Author: Peeters, Suzanne Hendrika Philomena
Title: Teaching and quality control in fetoscopic surgery
Issue Date: 2015-05-13

Chapter 1

Learning curve for fetoscopic laser surgery using cumulative sum analysis

Suzanne H. P. Peeters
Erik W. van Zwet
Dick Oepkes
Enrico Lopriore
Frans J. Klumper
Johanna M. Middeldorp

Acta Obstet Gynecol Scand. 2014 Jul; 93(7):705-11.

ABSTRACT

Objective. To identify a learning curve and monitor operator performance for fetoscopic laser surgery for twin-to-twin transfusion syndrome using cumulative sum analysis.

Design. Retrospective cohort study.

Setting. National tertiary referral center for invasive fetal therapy.

Population. A total of 340 consecutive monochorionic pregnancies with twin-to-twin transfusion syndrome treated with fetoscopic laser coagulation between August 2000 and December 2010.

Methods. A learning curve was generated using learning curve cumulative sum analysis and cumulative sum methodology to assess changes in double survival across the case sequence. Laser surgery was initially performed by two operators, joined by a third and fourth operator after 1 and 2 years, respectively.

Main outcome measures. Individual operator performance, double perinatal survival at 4 weeks.

Results. Overall survival of both twins occurred in 59% (201/340), median gestational age at birth was 32.0 weeks. Cumulative sum graphs showed that level of competence for double survival for the operators was reached after 26, 25, 26, and 35 procedures, respectively. Two operators kept their competence level and continued to improve after completing the initial learning process; two others went out of control at one point in time, according to the cumulative sum boundaries. A difference in learning effect was associated with number of procedures performed annually and previous experience with other ultrasound-guided invasive procedures.

Conclusions. This study shows that all operators reached a level of competence after at least 25 fetoscopic laser procedures and confirms the value of using the cumulative sum method both for learning curve assessment and for ongoing quality control.

INTRODUCTION

Mastering the skills necessary to perform surgical procedures with success, ease, and safety represents a learning curve for each surgeon. The learning curve itself ultimately represents the acquisition of competency, which affords an understanding of a new technique, technical modifications to the technique and improvements in support staff and perioperative care.

Learning curve patterns may vary depending on the type of surgical procedure, choice of outcome measurements[1], and from one surgeon to another.[2] Surgical training programs commonly prescribe a certain length of time or number of procedures performed to certify operators as proficient. A common question is: what is the actual number of procedures an individual operator has to perform to achieve satisfactory outcome results? The course of performance of an individual or group of surgeons can be plotted over time. In principle, novice surgeons are assumed to perform surgery less safely and efficiently than more experienced colleagues at the peak of their career.[3]

Although there is growing interest in tracking individual surgical outcome, traditional monitoring tools generally fail to consider the gradual learning process of starting surgeons. Cumulative sum (CUSUM) analysis is a graphical method for quality control to show changes in individual surgical performance. This technique is increasingly used as a management tool in medicine for competence monitoring.[4] The learning curve CUSUM (LC-CUSUM) test has been proposed to determine when a surgeon reaches a predefined level of performance while learning a new procedure.[5]

Twin-to-twin transfusion syndrome (TTTS) is one of the most common major complications of monochorionic twin pregnancies and carries a high risk of perinatal mortality and morbidity.[6] Fetoscopic laser coagulation is the treatment of choice [8,9] and with an increasing number of centers offering this procedure there is concern that, at least temporarily, less favorable outcomes will be seen because of learning curve effects. Identifying a learning curve for laser surgery is a logical step in the progression of incorporating this technique into the surgeons' armamentarium of procedures available in highly specialized Maternal–Fetal Medicine centers.[10]

We aim to use the CUSUM methodology to define the learning curve and monitor operator performance for fetoscopic laser coagulation.

