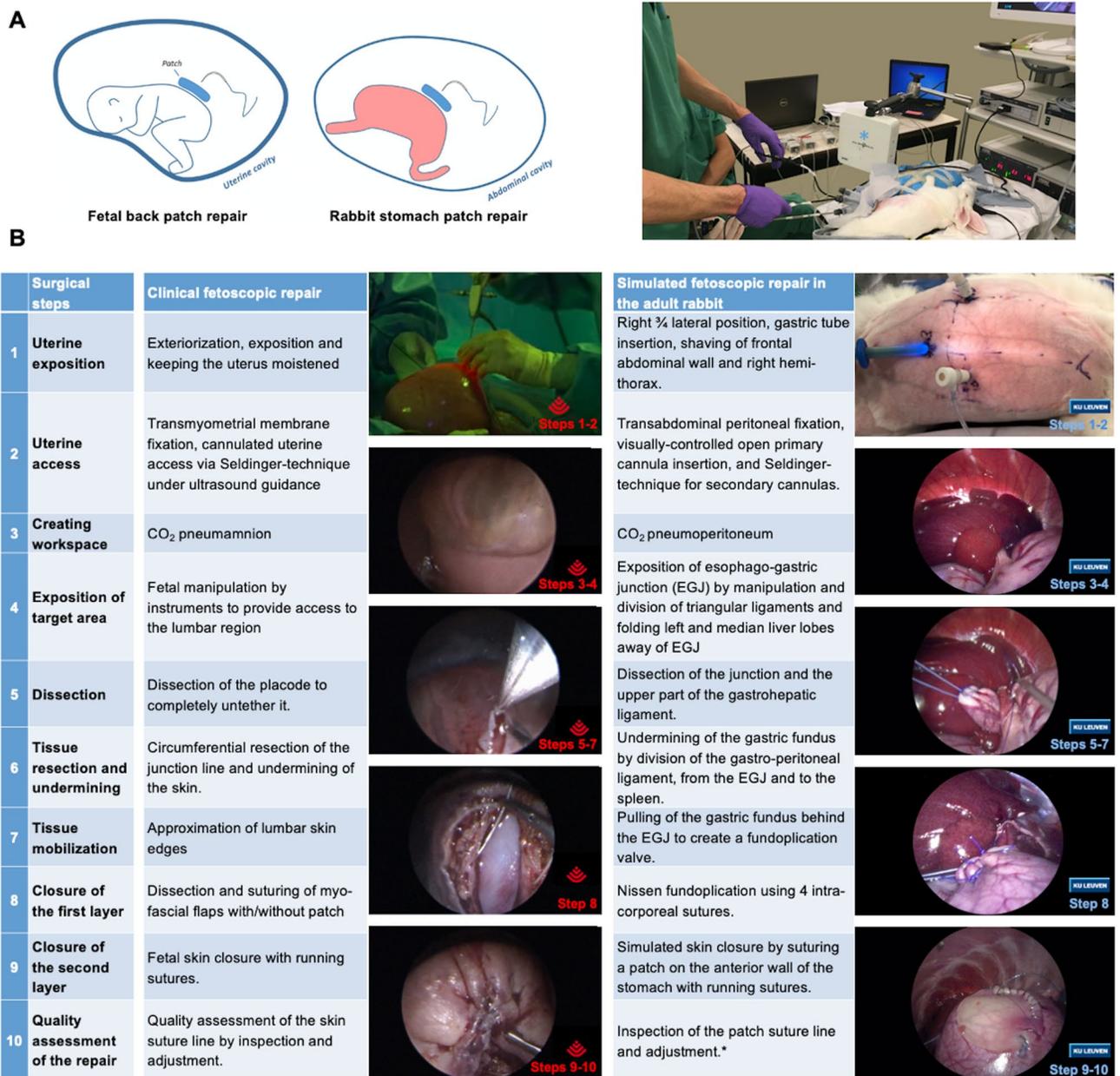


Figure 1. Comparison of clinical and simulated fetoscopic spina bifida repair in our rabbit training model. (A) Schematic drawing and external view of our model setup showing similarities in working space and presentation of the surgical target. (B) Comparison of the 10 essential steps performed during a clinical (left column) and simulated (right column) fetoscopic repair. Reproduced with permission of and copyright by the Texas Children’s Hospital, Houston, TX, USA and UZ Leuven. Artist drawing by Allan Javaux; copyright by UZ Leuven, Leuven, Belgium.

the FDA Drug Safety Communication about potential risks of general anesthesia in pregnant women³⁷ and an OSATS score $\geq 18/25$ ($> 70\%$)^{35,36}.

Validation study. *Construct validity.* To assess construct validity and therefore discriminate performance levels of our simulator, we determined and compared competency level of novices and experts during their first six cases applying two methods. First, the Competency Cumulative Sum (C-CUSUM) test³⁸ used the aforementioned binary outcome for surgical success and was set with a control limit of $h_c = 3$ (Supplementary Methods 5.1). Subsequently, we applied the group-splitting method⁹ by comparing performance using the five aforementioned clinical outcomes.

