Original Article

# Characteristics of the Learning Curve for Cesarean Section: A Retrospective Study

Takayuki Rikitake[1,4] Eijiro Hayata[3,4] Mineto Morita[3,4]
Masahiko Nakata[3,4] Kunichika Matsumoto[1,2] and Tomonori Hasegawa[1,2]*

[1]Department of Social Medicine, Toho University Graduate School of Medicine, Tokyo, Japan
[2]Department of Social Medicine, Toho University School of Medicine, Tokyo, Japan
[3]Department of Obstetrics and Gynecology, Toho University School of Medicine, Tokyo, Japan
[4]Department of Obstetrics and Gynecology, Toho University Omori Medical Center, Tokyo, Japan

**ABSTRACT**

***Introduction***: Indicators that can objectively evaluate the proficiency of obstetrics and gynecology trainees in the Cesarean section would be useful from a medical education perspective. We conducted a case-control study on Cesarean sections performed by trainees and specialists using 5 years of data from our hospital.

***Methods***: The cases involved 1,055 Cesarean deliveries performed at Toho University Omori Medical Center over 5 years from 2012 to 2017, including 745 singleton, low-risk, and low-urgency cases (259 cases performed by 15 trainees vs. 486 cases performed by 21 specialists), and examined the operation time, Apgar score, umbilical artery blood pH (UmA pH), and surgical blood loss. Furthermore, the operation time was divided into (a) time from the start of the operation to the delivery of the baby, (b) time from the delivery of the baby to the end of the operation, and (c) total operation time.

***Results***: No significant differences were observed in the Apgar score, UmA pH, or blood loss between the trainees and specialists except in operation time. However, the significant difference in operation time disappeared after 16 procedures had been performed. In the time series analysis, inflection points were observed at the 8th, 13th, and 18th surgical experiences.

***Conclusions***: The operation time can be used to objectively assess the proficiency of trainees. Trainees are required to perform ≥20 surgical cases to obtain stable skills. The learning process for the Cesarean section shows a sigmoid curve composed of the imitation, trial-and-error, improvement, and stable periods.

**Toho J Med 9 (1): 20–28, 2023**

**KEYWORDS: epidemiologic studies, Cesarean section, learning curve, medical education**

## Introduction

The Cesarean section is an essential surgical technique that must be mastered by obstetrics and gynecology trainees. Trainees must experience a certain number of surgical cases at designated education hospitals before they can be deemed proficient. Although improvement in the surgical technique should be evaluated based on objective indi-

*Corresponding Author: Tomonori Hasegawa, 5-21-16 Omori-nishi, Ota, Tokyo 143-8540, Japan, tel: +81−3−3762−4151
e-mail: health@med.toho-u.ac.jp
DOI: 10.14994/tohojmed.2022-018

Received Aug. 5, 2022: Accepted Oct. 25, 2022
Toho Journal of Medicine 9 (1), Mar. 1, 2023.
ISSN 2189−1990, CODEN: TJMOA2

cators,[1–5] it has conventionally been evaluated by a traditional teaching method called "see one, do one, teach one" without objective evaluations.[6] In recent years, it is advocated to create a safe environment in which a resident's skills can be rigorously assessed without exposing the patient to risk in the operating room.[7–11] With the laparoscopic and robot-assisted surgical approaches spreading, the objective structured assessment (OSATS) using virtual reality simulators and video assessment has been studied.[12–25] More recently, Observational Teamwork Assessment for Surgery, which evaluates nontechnical skills such as teamwork, leadership, and communication, is being developed from the perspective of preventing adverse events in surgery.[26, 27] From the viewpoint of patient safety, the improvement of the surgical skills of trainees should be objectively and appropriately evaluated.

Several large-scale reports of Cesarean section learning curves have been published overseas.[1–5] However, no studies have been published of the same size as those performed overseas on proficiency evaluations among Japanese obstetrics and gynecology trainees before obtaining specialist qualification. Overseas studies have examined the surgical time, Apgar score, umbilical artery blood pH (UmA pH), blood loss, postoperative complications, and length of hospital stay as indicators of technical proficiency. However, the evaluation items that could most appropriately appraise trainees' skills have not been established.[1–5]

This study established the most appropriate indicators for evaluating technical proficiency in the Cesarean section among trainee doctors at a designated education hospital in Japan.

## Methods

### Study design and ethical statement

We designed a retrospective case-control study conducted at a single facility (Toho University Omori Medical Center, Tokyo, Japan). The study protocol was approved by the Ethics Committee of Toho University Omori Medical Center (No. M17226).

