Operative competence in fetoscopic laser surgery for TTTS: a procedure-specific evaluation

SHP Peeters[1#], J Akkermans[1#], J Bustraan[2], J.M. Middeldorp[1], E Lopriore[3], R Devlieger[4], L Lewi[4], J Deprest[4], D Oepkes[1]

1. Department of Obstetrics, Division of Fetal Medicine, Leiden University Medical Center, the Netherlands
2. PLATO, Center for Research and Development in Education and Training, Faculty of Social Sciences, Leiden, the Netherlands
3. Department of Pediatrics, Division of Neonatology, Leiden University Medical Center, the Netherlands
4. Department of Obstetrics, Division of Fetal Medicine, University Hospitals Leuven, Department Development and Regeneration, KU Leuven, Belgium

# both of these authors contributed equally to this work.

Disclosure: None of the authors have a conflict of interest.

Corresponding author:

Suzanne Peeters, Dept. of Obstetrics, Leiden University Medical Center

K6-32, P.O. Box 9600,

2300 RC Leiden, the Netherlands

Phone: +31-71-5262896, Fax: +31-71-5266741

s.h.p.peeters@lumc.nl

Keywords: twin-to-twin transfusion syndrome, evaluation, training, simulator, teaching, laser therapy, monochorionic twins

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:10.1002/uog.15734

ABSTRACT

OBJECTIVE: Fetoscopic laser surgery for twin-twin transfusion syndrome is a procedure for which no objective tools exist to assess technical skills. To ensure that future fetal surgeons reach competence prior to performing the procedure unsupervised, we developed a performance assessment tool. The aim of this study was to validate this assessment tool for reliability and construct validity.

METHODS: A procedure-specific evaluation instrument containing all essential steps of the fetoscopic laser procedure was created using Delphi methodology. Eleven experts and 13 novices from three Fetal Medicine centers performed the procedure on the same simulator. Two independent observers assessed each surgery using the instrument (maximum score: 52). Inter-observer reliability was assessed using Spearman correlation. We compared performance of novices and experts to assess construct validity.

RESULTS: Inter-observer reliability was high ($r=0.974$, $p<0.001$). Checklist scores for experts and novices were significantly different: median score for novices was 28/52 (54%) while for experts 42/52 (81%) ($p<0.001$). Procedure time and fetoscopy time were significantly shorter ($p<0.001$) for experts. Residual anastomoses were found in 1/11 (9%) procedures performed by experts and in 9/14 (64%) performed by novices ($p=0.006$). Multivariate analysis showed that the checklist score independently from age and gender predicted competence.

CONCLUSIONS: The procedure-specific assessment tool for fetoscopic laser surgery shows a good inter-observer reliability and discriminates experts from novices. This instrument may therefore be a useful tool in the training curriculum for starting fetal surgeons. Further intervention studies with reassessment before and after training may increase the construct validity of the tool.

INTRODUCTION

Fetoscopic laser therapy is the preferred treatment modality for twin-twin transfusion syndrome (TTTS),[1-3] but is only offered in a few highly specialized Fetal Medicine centers around the world.[4] Although fetoscopic laser surgery is a complex procedure that has been in use for more than two decades, standardized surgical training programs for fetoscopic interventions are nonexistent and performance is often authority based, i.e. on personal experience, belief and individual preferences. Also, the learning curve is ill-defined, and varies between 21 to 75 cases (based different survival outcome measures such as minimal double survival rates of 54% or at least one survivor in 70% of cases) to acquire the necessary skills.[5-8] Therefore, there is a need for a reliable assessment tool of technical performance. Such a tool would be useful to monitor progress, provide constant feedback along the learning curve, to serve as an instrument for (re-)certification and offer standardized training.

We previously reported on a list of steps judged essential to the laser procedure based on the Delphi methods.[9] These steps were consensus based by a sample of international experts, making the final tool representative of international, rather than local practice. The aim of this prospective cohort study was to assess reliability and validity of this instrument in the context of simulated operating room performance. We hypothesized that, based on the systematic manner in which this tool was created; we would obtain an acceptable level of inter-observer reliability and that the instrument would discriminate the performance of experts from that of novices.