Criterion validity. Criterion validity compares performance of our innovative model to the ground truth which is the clinical procedure in our case since, to the best of our knowledge, no animal training model has been validated yet⁷. Herein we compared the learning curve (LC) of novices in our model to the LC of novices performing

percutaneous fSBA-repair⁷. It was determined by the predictive validity method³⁹ that uses the Learning Curve CUSUM (LC-CUSUM) test with similar parameters as the C-CUSUM yet with a control limit set at $h_{LC}=0.85$ (Supplementary Methods 5.2)⁷.

Face and content validity. To assess realism (face validity) and usefulness (content validity) of our model, an anonymous online survey (Supplementary Online Survey SurveyMonkey #6VNJMS9) was designed and sent to fetal surgeons ($n=49$) worldwide who are currently involved in clinical fetal surgery programs for SBA using an open fetal and/or fetoscopic approach (Supplementary Methods 5.3, Supplementary Table S1)⁷. 22% (11/49) of the surgeons were performing fSBA-repair in their fetal center at that time. All were invited to try our model out in our research center and under our supervision to get a realistic experience. For obvious geographical reasons, some experts only answered our anonymous online questionnaire (non-users) while others also tried it (users). According to current practice, all survey responses from expert users or non-users were included in the analysis⁴⁰. We also performed a subanalysis of data from users currently performing fetoscopic repair in humans.

Expert fetal surgeons were categorized according to their specialty (obstetricians and gynecologists, pediatric neurosurgeons and pediatric surgeons) and demographic data were captured by seven questions. We additionally asked six questions on face validity and five on content validity, using a 7-point and 5-point Likert scale respectively. We set validity thresholds at 4/7 (undecided) and 3/5 (neutral) for each scale (Supplementary Methods 5.3)³⁹.

Statistical analysis. We used GraphPad Prism version 7.00 for MacOS X (GraphPad, La Jolla, CA, USA) to analyze the data. Binomial and categorical variables were expressed as percentages with their frequency distribution. Continuous variables were tested for normal distribution using the D'Agostino-Pearson (omnibus K2) or Shapiro-Wilk normality tests. Continuous variables normally distributed were presented as mean and standard deviation (SD) and the others were expressed as median and range or interquartile range (IQR).

For face and content validity, categorical and continuous variables based on the response by the three clinical subspecialties involved in the survey were compared with one-way analysis of variance (ANOVA) or Kruskal-Wallis test as appropriate. For construct validity, the Fisher exact test was used to compare binomial variables. Continuous variables were compared with unpaired two-tailed t test or Mann Whitney test as appropriate. A p value <0.05 was considered significant. For the construct and criterion validity, we performed C- and LC-CUSUM analysis using an algorithm that we developed in MATLAB software (Mathworks, Natick, MA, USA) based on and verified with the model of Biau et al.⁷.

Results

Surgical procedures. The data below were collected from 52 completed single-port simulated fSBA-repairs by two novices ($n=34$ and $n=18$), and from 18 multi-port simulated fSBA-repairs by three experts ($n=6$, $n=6$ and $n=6$).

Construct validity. Figure 2A displays the evolution of the C-CUSUM score of each novice and expert during the first six cases. As C-CUSUM scores of the three experts remained below the competency control limit $h_C=3$, they were considered competent. In contrast, scores of both novices reached a higher score of 3.07 after the 6th case and were therefore not competent (graphs of Fig. 2A). The two groups were also significantly different for all surgical outcomes measured when applying the group-splitting method (table of Fig. 2A).

Criterion validity. LC-CUSUM analysis of all the cases performed demonstrated that novice 1 reached competency when his score was 0.89, exceeding the control limit of $h_{LC}=0.85$, after 33 out of 34 cases (Fig. 2B, left graph). To reduce the number of animals used, novice 2 performed only 18 procedures and did not reach that threshold. When considering the best-case scenario, it was predicted that this novice reached competency at 31 cases (Fig. 2B, right graph). These numbers involving laparoscopic fetal surgeons are lower than what has been reported as the minimum number required for competency in clinical fSBA-repair performed by non-laparoscopic fetal surgeons ($n \geq 56$)⁷.

Demographics of survey respondents. The response rate from the fetal surgeons for SBA to the online survey was 53% (26/49). 38% (10/26) of the respondents tried the model out to get a realistic experience and 27% (7/26) were performing fSBA-repair in their fetal center (Supplementary Table S2). The demographics of the different subspecialists involved were comparable, except that pediatric neurosurgeons had less experience in laparoscopic surgery (Table 1). None of the respondents was aware of computer simulators or high-fidelity models for fSBA-repair.