MATERIAL AND METHODS

At the Leiden University Medical Center, the national referral center for invasive fetal therapy, a retrospective study was performed on prospectively collected data on all monochorionic twin pregnancies with TTTS treated by fetoscopic laser coagulation of vascular anastomoses between August 2000 and December 2010. Inclusion criteria for laser surgery were: monochorionic pregnancy, gestational age between 13 and 28 weeks, TTTS Quintero stage 1 with severe clinical symptoms of polyhydramnios, or TTTS Quintero stage ≥2. For this analysis, cases were not included if mothers were clinically in labor at time of diagnosis (n = 17). None of the other pregnancies were excluded once the fetoscope had been introduced into the amniotic cavity, even if laser coagulation was not possible. Details on the procedure were previously reported (11).

All surgeons eligible for training and performing fetoscopic laser surgery were experienced maternal–fetal medicine specialists with an academic career focusing on fetal medicine. Preconditions for training included extensive knowledge and experience in fetal medicine, expertise in diagnostic procedures including highest level of ultrasound, amniocentesis, chorionic villus sampling, and fetal blood sampling. All fetal surgeons were well-trained gynecologic laparoscopists.

Laser surgery was initially performed by two operators, joined by a third and fourth operator after 1 and 2 years, respectively. In the initial stage each procedure was attended by at least two operators; however, only one was performing the fetoscopy and laser surgery, others could observe and help to determine vascular anastomoses. Two procedures were performed by a foreign operator in the first year of laser therapy and excluded for learning curve analyses.

The technique that was used for the laser procedure underwent minor adjustments through the years, similar for all operators. Some of the women (n = 84) included in this study also participated in the randomized Solomon trial (www.trialregister.nl; NTR1245). The data on obstetric and neonatal outcomes were derived from medical charts. Demographic characteristics, details on the fetoscopic surgery, further interventions, and outcomes were prospectively collected in our fetal database. Pregnancy outcomes included survival up to 4 weeks after birth, intrauterine fetal demise and preterm delivery.

All surgeries were classified as failed or successful. Criteria for classifying a procedure as failed were: fetal loss of one or both twins and the occurrence of perinatal death within 28 days after the procedure. For this analysis, double perinatal survival, defined as survival up to 4 weeks of age, was chosen as a primary outcome. Individual prior experience with other ultrasound guided invasive procedures, such as intrauterine blood transfusion, chorionic villus sampling, and amniocentesis per operator was assessed. Individual performance with fetoscopic laser coagulation was analyzed by describing the number of procedures performed in consecutive years and the corresponding success rates.

Statistical analysis

For each surgeon, two types of CUSUM charts were constructed, to assess the learning curve and to monitor performance according to the number of successful procedures;[5,12] these are described in more detail in Appendix S1. A CUSUM score was computed from the successive outcomes, with successes yielding an increase in the score and failures yielding a decrease in score. In graphs, CUSUM scores are plotted on the y-axis against the consecutive number of procedures on the x-axis. The horizontal limit lines in CUSUM-graphs determine the spacing between unacceptable (H0) and acceptable (H1) boundaries, i.e. the null and alternative hypotheses. Once the score has reached a predefined level (L), the test rejects the null hypothesis in favor of the alternative and performance is deemed acceptable. In LC-CUSUM graphs the learning phase is completed when limit "L1" is crossed with accumulation of successive scores. CUSUM scores are then reset at 0 before starting the monitoring phase.

Control limits decide whether performance is unacceptable. In this study an "expert level" was calculated, to define acceptable and unacceptable boundaries as input for the CUSUM analysis, based on three of the largest recently published studies[6,13,14] and the only randomized controlled trial on fetoscopic laser therapy in TTTS.[9] Combining outcomes of these recent studies with survival rates of 67%,[13] 54%,[14] and 57%,[6] respectively, we calculated the "expert level" to be a mean survival of both twins at 4 weeks after delivery of 59% (H1: $p = 0.59$). The unacceptable boundary was set as 36% (H0: $p = 0.36$).[9]

The CUSUM score emits a signal in case of persistent deviation towards a deterioration or improvement in performance; by reaching either lower (H0) or upper (H1) limits. Conversely, the performance is assumed to be acceptable as long as the CUSUM score remained within the limits. The thresholds L1 (for competence) and L2 (for incompetence) are set differently. As it is likely that performance will be less successful during the learning phase, and to minimize the probability that performance of an operator who is not yet competent is noted as acceptable; L1 = 3.4. To quickly identify suboptimal performance of an operator who already completed the learning phase; L2 = 2.5.