METHODS

Participants and study design

This study is part of the SILICONE project (**SI**mulator for **L**aser therapy and **I**dentification of **C**ritical steps of Operation: New Education program), conducted to develop a standardized training program for fetoscopic laser surgery for TTTS. In the first part of the project we determined the essential steps of treatment to develop an assessment instrument.[9] In the current part of the project, this instrument was validated using a silicone simulator involving the complete laser procedure.

This study was conducted in three Fetal Medicine centers: Leiden University Medical Center (the Netherlands), University Hospitals KU Leuven (Belgium) and Karolinska Institutet, Stockholm (Sweden) from September 2014 until December 2014. We recruited 24 volunteers with special interest in fetal therapy to participate in the study. All participants completed a questionnaire to establish baseline demographic characteristics and previous experience in fetoscopic surgery to measure potential confounding factors that affect performance. Participants were stratified into 3 groups with regard to the level of previous experience; expert or novice or intermediate.

An expert was defined as a physician who currently practices fetoscopic laser surgery for TTTS and has performed at least 25 fetoscopic laser procedures independently.[8] Novices included fetal medicine specialists without practical fetal therapy experience OR obstetricians attending a fellowship in perinatology OR senior residents with special interest in perinatology and minimal invasive therapy. All novices were experienced sonographers and had appropriate knowledge of TTTS and its treatment options, but had never performed a fetoscopic laser procedure and had little or no previous experience with other ultrasound-guided invasive procedures (amniocentesis, chorionic villus sampling and/or intrauterine transfusion). Practicing fetal surgeons that were still in their learning curve(e.g. performed between 1-25 fetoscopic laser procedures) were excluded.

For secondary analyses experts were categorized into 2 groups: intermediate expert level (performed < 50 procedures) and senior expert level (performed > 50 procedures).

Assessment

All participants (irrespective of the level of expertise) performed a similar assignment on the simulator. The scenario involved a patient of 17 weeks' gestation with stage 3 TTTS referred for laser therapy. The assignment included the complete fetoscopic laser procedure; starting from the moment the operation room is entered, until the surgery was finished and direct post-operative management was ordered. Three different items were scored: 'time', 'checklist with essential steps of procedure' and 'complete identification of vascular equator'.

All participants were evaluated by 2 independent observers (S.P. and J.A.), using the assessment instrument created by the Delphi consensus.[9] This list of essential steps was modified into a checklist adjusted to the simulated scenario. A detailed description of the instrument is available in the appendix. Each item was awarded 1 point if it was done properly (range 0-52). Procedure time, defined as 'the moment the surgeon enters the operating room until the moment that direct post-operative management is ordered' and fetoscopy time, defined as 'the moment the fetoscope is introduced for the first time until final removal' were recorded. A map of the placental architecture was used by the assessors to mark the coagulated anastomoses.

Simulated scenario

To explain the task, all participants were shown a standardized multimedia presentation outlining the background and aim of the study, as well as the performance metrics (time, missed essential steps and complete coagulation of the vascular equator). Finally, the context of the scenario (including patient characteristics, findings of diagnostic procedure and pre-surgical management) was presented.