Face validity. The three subspecialties of respondents considered our live model realistic since all survey questions reached scores $\geq 4/7$ in $\geq 60\%$ of cases and recommended it for its realism ($\geq 84\%$ of scores $\geq 4/7$; Table 2). Good to exceptional scores (63 to 100% of scores $\geq 4/7$) were reached for questions that required a living and breathing animal. These encompassed the surgical target, surgical steps such as dissection, resection and suturing, and depth perception (Table 2). Despite some discrepancies, there were no significant differences in ratings between the subspecialists. In addition, the seven fetoscopic experts who used our model confirmed its realism as 95% of survey questions scored $\geq 4/7$. They also recommended it for its realism (100% of scores $\geq 4/7$; Supplementary Table S3).

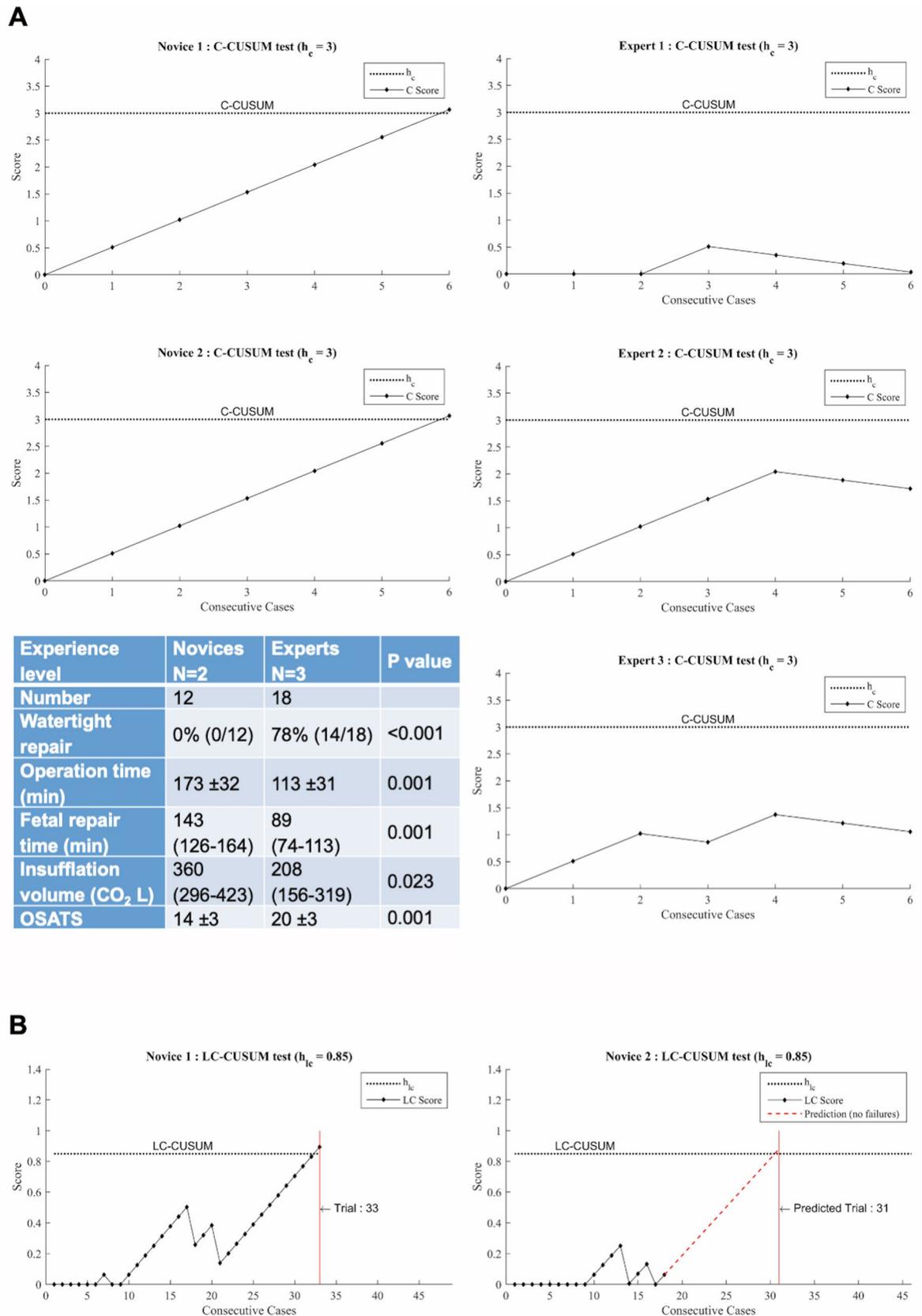


Figure 2. Construct and criterion validity of our training model. **(A)** Construct validity: competency assessment of novices and experts using competency cumulative sum (C-CUSUM) test (graphs), and comparison of clinical outcomes of the first six cases of novices and experts using the group splitting method (table). **(B)** Criterion validity: learning curve assessments of novices and experts using learning curve CUSUM (LC-CUSUM) test (graph).