Factors that may influence the outcome of the procedure such as gestational age, Quintero stage, placenta localization, and introduction technique were compared for all procedures performed by each operator by using chi-squared test and one-way analysis of variance. Taking these patients' risk factors into consideration, all CUSUM scores were calculated based on risk-adjusted case failure. Therefore operative failures in high-risk patients are not as visually significant on the CUSUM chart as they would be in low-risk patients. A p-value ≤ 0.05 was considered to indicate statistical significance.

Statistical analysis was performed with IBM SPSS Statistics for Windows version 21.0 (IBM Corp., Armonk, NY, USA). No ethics approval was required for this study.

RESULTS

In total, 340 monochorionic pregnancies treated with fetoscopic laser coagulation were included in this study. Characteristics and outcomes of all treated pregnancies sorted by procedures per operator are described in Table 1. Operator 1 performed significantly more combined open laparoscopy for anterior placenta in the starting years of fetoscopy.[15] Other possible confounding factors, such as Quintero stage and placenta localization, were not significantly different between all operators, nonetheless high-risk patients (i.e. advanced Quintero stage, anterior placenta) were identified to calculate adjusted-risk CUSUM scores. Secondary analysis of preoperative characteristics over the consecutive years showed no significant differences. Overall perinatal survival rates of at least one twin and double survival were 86% (294/340) and 59% (201/340), respectively.
Survival rate at birth, including triplet pregnancies, was 76% (528/690), equally divided among recipient twins in 72% (244/340) and donor twins in 75% (256/340). Neonatal death occurred in 21 neonates [from 19 pregnancies, 6% (19/340)]. Delivery occurred before 24 weeks in 11% (36/340), before 28 weeks in 18% (61/340), before 32 weeks in 41% (139/340), and before 34 weeks in 57% (194/340) of pregnancies. Ten triplet pregnancies were treated, all were dichorionic triamniotic triplets. In five triplets, all children survived at birth, in the five other cases single intrauterine fetal demise occurred. One former recipient died within 4 weeks after birth. Figure 1 shows the number of procedures performed per individual operator and fetal outcome using LC-CUSUM and CUSUM plots. The graphical data demonstrate a range of time over which competence was attained and kept. Operators 1 and 2, starting at almost the same time in 2000, had similar learning curves. The learning phase of operators 1 and 2 ended after procedure 26 and 25, respectively, demonstrated by crossing the L1 threshold. For operator 1 an alarm was set at the 95th procedure (red circle), indicating inadequate performance. The performance of operator 2 remained in between alarms and therefore performance was deemed adequate. Operator 3 finished the learning curve after 26 procedures; however, after setting stricter boundaries, performance became unacceptable after the 38th procedure because of accumulated failures. Performance improved directly hereafter, as shown by the dotted line in the plot. Operator 4, starting 2 years after the first operator was declared competent after 35 procedures and remained competent. In Table 2, number of procedures performed, individual success rates and previous experience with ultrasound-guided procedures are summarized.