Simulator characteristics

The simulator used for this study has previously been described[10] (Francis LeBouthillier, Surgical touch, Toronto, Canada), but was modified with a highly realistic silicone copy of a 17 week monochorionic twin placenta and twin fetuses (R. Bakker, Manimalworks, Rotterdam, The Netherlands). The silicone topping on the model mimics the abdominal wall. Inside there is a mimic of a uterus, which contains water and the placenta. The individual layers of the abdominal wall, the uterus and placenta have sonographic and compliance properties that mimic the clinical situation. The model allows an operator to practice ultrasound examination of a monochorionic pregnancy, required to select the best site for introduction of the instruments. The model also provides a realistic intrauterine environment, optimal to practice manual dexterity skills and to train navigation along the placental surface. Moreover, the addition of a "stuck" donor twin on the placenta simulates the inability to oversee the complete vascular equator. The addition of a "free-floating" recipient simulates a realistic complex situation of floating fetal extremities and umbilical cord in the recipients' sac. All necessary instruments (i.e. fetoscope, introduction set, endoscopy tower etc.) were used from the local Fetal Medicine center so that participants perform their tasks in a setting that was identical to what would be their clinical environment. Figure 1 shows a participant performing the procedure on the simulator model.

Statistical analysis

Demographics, procedure- and fetoscopy time, checklist score and presence of residual anastomoses were compared between experts and novices. Due to the small sample size and non-normality of the data, the Mann Whitney U test was used to test for differences between groups for the continuous variables. To test for differences between groups on non-ordinal categorical outcomes, such as presence or absence of experience, Fisher's exact test was used.

Spearman correlation coefficient was used to measure the inter-observer reliability. A correlation of 0.9 or higher was considered to be indicative of an excellent agreement.

We used a multivariate regression analysis to determine independent predictors for the construct validity of the instrument. Construct validity refers to the degree to which any measurement approach or instrument succeeds in describing or quantifying what it is designed to measure. Moreover, to evaluate the accuracy with which scores on a given instrument can classify groups that are already known to differ on a criterion measure (i.e. experts and novices). In other words, if experts are the ones with the construct (surgical skills) and the novices are the ones without the construct; construct validity determines whether the instrument identifies the presence or absences of the construct (surgical skills).

A two-sided p-value <0.05 was considered to indicate statistical significance. Statistical analyses were performed with IBM. SPSS version 21.0 (IBM SPSS Statistics for Windows, Version 21.0 Armonk, New York: IBM Corp.) Since no patients were involved, no formal ethical approval and written informed consent was needed for this study.

RESULTS

In this study, 24 fetoscopic simulated laser surgeries were analyzed. They were performed by 11

(46%) experts and 13 (54%) novices. Eleven participants were male, 13 were female. Although 4/13 (31%) of the novices in the study had previous limited experience with invasive obstetric procedures (e.g. amniocentesis, chorionic villus sampling, intrauterine transfusion etc.) none had previously performed the fetoscopic laser procedure for TTTS. In the group of experts, 5/11 (45%) had performed >100 procedures with a median of 10 procedures (range 8-20) annually. The demographics of the participants are shown in table 1.

Overall median procedure time was 40 minutes (range: 26-50 minutes). Experts were able to complete the procedure in 32 minutes, versus 43 minutes ($p=0.003$) by novices. Fetoscopy time was also significantly different between the groups. Median fetoscopy time for all participants was 17 minutes, (range: 10-27 minutes): 11 minutes for experts versus 20 minutes for novices ($p<0.001$). Residual anastomoses were found in 10/25 (40%) procedures, 1/11 (9%) performed by experts and in 9/14 (64%) performed by novices ($p=0.005$).

Secondary analyses were performed regarding level of expertise in the expert group comparing the results for intermediate and senior experts. Procedure time and fetoscopy time were not significantly different between the groups (32 minutes versus 31 minutes $p=0.776$ and 12 minutes versus 11 minutes, $p=0.376$), as well as surgical performance score 45/52 (87%) versus 49/52 (94%) $p=0.630$.

Reliability

The overall inter-observer reliability of the two raters' total scores (J.A. and S.P.) for the fetoscopic laser procedure was excellent (r_s): 0.974 *(*$p<0.001$) (Figure 2).

Agreement was less but still strong in the domains concerning 'direct post-operative management' (r_s: 0.722; $p<0.001$) and 'assessment during procedure' (r_s: 0.789; $p<0.001$) as displayed in table 2.

Agreement for the two raters remained high amongst intermediate experts (r_s): (0.866) and senior experts (r_s): (0.938).