Specialty	Obstetrician and gynecologist	Pediatric neurosurgeon	Pediatric surgeon	p-value
Number	11	6	9	
Number of years of experience as a				
Specialist	16 ± 13	22 ± 9	17 ± 8	0.566
Laparoscopic surgeon	7 (10)	0 (1)	13 (10)	0.003
Handedness				
Right-Handed	11	5	7	
Left-Handed	0	1	2	
Number of fetal SB open repairs in humans				
As first surgeon	1–30 (0;1–30)	31–60 (1–30;91–120)	0 (0;31–60)	0.230
As second surgeon	1–30 (0;31–60)	1–30 (0;1–30)	1–30 (1–30;61–90)	0.363
Number of fetal SB fetoscopic repairs in humans				
As first surgeon	0 (0;0)	0 (0;1–30)	0 (0;1–30)	0.977
As second surgeon	0 (0;0)	0 (0;0)	0 (0;1–30)	0.779
Number of training sessions on an open fetal surgery simulator				
Virtual reality	0 ± 0	0 ± 0	0 ± 0	NA
Box trainer	0 (0;1–30)	0 (0;0)	0 (0;0)	0.108
Animal model	0 (0;1–30)	0 (0;0)	0 (0;31–60)	0.643
Human cadaver	0 (0;0)	0 (0;0)	0 (0;0)	0.722
Number of training sessions on a fetoscopic surgery simulator				
Virtual reality	0 ± 0	0 ± 0	0 ± 0	NA
Box trainer	0 (0;1–30)	1–30 (0;1–30)	0 (0;1–30)	0.449
Animal model	0 (0;1–30)	0 (0;1–30)	0 (0;1–30)	0.648
Human cadaver	0 ± 0	0 ± 0	0 ± 0	NA

Table 1. Demographics of the three subspecialties involved in fetal surgery. Number of surgeries per specialty are displayed in median and range based on 5 block of 30 cases as per the online survey (0, 1–30, 31–60, 61–90, 91–120, ≥ 121). For example 31–60 (1–30;91–120) signifies that a mean of 1–30 cases was performed and the number of cases varied between 1–30 and 91–120.

Content validity. There were no significant differences in ratings between the subspecialists. Respondents considered the model useful and would recommend it for training ($\geq 71\%$ of scores $\geq 3/5$) and improving complex fetoscopic skills, such as instrument handling and suturing ($\geq 83\%$ and 100% of score $\geq 3/5$ respectively; Table 2). Average ($\geq 50\%$) scores varied among the three subspecialties and were obtained when we asked whether the model exposes to stress similar to that in clinical conditions and is useful to train for fetal positioning ($\geq 50\%$ and $\geq 60\%$ of scores $\geq 3/5$ respectfully). In contrast high scores ($\geq 83\%$ of scores $\geq 3/5$) concerned the usefulness of the model for instrument handling, suturing, self-confidence, insufflation, tissue mobilization, skin closure and quality assessment of the patch repair. Finally, the seven fetoscopic experts who used our model confirmed its usefulness as 98% of survey questions scored $\geq 3/5$. They also recommended it for its training (100% of scores $\geq 3/5$; Supplementary Table S3).

Discussion

Main findings. We developed and validated a high-fidelity training model for fetoscopic SBA repair in live rabbits. We first demonstrated that competency of laparoscopic fetal surgeons was reached at 33 consecutive cases. That number is lower than what has been reported for non-laparoscopic fetal surgeons performing clinical multi-port fSBA-repair. Surveyed fetal experts also proved face and content validity.

Clinical interpretation. A recent systematic review with available individual patient data on multi-port fSBA-repair demonstrated that the LC to reach competency was at least 56 cases⁷. This is of the same order of magnitude of other complex multi-port laparoscopic surgeries such as colectomy^{41,42} or sacrocolpopexy⁴³ performed by surgeons without previous experience in these techniques. In our high-fidelity model the LC was 33 cases for novices, more precisely surgeons experienced in multi-port laparoscopic surgery but non-experienced with single-port surgery. That number is in keeping with other complex single-port procedures such as colectomy performed by surgeons experienced in multi-port laparoscopic surgery⁴⁴. It is also similar to other advanced, yet less complex multi-port endoscopic procedures than those above, such as cholecystectomy⁴⁵, pyloromyotomy⁴⁶ or the most common fetoscopic operation (laser coagulation for TTTS)²⁴. They are described as less complex as they do not require suturing and extensive dissection (or in case of TTTS none at all) skills. We therefore surmise that the number of 33 reached in our simulator is the LC of surgeons experienced with complex multi-port laparoscopy before translation to clinical practice. It may represent an underestimation for surgeons non-experienced with complex multi-port laparoscopy. Clinically, the challenges of fetal surgery are greater than what can be simulated, such as the complex pathologic anatomy of the lesion, the frailty of human fetal tissue, the interference of fetal monitoring, the presence and vicinity of the placenta, the large number of