	All pregnancies (n = 340)	Operator 1 (n = 110)	Operator 2 (n = 109)	Operator 3 (n = 72)	Operator 4 (n = 49)	*p*-value
Gestational age at laser surgery (week), median (range)	20+0 (13+4-29+4)	19+6 (13+4-29+4)	20+1 (14+3-27+4)	20+4 (14+4-29+0)	19+6 (14+5-25+6)	0.48
Gestational age at birth (weeks), median (range)	32+0 (15+6-41+2)	31+6 (18+0-40+0)	32+2 (15+0-41+2)	31+6 (16+4-40+6)	31+5 (17+0-39+1)	0.88
Preterm birth <24 weeks gestational age, n (%)	36 (11)	13 (12)	7 (6)	11 (15)	5 (10)	
Quintero stage I, n (%)	35 (10)	13 (12)	7 (6)	7 (9)	8 (16)	0.11
Quintero stage II, n (%)	109 (32)	43 (39)	31 (28)	22 (31)	13 (27)	
Quintero stage III, n (%)	181 (53)	46 (42)	68 (62	40 (56)	27 (55)	
Quintero stage IV, n (%)	15 (5)	8 (7)	3 (4)	3 (4)	1 (2)	
Placental localization posterior, n (%)	187 (55)	50 (45)	58 (53)	47 (65)	32 (65)	0.09
Placental localization anterior, n (%)	137 (40)	55 (50	44 (40)	22 (31)	16 (33)	
Placental localization lateral, n (%)	16 (5)	5 (5)	7 (7)	3 (4)	1 (2)	
Percutaneous insertion, n (%)	304 (89)	83 (75)	104 (95)	69 (96)	48 (98)	0.01
Laparotomy, n (%)	10 (3)	6 (6)	2 (2)	2 (3)	0 (0)	
COLFAP, n (%)	26 (8)	21 (19)	3 (3)	1 (1)	1 (2)	

Table 1: Characteristics of 340 pregnancies complicated by twin-to-twin transfusion syndrome and treated with laser surgery sorted by operator.
Legend: COLFAP, combined laparoscopy and fetoscopy in cases with completely anterior placenta.