The inter-observer variability did not significantly change over time (data not shown).

Construct validity

Rater 1's median score for novices on the assessment tool was 29/52 (56%) (range: 20-43), compared to an median expert score of 47/52 (90%) (range: 44-50) ($p<0.001$). Rater 2's median novice score similarly demonstrated statistically significant differences between novice and expert performance [30/52 (58%) (range: 19-45) versus 48/52 (92%) (range: 43-52)] ($p<0.001$).

The overall median checklist scores (combining the scores of the two raters) were 28/52 (54%) 20-44) in novices versus 42/52 (81%) (44-51) in experts ($p<0.001$) and were significantly associated with the presence of residual anastomoses as demonstrated in figure 3 ($p=0.002$). Sensitivity-specificity analysis showed an area under the curve of 0.861. Multivariate analyses showed that age (b_1: 0.203; $p= 0.351$) and gender (b_1: 0.088; $p=0.539$) of participants were not significantly associated with checklist scores and level of experience.

DISCUSSION

This study assessed the inter-observer reliability and construct validity of a procedure-specific evaluation tool for fetoscopic laser surgery of TTTS, created using the Delphi methodology.[9] Our instrument effectively distinguished performance of experts and novices with an acceptable level of inter-observer reliability.

Any discussion of evaluation or assessment must address issues of validity and reliability. The instrument will only be useful to educators or surgeons as a measure of competence when it does measure the construct that it intends to measure (validity) and when the results that are obtained are consistent and therefore meaningful (reliability). Inter-observer reliability refers to a degree to which difference in score on the tool reflects a difference in quality of performance rather than a difference between the raters. A high level of inter-observer reliability allows evaluation of skills by different observers and will be minimally affected by the variability of the rater.[11]

Till today, trainees in fetal surgery are educated according to the “master–apprentice” principle. Direct observation by experts alone may not be a reliable method of assessment and may lead to recall bias due to the retrospective nature of the evaluation. Use of fixed criteria such as a validated checklist by observing experts can address these concerns.[12, 13] Additionally, task-specific checklists provide trainees with detailed methods on how to perform the procedure and enable formative feedback and deliberate practice. To achieve standardization and wide implementation, an assessment tool must be reflective of practice among many institutions; therefore we included participants from three major Fetal Medicine centers.

Validation of assessment tools for training has been done frequently in other medical areas,[14-17] but never in the field of fetal therapy. Observation of surgical skills without structured criteria has poor reliability and will result in a low level of agreement among the raters.[18] The values for inter-observer reliability in this study indicate that our evaluation tool reaches the cut-off of 0.8 deemed acceptable for assessment.[11]

The purpose of this study was to validate the evaluation tool for surgeon's technical performance using a highly realistic simulator. Objective feedback to fetal surgeons on their performance based on highly reliable assessment tools could also be of great value for ongoing assessment and lifelong learning. Developing similar assessment tools for other invasive obstetric procedures will make it possible to teach and evaluate procedures using disseminated learning materials. Since we want to make the curriculum competency based, it is also important to define expert benchmark levels of proficiency for the final curriculum.

Procedure-specific checklists have been shown to be less reliable and less construct valid than global rating scales[19] However, a global assessment scale can make an instrument indistinctly and have an apparent precision, since items are rated on scales (e.g. 1-10) instead of 'achieved' or 'failed'. For feedback purposes it is sufficient to know at a glance which elements need improvement (instead of adding values to the assessed items).

Procedure time and fetoscopy time were significantly lower in the expert group compared to novices. This may be explained by the often interrupted flow of thoughts when performing a procedure for the first time. Surgical steps need to be carried out consciously for novices, as opposed to automatically for experts, making a procedure-specific tool even more valuable for training purposes and combines efficacy (closing all anastomoses) with safety (avoid complications).