Specialty	Obstetrician and gynecologist	Pediatric neurosurgeon	Pediatric surgeon	P value
Number	11	6	9	
Face validity	Percentages of scores $\geq 4/7$ on the Likert scale			
Overall realism				
Recommendation for realism	88%	94%	84%	0.874
Surgical scene				
Animal positioning	80%	83%	75%	0.916
Position of the video monitor	100%	100%	100%	0.216
Position of 1st surgeon	100%	100%	88%	0.557
Position of 2nd surgeon	100%	100%	75%	0.933
Surgical cavity				
Humidified environment (Fluid-gas interface)	90%	100%	88%	0.804
Workspace	90%	100%	88%	0.034*
Vision	100%	100%	88%	0.419
Surgical target				
Mimic of the fetal lumbar region	70%	84%	73%	0.961
Instrumentation set				
Endoscope	90%	100%	100%	0.440
Grasping forceps	90%	100%	100%	0.253
Scissors	100%	100%	100%	0.648
Coagulating and dissecting hook	100%	100%	100%	0.646
Dissector	100%	100%	100%	0.349
Needle holders	100%	100%	100%	0.167
Surgical steps				
Exposition	80%	83%	75%	0.750
Port insertion	60%	83%	88%	0.201
Insufflation	100%	100%	100%	0.753
Fetal positioning	70%	83%	63%	0.834
Dissection	70%	83%	63%	0.964
Resection	70%	83%	63%	0.889
Mobilization	80%	100%	63%	0.984
Patch	90%	83%	63%	0.521
Skin	90%	83%	75%	0.159
Quality assessment	100%	83%	88%	0.507
Depth perception				
Mimic of clinical conditions (live motions)	80%	100%	100%	0.110
Content validity	Percentages of scores $\geq 3/5$ on the Likert scale			
Overall usefulness				
Recommendation for training	80%	83%	71%	0.332
Overall difficulty				
As difficult as in humans	80%	100%	71%	0.815
Similar stress	50%	50%	86%	0.104
Instrument handling				
Improvement of instrument handling skills	100%	83%	100%	0.690
Suturing				
Improvement of suturing skills	100%	100%	100%	0.832
Self-confidence				
Overall confidence	100%	100%	100%	0.180
Surgical Tasks				
Exposition	80%	83%	71%	0.501
Port insertion	90%	83%	71%	0.568
Insufflation	100%	83%	100%	0.718
Fetal positioning	60%	83%	71%	0.712
Dissection	80%	83%	71%	0.595
Resection	80%	83%	71%	0.642
Continued				

Face validity	Percentages of scores $\geq 4/7$ on the Likert scale			
Mobilization	80%	83%	86%	0.829
Patch	90%	83%	71%	0.939
Skin	90%	83%	100%	0.385
Quality assessment	100%	83%	100%	0.406

Table 2. Face and content validity of the model. There were no significant differences between each of the 3 specialties of fetal surgeons.

people and specialties around the operation table, or simply the stress of operating on two patients. These points were suggested by 6/26 (23%) of the survey participants (3/6 pediatric neurosurgeons, 2/9 pediatric surgeons and 1/11 obstetrician and gynecologist).

Strength and limitations. We acknowledge a number of limitations to our study. First, some surveyed fetal surgeons raised concerns regarding the realism of specific aspects of our model. Indeed, the simulation procedure does not mimic the precise dissection and gentle manipulation of the dura, musculo-fascial flaps and fetal skin. However, manipulation, dissection and suturing of the stomach are quite comparable to these clinical steps, as the rabbit stomach can be easily damaged and perforated. This way, those steps unmask potential clinically relevant complications. Secondly, three laparoscopic fetal surgeons were involved in our fetal surgery training to reduce the number of animal required hence following the ethical standards of the NC3Rs-guidelines^{18,19}. Our competency analysis allowed us to confirm our hypothesis about the competency level of novices and experts. Since experts 1 and 2 were previously novices 1 and 2, these surgeons became experts in three-port surgery in our rabbit training model—yet not for clinical fSBA-repair—following their training as novices in single-port surgery. Finally, we only tested our model for either a single- and three-port²³. A two-port approach²⁶ currently practiced by some centers can easily be adapted.

Our study also has considerable strengths. Firstly, in the development of a simulator, we followed the consensus guidelines for animal research and validation of animal models^{18,39} and surgical simulators^{13,14,22}. Secondly, we applied robust methods for assessing subjective and objective validity. Thirdly, we measured the LC and competency level of both experts and novices with standardized methodology³⁸. Finally, our observations seem clinically relevant as we come to numbers that are comparable to what has been demonstrated for complex clinical laparoscopic procedures performed by trained laparoscopic surgeons.