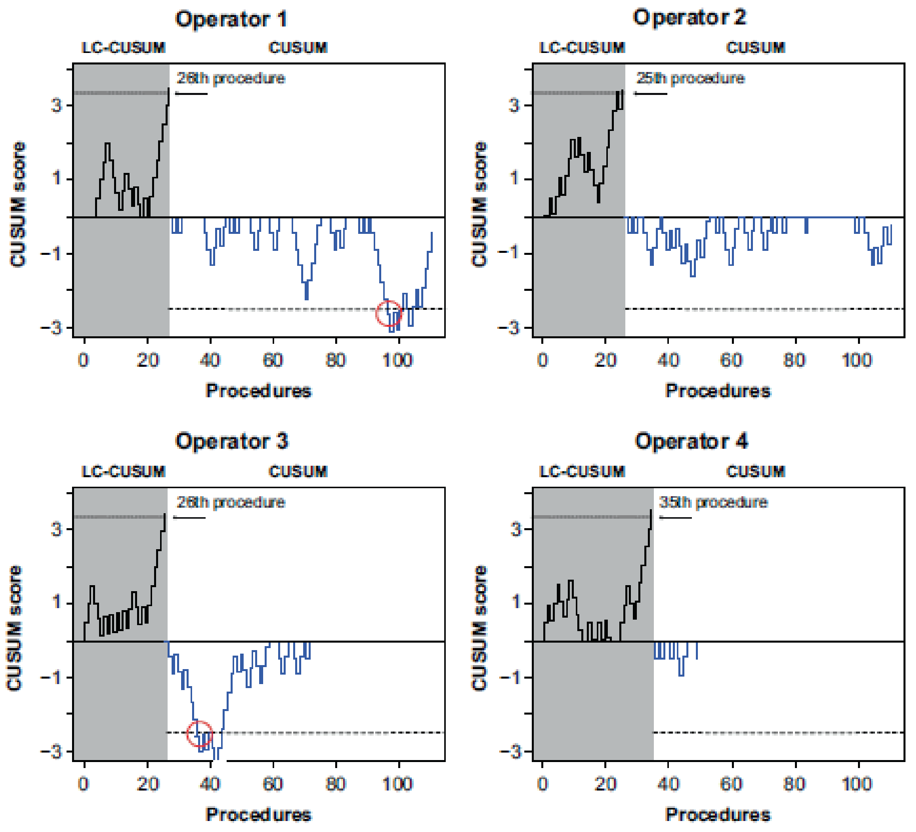

Figure 1
The individual learning curve cumulative sum (LC-CUSUM) and cumulative sum (CUSUM) plots for fetoscopic laser coagulation in twin-twin transfusion syndrome in four operators representing double survival at 4 weeks.
Legend: The consecutive number of performed procedures on the x-axis is plotted against the actual CUSUM value on the y-axis. LC-CUSUM and CUSUM scores are plotted in terms of the successive procedures, evaluating performance by double survival. LC-CUSUM is applied until acceptable performance has been reached and CUSUM is used thereafter to ensure that adequate level is maintained. For the LC-CUSUM (black line), as long as the score remains below the limit "L1" (dotted line), the operator is not considered proficient; when the LC-CUSUM score crosses this limit, the operator has completed the learning phase. For the CUSUM (blue line), as long as the score remains under the limit, the operator is considered to maintain an acceptable performance. The dotted blue lines represent the CUSUM score from operators that showed unacceptable performance as a result of a peak in failure rate (red circle) during their CUSUM monitoring phase.

Operator	Period	Procedures (n)	At least one survivor		Double survival		Annual procedures (mean number)	Previous experience (years)*
			n	%	n	%		
Operator 1	Aug 2000 to Dec 2010	110	94	85	66	60	11	9
Operator 2	Dec 2000 to Dec 2010	109	100	92	69	63	11	8
Operator 3	Jan 2002 to Dec 2010	72	58	81	40	56	8	3
Operator 4	Oct 2002 to Dec 2010	47	42	89	26	53	7	3
Total procedures	Aug 2000 to Dec 2010	340	294	86	201	59	34	na

Table 2 Operator's success and experience.
Legend: *Determined as experience with other ultrasound-guided invasive procedures (intraperitoneal transfusion, chorionic villus sampling and/or amniocentesis).

DISCUSSION

Our study shows an increase in survival rates with growing operator experience, using an LC-CUSUM method to assess individual learning curves. In addition, the CUSUM analysis for continuous monitoring of individual performance proved both feasible and highly insightful. The number of procedures required to reach an adequate level of performance ranged between 26 and 35 in our four operators. The minimal individual variation in learning profiles may be explained by the "group" learning effect, by influence of general expertise and exchange of experience.

Two operators showed a peak in failure rate during their CUSUM monitoring phase. Such an event may have multiple explanations, which cannot be inferred from the CUSUM analysis but requires in depth analysis of the procedures itself, such as difference in technical details with other operators, and particular case mix during this period. This is an illustrative example of the use of CUSUM plots for monitoring of performance among experienced operators, a practice which these operators continued beyond the study period. In particular, CUSUM enables almost real-time evaluation, in contrast to other less sophisticated methods, such as annual or 3-monthly reports. The use of CUSUM prevents potentially hazardous delay in taking action to improve outcome. In the case of real-time assessment one should consider monitoring and coaching these operators according to the LC-CUSUM standards, until they again reach acceptable levels of performance. However, awareness of underperformance alone may already improve outcomes.

Due to the retrospective design of this study, ongoing CUSUM scores were plotted and rapid recovery was demonstrated within a short time. The difference in learning curves between all operators was probably the result of operator 4. Operator 4 reached a level of competence after 35 procedures. The slow initial accrual of cases by operator 4 could certainly have contributed to the upward trend of the CUSUM curve for these first 35 procedures. Operators 1 and 2 performed approximately 12 procedures per year, whereas operator 4 started off performing only four cases annually, although often supervised by an experienced colleague. Compared with the other operators, also previous experience with other ultrasound-guided invasive procedures was considerably different for operator 4. During analysis of preliminary results of this study in 2010 we already noted this effect and assured that surgeries were more equally divided among operators by making logistical changes in our clinic. Equal division of the number of procedures performed by each operator annually should be taken into account in case of learning curve assessment of rare procedures such as laser surgery in TTTS.