A limitation of this study is that a few items identified through the prior Delphi consensus could not be analyzed during the simulator experiments since they take place in the diagnostic and pre-operative phase of the procedure. These steps include: "diagnostic procedure" (e.g. ultrasound examination at out-patient clinic confirming diagnosis and determine treatment options), "pre-surgical management" (e.g. prescription of procedure related medication etc.) and "follow-up ultrasound examination".

Therefore the construct validity and reliability measurement of this tool does not include these particular steps.

Due to nature of the procedure, we were unable to assess the validity of the instrument in surgery on real patients; therefore the simulator was used. Even though the simulator was regarded highly realistic, clinical features such as ‘tissue reaction after firing the laser’ and ‘complications such as bleeding’ could not be simulated. On the other hand, assessment using a simulator model can also be advantageous, since the lack of standardization in real patients makes consistent assessment of technical skills difficult. Advantages of the simulator model include the fact that tasks can be presented consistently to many trainees, who can operate independently, objective assessment by more than one faculty member is possible and there is no intrusion on operating room time, which has financial and ethical advantages.[20]

Quite often, even experienced operators work as a 'team' and this team may have experience where the sum is greater than the individual parts. This effect is hard to quantify and was not measured in this study. For this study, participants were assessed live in the operating room, therefore observers were able to oversee all steps, in contrast to only fetoscopic view or single camera position. This allowed us to evaluate the complete procedure, including all its facets such as sterility and handling of the instruments. Unfortunately, this element of our study prevented blinding the raters for the level of experience.

The construct validity of the instrument could be further assessed with a study with a pre- and post-training design. Correlation with a learning curve would further support its validity. Future studies should focus on the development and validation of a training curriculum aimed at improving the operative and technical skills of trainees in fetal therapy. Finally, additional studies should be performed to assess how well instructors can evaluate clinical skills when observing surgeons working with real patients and how to implement this into clinical practice.

ACKNOWLEDGEMENTS

We would like to thank all the participants from the MFM centers who generously shared their time, experience and materials for the purpose of this project.

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

FIGURES

Figure 1 Participant performing procedure on simulator for fetoscopic laser surgery