Conclusion

We developed and validated a high-fidelity model for fetoscopic layered SBA repair. It was used to determine the learning curve of laparoscopic fetal surgeons, which was in the range of other complex endoscopic procedures. We propose the use of this model to determine and shorten the learning curve of laparoscopic fetal surgeons, and aid retention of operative skills.

Received: 24 July 2020; Accepted: 15 February 2021

Published online: 17 March 2021

References

- Adzick, N. S. *et al.* A randomized trial of prenatal versus postnatal repair of myelomeningocele. *N. Engl. J. Med.* **364**, 993–1004. <https://doi.org/10.1056/NEJMoa1014379> (2011).
- Farmer, D. L. *et al.* The Management of Myelomeningocele Study: Full cohort 30-month pediatric outcomes. *Am. J. Obstet. Gynecol.* **218**(256), e251–256. <https://doi.org/10.1016/j.ajog.2017.12.001> (2018).
- Joyeux, L. *et al.* Fetoscopic versus open repair for spina bifida aperta: A systematic review of outcomes. *Fetal Diagn. Ther.* **39**, 161–171. <https://doi.org/10.1159/000443498> (2016).
- Kabagambe, S. K., Jensen, G. W., Chen, Y. J., Vanover, M. A. & Farmer, D. L. Fetal surgery for myelomeningocele: A systematic review and meta-analysis of outcomes in fetoscopic versus open repair. *Fetal Diagn. Ther.* **43**, 161–174. <https://doi.org/10.1159/000479505> (2018).
- Lapa Pedreira, D. A. *et al.* Percutaneous fetoscopic closure of large open spina bifida using a bilaminar skin substitute. *Ultrasound Obstet. Gynecol.* **52**, 458–466. <https://doi.org/10.1002/uog.19001> (2018).
- Belfort, M. A. *et al.* Comparison of two fetoscopic open neural tube defect repair techniques: Single- vs three-layer closure. *Ultrasound Obstet. Gynecol.* **56**, 532–540. <https://doi.org/10.1002/uog.21915> (2020).
- Joyeux, L. *et al.* Learning curves of open and endoscopic fetal spina bifida closure: Systematic review and meta-analysis. *Ultrasound Obstet. Gynecol.* **55**, 730–739. <https://doi.org/10.1002/uog.20389> (2020).
- Harrysson, I. J. *et al.* Systematic review of learning curves for minimally invasive abdominal surgery: A review of the methodology of data collection, depiction of outcomes, and statistical analysis. *Ann. Surg.* **260**, 37–45. <https://doi.org/10.1097/SLA.0000000000000596> (2014).
- Khan, N., Abboudi, H., Khan, M. S., Dasgupta, P. & Ahmed, K. Measuring the surgical “learning curve”: Methods, variables and competency. *Bju Int.* **113**, 504–508. <https://doi.org/10.1111/bju.12197> (2014).
- Cohen, A. R. *et al.* Position statement on fetal myelomeningocele repair. *Am J Obstet Gynecol* **210**, 107–111. <https://doi.org/10.1016/j.ajog.2013.09.016> (2014).
- Szasz, P., Louridas, M., de Montbrun, S., Harris, K. A. & Grantcharov, T. P. Consensus-based training and assessment model for general surgery. *Br. J. Surg.* **103**, 763–771. <https://doi.org/10.1002/bjs.10103> (2016).