To optimize the learning curve in our center, in the initial phase each procedure was attended by at least two operators. In addition, all treated cases were discussed monthly in a multidisciplinary setting, with exchanging of ideas and suggestions for improvement of technique and prevention of complications. Another most helpful tool, in our view, was the systematic evaluation of each treated placenta through careful placental injection of colored dye.[16]

The learning curve in our series represents the improvement of both the operators, from experience and practice, and the performance of the entire team at managing pregnancies involving TTTS. Teamwork, discussion (including international audits),[17] stimulation, controllability, and continuity may be beneficial factors. Previous authors have shared their opinion on a learning curve for fetoscopic laser coagulation in TTTS. Julian De Lia, the pioneer of laser surgery in TTTS, describes that the end of his learning curve was reached after 33 procedures, although without further explanation.[18] Hecher et al. reported a learning curve of 75 procedures and found a significant increase in perinatal survival over time.[19] Defining a learning curve with a predefined number of cases fails to take into account that individual operators may not achieve proficiency with a fixed sample size. Papanna et al. used similar LC-CUSUM and CUSUM analysis to determine the learning curve for three operators, reaching a level of competence after 60, 21, and 21 procedures, respectively.[20] However, this study did not include risk adjustment for high-risk patients, which may highly influence the slope of the CUSUM plots. Results of our center with four operators showed similar pregnancy outcomes as published series with one operator[19] and the presumed disadvantage of dilution of procedures does not seem to exist in our unit, likely due to the team approach.

Case selection, by either treating predominantly high risk or low-risk cases during the learning phase, may bias learning curve results. This effect should be taken into account during assessment of competence. Setting appropriate thresholds for acceptable and unacceptable performance is difficult. In this study we used literature-based expert levels, but an alternative may include historical data from the own unit with advantage of having comparable case-mix. Likewise, the choice of design parameters for CUSUM is critical to its performance, other surgical parameters such as operating time or presence of residual anastomoses may be useful endpoints to classify the operation as successful or failed. One of the limitations in this study is that the effect on an individual learning curve by assisting another operator was not measured. This however may have a positive influence on individual expertise levels. This effect was described by Kolkman et al., who showed that mentor traineeship can accelerate the learning curve of advanced laparoscopic procedures.[21] The other attributable effect that was not measured included the extent of the operators' experience with obstetric ultrasound, invasive fetal diagnostic and therapeutic procedures, and endoscopy prior to starting laser therapy. These skills probably contributed to a steeper individual learning curve.

In summary, this study evaluated the learning curve for fetoscopic laser coagulation as an example of minimal invasive intrauterine treatment. CUSUM analysis is well suited to the assessment of procedures with a binary outcome, but accurate and appropriate standards of practice must be determined before assessment to ensure the correct identification of underperformance. A prospective study would be able to evaluate the value of the CUSUM technique as a continuous audit system, allowing urgent real-time feedback to improve the quality of surgery. Determination of an accurate learning curve, as well as evaluation of individual surgeons, will be of great value with relevance to other procedures to decrease medical error and substandard performance, and to improve quality.

ACKNOWLEDGEMENTS

We gratefully acknowledge Dr. I.L. van Kamp and Dr. I.T. Lindenburg, Department of Obstetrics and Dr. R. Wolterbeek, Department of Biostatistics at the Leiden University Medical Center for their contributions to the statistical analysis.