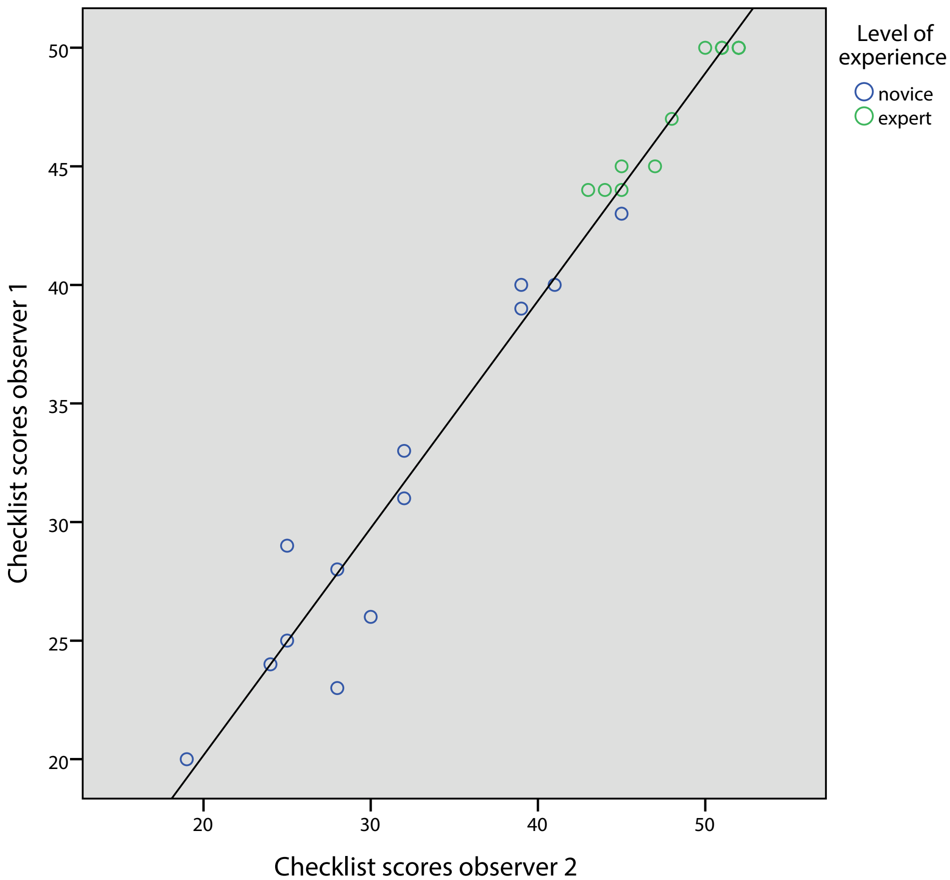

Figure 2 Correlation of checklist scores of the two observers

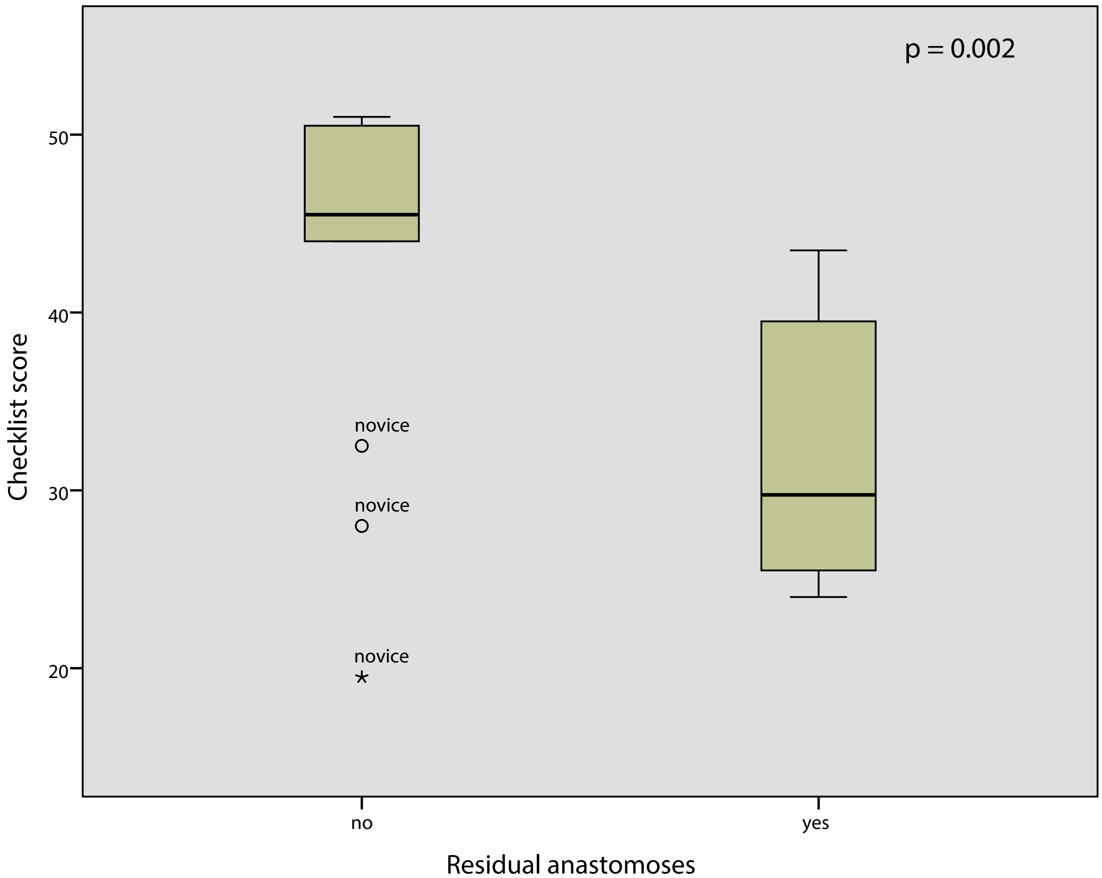

Figure 3 Checklist score and presence of residual anastomoses

APPENDIX

Appendix 1 Evaluation instrument

Disclosure of interests

None of the authors have a conflict of interest.