12. McCulloch, P. *et al.* No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* **374**, 1105–1112. [https://doi.org/10.1016/S0140-6736\(09\)61116-8](https://doi.org/10.1016/S0140-6736(09)61116-8) (2009).
13. Hirst, A. *et al.* No Surgical innovation without evaluation: Evolution and further development of the IDEAL framework and recommendations. *Ann. Surg.* **269**, 211–220. <https://doi.org/10.1097/SLA.0000000000002794> (2019).
14. Belfort, M. A., Whitehead, W. E., Bednov, A. & Shamshirsaz, A. A. Low-fidelity simulator for the standardized training of fetoscopic meningomyelocele repair. *Obstet. Gynecol.* **131**, 125–129. <https://doi.org/10.1097/AOG.0000000000002406> (2018).
15. Miller, J. L. *et al.* Ultrasound based three-dimensional medical model printing for multispecialty team surgical rehearsal prior to fetoscopic myelomeningocele repair. *Ultrasound Obstet. Gynecol.* <https://doi.org/10.1002/uog.18891> (2017).
16. Joyeux, L. *et al.* Safety and efficacy of fetal surgery techniques to close a spina bifida defect in the fetal lamb model: A systematic review. *Prenat. Diagn.* **38**, 231–242. <https://doi.org/10.1002/pd.5222> (2018).
17. Joyeux, L. *et al.* Validation of the Fetal Lamb model of spina bifida. *Sci. Rep.* **9**, 9327. <https://doi.org/10.1038/s41598-019-45819-3> (2019).
18. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8**, e1000412. <https://doi.org/10.1371/journal.pbio.1000412> (2010).
19. NC3Rs. *Conducting a Pilot Study*. <https://www.nc3rs.org.uk/conducting-pilot-study#data> (2016).
20. Carter, F. J. *et al.* Consensus guidelines for validation of virtual reality surgical simulators. *Surg. Endosc.* **19**, 1523–1532. <https://doi.org/10.1007/s00464-005-0384-2> (2005).
21. Schout, B. M., Hendrikx, A. J., Scheele, F., Bemelmans, B. L. & Scherpbier, A. J. Validation and implementation of surgical simulators: A critical review of present, past, and future. *Surg. Endosc.* **24**, 536–546. <https://doi.org/10.1007/s00464-009-0634-9> (2010).
22. Dawson, B. & Trapp, R. G. *Basic and Clinical Biostatistics* 287–289 (McGraw-Hill Companies, London, 2004).
23. Gine, C. *et al.* Fetoscopic two-layer closure of open neural tube defects. *Ultrasound Obstet. Gynecol.* **52**, 452–457. <https://doi.org/10.1002/uog.19104> (2018).
24. Peeters, S. H. *et al.* Learning curve for fetoscopic laser surgery using cumulative sum analysis. *Acta Obstet. Gynecol. Scand.* **93**, 705–711. <https://doi.org/10.1111/aogs.12402> (2014).
25. Jani, J. C. *et al.* Severe diaphragmatic hernia treated by fetal endoscopic tracheal occlusion. *Ultrasound Obstet. Gynecol.* **34**, 304–310. <https://doi.org/10.1002/uog.6450> (2009).
26. Belfort, M. A. *et al.* Fetoscopic open neural tube defect repair: Development and refinement of a two-port, carbon dioxide insufflation technique. *Obstet. Gynecol.* **129**, 734–743. <https://doi.org/10.1097/AOG.0000000000001941> (2017).
27. Salomon, L. J., Bernard, J. P. & Ville, Y. Estimation of fetal weight: reference range at 20–36 weeks' gestation and comparison with actual birth-weight reference range. *Ultrasound Obstet. Gynecol.* **29**, 550–555. <https://doi.org/10.1002/uog.4019> (2007).
28. Ovaere, C. *et al.* Prenatal diagnosis and patient preferences in patients with neural tube defects around the advent of fetal surgery in Belgium and Holland. *Fetal Diagn. Ther.* **37**, 226–234. <https://doi.org/10.1159/000365214> (2015).
29. Heuer, G. G., Adzick, N. S. & Sutton, L. N. Fetal myelomeningocele closure: Technical considerations. *Fetal Diagn. Ther.* **37**, 166–171. <https://doi.org/10.1159/000363182> (2015).
30. Luks, F. I., Peers, K. H., Deprest, J. A. & Lerut, T. E. Gasless laparoscopy in infants: The rabbit model. *J. Pediatr. Surg.* **30**, 1206–1208. [https://doi.org/10.1016/0022-3468\(95\)90023-3](https://doi.org/10.1016/0022-3468(95)90023-3) (1995).
31. Kirlum, H. J., Heinrich, M., Tillo, N. & Till, H. Advanced paediatric laparoscopic surgery: Repetitive training in a rabbit model provides superior skills for live operations. *Eur. J. Pediatr. Surg.* **15**, 149–152. <https://doi.org/10.1055/s-2005-837600> (2005).
32. Quintero, R. A. *et al.* Hydrolaparoscopy in the rabbit: A fine model for the development of operative fetoscopy. *Am. J. Obstet. Gynecol.* **171**, 1139–1142. [https://doi.org/10.1016/0002-9378\(94\)90052-3](https://doi.org/10.1016/0002-9378(94)90052-3) (1994).
33. Chitty, L. S., Altman, D. G., Henderson, A. & Campbell, S. Charts of fetal size: 3. Abdominal measurements. *Br. J. Obstet. Gynaecol.* **101**, 125–131. <https://doi.org/10.1111/j.1471-0528.1994.tb13077.x> (1994).
34. Jamieson, G. G., Watson, D. I., Britten-Jones, R., Mitchell, P. C. & Anvari, M. Laparoscopic Nissen fundoplication. *Ann. Surg.* **220**, 137–145. <https://doi.org/10.1097/0000658-199408000-00004> (1994).
35. Martin, J. A. *et al.* Objective structured assessment of technical skill (OSATS) for surgical residents. *Br. J. Surg.* **84**, 273–278. <https://doi.org/10.1046/j.1365-2168.1997.02502.x> (1997).
36. Swift, S. E. & Carter, J. F. Institution and validation of an observed structured assessment of technical skills (OSATS) for obstetrics and gynecology residents and faculty. *Am. J. Obstet. Gynecol.* **195**, 617–621. <https://doi.org/10.1016/j.ajog.2006.05.032> (2006) (**discussion 621–613**).
37. FDA, U. S. *FDA Review Results in New Warnings About Using General Anesthetics and Sedation Drugs in Young Children and Pregnant Women*. <https://www.fda.gov/Drugs/DrugSafety/ucm532356.htm> (2017).
38. Biau, D. J. & Porcher, R. A method for monitoring a process from an out of control to an in control state: Application to the learning curve. *Stat. Med.* **29**, 1900–1909. <https://doi.org/10.1002/sim.3947> (2010).
39. Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R. & Young, S. L. Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Front. Public Health* **6**, 149. <https://doi.org/10.3389/fpubh.2018.00149> (2018).
40. Connell, J. *et al.* The importance of content and face validity in instrument development: Lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Qual. Life Res.* **27**, 1893–1902. <https://doi.org/10.1007/s11136-018-1847-y> (2018).
41. Jamali, F. R. *et al.* Evaluating the degree of difficulty of laparoscopic colorectal surgery. *Arch. Surg.* **143**, 762–767. <https://doi.org/10.1001/archsurg.143.8.762> (2008) (**discussion 768**).
42. Tekkis, P. P., Senagore, A. J., Delaney, C. P. & Fazio, V. W. Evaluation of the learning curve in laparoscopic colorectal surgery: Comparison of right-sided and left-sided resections. *Ann. Surg.* **242**, 83–91. <https://doi.org/10.1097/01.sla.0000167857.14690.68> (2005).
43. Claerhout, F. *et al.* Analysis of the learning process for laparoscopic sacrocolpopexy: Identification of challenging steps. *Int. Urogynecol. J.* **25**, 1185–1191. <https://doi.org/10.1007/s00192-014-2412-z> (2014).
44. Kim, C. W. *et al.* Learning curve for single-port laparoscopic colon cancer resection: A multicenter observational study. *Surg. Endosc.* **31**, 1828–1835. <https://doi.org/10.1007/s00464-016-5180-7> (2017).
45. Moore, M. J. & Bennett, C. L. The learning curve for laparoscopic cholecystectomy. The Southern Surgeons Club. *Am. J. Surg.* **170**, 55–59. [https://doi.org/10.1016/s0002-9610\(99\)80252-9](https://doi.org/10.1016/s0002-9610(99)80252-9) (1995).
46. Oomen, M. W., Hoekstra, L. T., Bakx, R. & Heij, H. A. Learning curves for pediatric laparoscopy: How many operations are enough? The Amsterdam experience with laparoscopic pyloromyotomy. *Surg. Endosc.* **24**, 1829–1833. <https://doi.org/10.1007/s00464-010-0880-x> (2010).