REFERENCES

1. Carty MJ, Chan R, Huckman R, Snow D, Orgill DP. A detailed analysis of the reduction mammaplasty learning curve: a statistical process model for approaching surgical performance improvement. Plast Reconstr Surg. 2009;124:706–14.
2. Lindenburg IT, Wolterbeek R, Oepkes D, Klumper FJ, Vandenbussche FP, van Kamp IL. Quality control for intravascular intrauterine transfusion using cumulative sum (CUSUM) analysis for the monitoring of individual performance. Fetal Diagn Ther. 2011;29:307–14.
3. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. Ann Intern Med. 2005;142:260–73.
4. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. Qual Saf Health Care. 2007;16:203–7.
5. Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. Br J Surg. 2008;95:925–9.
6. Baud D, Windrim R, Keunen J, Kelly EN, Shah P, Van Mieghem MT, et al. Fetoscopic laser therapy for twin–twin transfusion syndrome before 17 and after 26 weeks' gestation. Am J Obstet Gynecol. 2013;208:197.
7. van Klink JM, Koopman HM, Oepkes D, Walther FJ, Lopriore E. Long-term neurodevelopmental outcome in monochorionic twins after fetal therapy. Early Hum Dev. 2011;87:601–6.
8. Crombleholme TM, Shera D, Lee H, Johnson M, D'Alton M, Porter F, et al. A prospective, randomized, multicenter trial of amnioreduction vs selective fetoscopic laser photocoagulation for the treatment of severe twin–twin transfusion syndrome. Am J Obstet Gynecol. 2007;197:396–9.
9. 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. N Engl J Med. 2004;351:136–44.
10. Lewi L, Devlieger R, De CL, Deprest J. Twin–twin transfusion syndrome: the good news is; there is still room for improvement. Acta Obstet Gynecol Scand. 2012;91:1131–3.
11. Middeldorp JM, Sueters M, Lopriore E, Klumper FJ, Oepkes D, Devlieger R, et al. Fetoscopic laser surgery in 100 pregnancies with severe twin-to-twin transfusion syndrome in the Netherlands. Fetal Diagn Ther. 2007;22:190–4.
12. Biau DJ, Porcher R, Salomon LJ. CUSUM: a tool for ongoing assessment of performance. Ultrasound Obstet Gynecol. 2008;31:252–5.
13. Chmait RH, Kontopoulos EV, Korst LM, Llanes A, Petisco I, Quintero RA. Stage-based outcomes of 682 consecutive cases of twin–twin transfusion syndrome treated with laser surgery: the USFetus experience. Am J Obstet Gynecol. 2011;204:393–6.
14. Valsky DV, Eixarch E, Martinez-Crespo JM, Acosta ER, Lewi L, Deprest J, et al. Fetoscopic laser surgery for twin-to-twin transfusion syndrome after 26 weeks of gestation. Fetal Diagn Ther. 2012;31:30–4.
15. Middeldorp JM, Lopriore E, Sueters M, Jansen FW, Ringers J, Klumper FJ, et al. Laparoscopically guided uterine entry for fetoscopy in twin-to-twin transfusion syndrome with completely anterior placenta: a novel technique. Fetal Diagn Ther. 2007;22:409–15.
16. Lopriore E, Slaghekke F, Middeldorp JM, Klumper FJ, van Lith JM, Walther FJ, et al. Accurate and simple evaluation of vascular anastomoses in monochorionic placenta using colored dye. J Vis Exp. 2011;55:e3208.
17. Lindgren P, Westgren M. Twin–twin transfusion syndrome and fetal medicine centers. Acta Obstet Gynecol Scand. 2013;92:362.
18. De Lia JE, Kuhlmann RS, Lopez KP. Treating previable twin–twin transfusion syndrome with fetoscopic laser surgery: outcomes following the learning curve. J Perinat Med. 1999;27:61–7.
19. 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. Eur J Obstet Gynecol Reprod Biol. 2000;92:135–9.
20. 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. Am J Obstet Gynecol. 2011;204:218–9.
21. Kolkman W, Engels LE, Smeets MJ, Jansen FW. Teach the teachers: an observational study on mentor traineeship in gynecological laparoscopic surgery. Gynecol Obstet Invest. 2007;64:1–7.

APPENDIX 1.

CUSUM methodology

The cumulative sum (CUSUM) methodology is a graphical method to assess changes in individual surgical performance (1). CUSUM sequentially tests the null hypothesis that the process is in control, i.e. its mean is equal to a given target. Thus, it detects when the process changes to an out of control state. The learning curve CUSUM (LC-CUSUM) was developed based on two one-sided test procedures where the null hypothesis is that the process is out of control (2).