Contribution to authorship

All authors made substantial contributions to conception and design, and/or acquisition of data, and/or analysis and interpretation of data; participated in drafting the article or critical revising the content and gave final approval of the submitted version.

Details of ethics approval

The Medical Ethics committee of the Leiden University Medical Center declares that no formal ethical approval and written informed consent is needed for this study.

REFERENCES

1. Senat MV, Deprest J, Boulvain M, Paupe A, Winer N, Ville Y. Endoscopic laser surgery versus serial amnioreduction for severe twin-to-twin transfusion syndrome. *The New England journal of medicine*. 2004;351:136-144
2. Robyr R, Quarello E, Ville Y. Management of fetofetal transfusion syndrome. *Prenatal diagnosis*. 2005;25:786-795
3. Slaghekke F, Lopriore E, Lewi L, Middeldorp JM, van Zwet EW, Weingertner AS, Klumper FJ, Dekoninck P, Devlieger R, Lanna MM, Deprest J, Favre R, Oepkes D, Lopriore E.Fetoscopic laser coagulation of the vascular equator versus selective coagulation for twin-to-twin transfusion syndrome: An open-label randomised controlled trial. *Lancet*. 2014;383:2144-2151
4. Akkermans J, Peeters SH, Middeldorp JM, Klumper FJ, Lopriore E, Ryan G, Oepkes D. A world-wide survey on laser surgery for twin-twin transfusion syndrome. *Ultrasound in obstetrics & gynecology :* 2015 Feb;45(2):168-74.
5. De Lia JE, Kuhlmann RS, Lopez KP. Treating previable twin-twin transfusion syndrome with fetoscopic laser surgery: Outcomes following the learning curve. *Journal of perinatal medicine*. 1999;27:61-67
6. Hecher K, Diehl W, Zikulnig L, Vetter M, Hackeloer BJ. Endoscopic laser coagulation of placental anastomoses in 200 pregnancies with severe mid-trimester twin-to-twin transfusion syndrome. *European journal of obstetrics, gynecology, and reproductive biology*. 2000;92:135-139
7. Papanna R, Biau DJ, Mann LK, Johnson A, Moise KJ, Jr. Use of the learning curve-cumulative summation test for quantitative and individualized assessment of competency of a surgical procedure in obstetrics and gynecology: Fetoscopic laser ablation as a model. *American journal of obstetrics and gynecology*. 2011;204:218 e211-219
8. Peeters SH, Van Zwet EW, Oepkes D, Lopriore E, Klumper FJ, Middeldorp JM. Learning curve for fetoscopic laser surgery using cumulative sum analysis. *Acta obstetricia et gynecologica Scandinavica*. 2014 Jul;93(7):705-11
9. Peeters SH, Akkermans J, Westra M, Lopriore E, Middeldorp JM, Klumper FJ, Lewi L, Devlieger R, Deprest J, Kontopoulos EV, Quintero R, Chmait RH, Smoleniec JS, Otaño L, Oepkes D. Identification of essential steps in laser procedure for twin-to-twin transfusion syndrome using the delphi methodology: Silicone study. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*. 2015 Apr;45(4):439-46.
10. Pittini R, Oepkes D, Macrury K, Reznick R, Beyene J, Windrim R. Teaching invasive perinatal procedures: Assessment of a high fidelity simulator-based curriculum. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*. 2002;19:478-483
11. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: Rigorous science for the assessment of surgical education and training. *Surgical endoscopy*. 2003;17:1525-1529
12. Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. *American journal of surgery*. 2007;193:551-555; discussion 555
13. Hiemstra E, Kolkman W, Wolterbeek R, Trimbos B, Jansen FW. Value of an objective assessment tool in the operating room. *Canadian journal of surgery. Journal canadien de chirurgie*. 2011;54:116-122