Acknowledgments

We are very grateful to Rosita Kinnart from the Center for Surgical Technologies for her precious help in preparing the surgeries. We also thank the following colleagues for their technical support: Dr. Lennart Van der Veeken, Dr. David Basurto, Dr. Johanna Parra, Dr. Enrico Corno and Dr. Yada Kunpalin.

Author contributions

All authors of this multicenter study contributed substantially to the submitted manuscript. L.J. and A.J. designed the entire study with the help of M.P.E., P.D.C., E.V.P. and J.D. L.J., J.D. and P.D.C. performed the surgeries and animal preparations with help of M.P.E., T.M., R.S.D., G.G. The MATLAB algorithm was developed by A.J. A.J. and L.J. performed the cumulative sum analysis using the custom-made algorithm made by A.J. and validated. F.D.B., R.S.D. and G.V.D.B. independently analyzed and rated the videos of the surgeries and scored them using the OSATS rating scale. L.J. and S.V. created the online questionnaire for the face and content validities. L.J. and A.J. performed the statistical analysis of the data and wrote the manuscript. L.J., A.J. and J.D. drafted and edited the manuscript. All co-authors made substantial contributions to the study design, acquisition and interpretation of data, critically revised and approved the final submitted manuscript.

Funding

LJ is supported by an Innovative Engineering for Health award by the Wellcome Trust (WT101957) and the Engineering and Physical Sciences Research Council (EPSRC) (NS/A000027/1). AJ is supported by a grant from The Research Foundation – Flanders (FWO) [SB/1S63117N]. MPD and TM were funded with support of the Erasmus + Programme of the European Union (Framework Agreement number: 2013–0040). FDB is supported by a Fulbright Research grant. PDC is supported by the National Institute for Health Research (NIHR-RP-2014-04-046) and the BRC NIHR GOSH. JD was a Clinical Researcher of the FWO (FWO-1.8.012.07) and is currently supported by the Great Ormond Street Hospital Charity fund.

Competing interests

The authors declare no conflict of interests. All the instrumentation from the company KARL STORZ were part of a research loan (agreement 1705230948). Above mentioned support of KARL STORZ is in no way affiliated with any other service or procurement decisions on the part of the contractual parties.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-85607-6>.

Correspondence and requests for materials should be addressed to L.J. or J.D.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021