We introduced a null- and alternative hypothesis, but it should be stressed that a CUSUM analysis itself is not a hypothesis test. Irrespective of the level of expertise of the trainee, his or her CUSUM will eventually cross the threshold. Hence, the null hypothesis is always rejected and the probability of a type I error (α) is 1. Consequently, the probability of a type II error (β) is 0 (3). For this reason, the performance of a CUSUM procedure must be quantified differently.

Biau et al. recommended using simulation of average run lengths (ARL) (4). A simulated cohort of surgeons is used in which each surgeon was assumed to have a probability of making an error in a single case. The cumulated score for each case in the series was calculated until it crossed the terminating barrier L1, or until the maximum number of cases was reached. The run length is therefore defined as the number of cases until threshold L1 is crossed. This is a random variable, whose distribution depends on the skill of the surgeon. The ARL under H0 and H1 is sometimes used to quantify the performance of the CUSUM procedure. However, since the distribution of the run length is highly skewed, we prefer the Median Run Length (MRL).

Ideally, we would like the MRL to be very short under H1 (if operator is underachieving this will easily be detected), and very long under H0 (since it is likely that performance will be less successful while in the learning phase, if MRL under H0 is long, the probability of declaring a surgeon competent who in fact is not becomes small). By raising or lowering the threshold L1 we can change the MRL, but increasing the MRL under H0 means also increasing it under H1.

We used a Monte Carlo simulation (10,000 runs) to determine the MRL for our set of parameters. We found:

MRL for competence under H0: (p=0.36) 263
MRL for competence under H1: (p=0.59) 23

When a surgeon has been declared "competent" everything is turned around and (starting from zero) a CUSUM is constructed for:

H0: p=0.59 and H1: p=0.36

Now we continue on-going monitoring of performance until the CUSUM crosses a pre-defined limit L2=2.5 When that happens, the performance is discussed with the surgeon and possibly remedial actions are taken. Again, we computed the MRL:

MRL for incompetence under H0: (p=0.36) 89
MRL for incompetence under H1: (p=0.59) 15

We can turn a CUSUM into a hypothesis test, by choosing a fixed number 'n' and rejecting the null hypothesis if the threshold is crossed before the n-th case. In our case, we choose n=30, since this is a generally accepted number of procedures after which the learning curve should be completed (5;6). For the learning phase we found the probability of a type I error (α) is 0.05 and the power (1- β) 0.72. Therefore according to this analysis the probability to reject the hypothesis that performance is unacceptable, when performance is indeed unacceptable is low and the probability to reject the hypothesis that performance is unacceptable when in fact it is acceptable is high.

For the second phase, we found that the probability of a type I error is 0.17 and the power 0.87.

LC-CUSUM holds the feature that a barrier at 0 cannot be crossed and the score remains at 0 if the operator accumulates successive failures. In this way a starting surgeon will not have to compensate unnecessarily for the accumulated failures when starting a procedure. Since the performance of the surgeon is out of control at the beginning, an upper limit indicating this inadequate procedure is unnecessary. After an operator crossed boundary 'h' he or she has reached a level of competency and further performance is monitored with a CUSUM test.

REFERENCES

1. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R. Quality control of surgical and interventional procedures: a review of the CUSUM. Qual Saf Health Care 2007 Jun;16(3):203-7.
2. Biau DJ, 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 2010 Aug 15;29(18):1900-9.
3. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. Stat Methods Med Res 2003 Mar;12(2):147-70.
4. Biau DJ, Williams SM, Schlup MM, Nizard RS, Porcher R. Quantitative and individualized assessment of the learning curve using LC-CUSUM. Br J Surg 2008 Jul;95(7):925-9.
5. 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. Am J Obstet Gynecol 2011 Mar;204(3):218-9.
7. De Lia JE, Kuhlmann RS, Lopez KP. Treating previable twin-twin transfusion syndrome with fetoscopic laser surgery: outcomes following the learning curve. J Perinat Med 1999;27(1):61-7.