14. Palter VN, Grantcharov TP. A prospective study demonstrating the reliability and validity of two procedure-specific evaluation tools to assess operative competence in laparoscopic colorectal surgery. *Surgical endoscopy*. 2012;26:2489-2503

15. McEvoy MD, Hand WR, Furse CM, Field LC, Clark CA, Moitra VK, Nietert PJ, O'Connor MF, Nunnally ME.. Validity and reliability assessment of detailed scoring checklists for use during perioperative emergency simulation training. *Simulation in healthcare : journal of the Society for Simulation in Healthcare*. 2014;9:295-30316. Alici F, Buerkle B, Tempfer CB. Objective structured assessment of technical skills (osats) evaluation of hysteroscopy training: A prospective study. *European journal of obstetrics, gynecology, and reproductive biology*. 2014;178:1-5

17. Tolsgaard MG, Ringsted C, Dreisler E, Norgaard LN, Petersen JH, Madsen ME, Freiesleben NL, Sørensen JL, Tabor A. Sustained effect of simulation-based ultrasound training on clinical performance: A randomized trial. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*. 2015

18. Reznick RK. Teaching and testing technical skills. *American journal of surgery*. 1993;165:358-361

19. Beard JD, Choksy S, Khan S. Assessment of operative competence during carotid endarterectomy. *The British journal of surgery*. 2007;94:726-730

20. Goff BA, Lentz GM, Lee D, Houmard B, Mandel LS. Development of an objective structured assessment of technical skills for obstetric and gynecology residents. *Obstetrics and gynecology*. 2000;96:146-150

Table 1. Demographics of study participants

Demographics	Expert n/N (%)	Novices n/N (%)	p value
Gender			
Male	8/11 (73)	3/13 (23)	0.015
Female	3/11 (27)	10/13 (77)	
Age			
(median in years, range)	52 (35-59)	32 (28-42)	<0.001
Experience with invasive obstetric procedures			
Has experience with invasive obst. procedures	11/11 (100)	4/13 (31)	0.001
years (median, range)	15 (7-23)	3 (1-8)	0.003
Type of invasive obstetric procedures			
Amniocentesis	11/11 (100)	3/13 (23)	
Chorionic villus sampling	11/11 (100)	3/13 (23)	
Intrauterine transfusion	8/11 (73)	1/13 (8)	
Fetal shunt placement	8/11 (73)	0	
Bipolar cord occlusion	11/11 (100)	0	
Open fetal surgery	4/11 (36)	0	
Other	4/11 (36)	0	
No. of FLS attended (incl. assisting or watching procedure)			
None	0	2/13 (15)	0.001
< 10 procedures	0	7/13 (54)	
10-25 procedures	0	0	
25-50 procedures	1/11 (9)	2/13 (15)	
50-100 procedures	1/11 (9)	0	
>100 procedures	9/11 (82)	2/13 (15)	
Experience with simulator training			
Never	2/11 (18)	1/13 (8)	0.447
A few times	4/11 (36)	8/13 (62)	
Regularly	5/11 (45)	4/13 (30)	

Table 2. Inter-observer reliability by domain

	Domain	No. of steps	R_s	p value
A	Preparation in operating room	7	0.956	<0.001
B	Ultrasound examination (together with sonographer)	7	0.862	<0.001
C	Pre-operative preparations	7	0.943	<0.001
D	Positioning and connection of instruments (pre-insertion)	6	0.977	<0.001
E	Insertion	5	0.947	<0.001
F	Orientation	8	0.857	<0.001
G	Laser coagulation	4	0.862	<0.001
H	Assessment during procedure	3	0.789	<0.001
I	Amniodrainage	2	1.000	<0.001
J	Closure	1	0.845	<0.001
K	Direct post-operative management	2	0.722	<0.001
	Overall	52	0.974	<0.001

R_s : Spearman correlation coefficient