### Participants, classification, and data collection

We reviewed the medical records of cases delivered after 37 weeks' gestation from April 2012 to March 2017 at Toho University Medical Center Omori Hospital (n = 4,924). We then extracted cases involving singleton, low-risk pregnancies delivered by Cesarean section due to a less urgent indication. The exclusion criteria were (1) vaginal delivery, (2) multiple pregnancies, (3) high-risk pregnancy (placenta previa, severe pregnancy hypertension, etc.), and (4) Cesarean section performed for highly urgent indications (fetal distress, HELLP ＜Hemolysis, Elevated Liver enzymes and Low Platelets＞ syndrome, placental abruption, uterine rupture, umbilical cord prolapse, etc.). The cases were divided into two groups: a trainee group in which the trainees operated as the first surgeon and a specialist group in which the specialists operated as the first surgeon. A trainee was defined as "a doctor who has completed the junior resident stages and has newly started obstetrics and gynecology training," and a specialist was defined as a "doctor who has obtained the obstetrics and gynecology specialist qualification of the Japan Society of Obstetrics and Gynecology." The extracted cases were given new case numbers, and an anonymous database was created. In the event of data loss, we referenced the delivery records and electronic medical records. Cases with obvious data loss were further excluded.

### Surgical procedures

The surgical procedure at our hospital is standardized. The trainee observes the specialist's surgeries approximately 10-20 times before his or her first operation, learns the surgical procedure of the specialist, and performs his or her first operation applying the specialist's procedure under the specialist's guidance. Therefore, both trainees and specialists perform the same procedure. Until the trainee becomes a specialist, he or she always performs surgery under the guidance of a specialist.

The standard surgical procedure is performed as follows: The abdomen is opened using the Pfannenstiel incision method in the first Cesarean section or using the same incision method as the previous Cesarean section in the second and subsequent Cesarean sections. Incision of the myometrium is performed at the lower uterine transection, and a uterine suture is performed with two layers (first layer: nodular suture; second layer: continuous suture). The abdominal wall is then sutured with three layers (peritoneum, fascia, and subcutaneous tissue), and finally, the dermis is buried and sutured, and the surgical procedure is completed.

### Endpoint

The baseline characteristics collected included maternal age, gestational age, and newborn birth weight. We set the following four items as indicators for objectively evaluating technical proficiency: (1) operation time, (2) Apgar Score at 1 and 5 min, (3) UmA pH, and (4) blood loss. The operation time was divided into three sections: (a) the time from the start of the operation (first incision) to the delivery of the baby (incision-delivery time, I-D time); (b) the time from the delivery of the baby to the end of the operation (delivery-suture time, D-S time); and (c) the total operation time (incision-suture time, I-S time). I-D and (b) D-S times are continuous. Thus,

$$I\text{-}D\ time + D\text{-}S\ time = I\text{-}S\ time \quad (1)$$

At Toho University Hospital, patients are to be hospitalized for 7 days after Cesarean section as per the clinical pathway; thus, the duration of hospitalization after surgery was not included among the evaluation items.

We assumed that items with significant differences could be used as proficiency evaluation indicators and thus drew learning curves for each item. First, the surgical cases of the trainees were numbered in order of surgical operation and divided into groups of five (1-5, 6-10, 11-15, etc.). Using this grouping, a significant difference test was performed between the trainee and specialist groups to examine at which grouping the significant difference between the two groups disappears. Next, we calculated the mean time of each case grouping and created a learning curve plot. Performing a smoothing operation using the fifth-order moving average method, we illustrated the trends in the mean times.

A time series analysis was performed using the autocorrelation coefficient for the learning curve created. This analysis can help predict trends and future changes in the data. For example, a shortening trend in the time required for surgery as experience increases may suggest an inflection point somewhere. We therefore treated the mean data of the trainee operation time as time series data. In the time series analysis, a regression model was constructed using an autocorrelation coefficient using the autoregressive (AR) and moving average (MA) models. The AR model is used to regress to a previous value at time point "p" with respect to the stationary portion, whereas the MA model is used to regress to a previous value at time point "q" with respect to the nonstationary portion. The general form combining the AR(p) and the MA(q) models is called the ARMA(p, q) model, which allows the creation of a model that can flexibly explain stationary time series data. Since the ARMA model cannot be applied to nonstationary time series data as it is, the ARMA (p, q) model is applied to the stationary series by first taking the difference "d" times. This is called the ARIMA (p, d, q) model, with which a time series model can be constructed even for nonstationary models. The Box-Jenkins method was used to determine the optimal order of the variables "p," "d," and "q" in applying the ARIMA (p, d, q) model to the time series data of the trainee operation times. Akaike's Information Criterion was used to select the model that could best determine the optimal order of "p," "d," and "q." The ARIMA (p, d, q) model was applied to I-D, D-S, and I-S times.

We also investigated the existence of a statistically significant inflection point using join point analysis for the time series data explained by the ARIMA model. Join point analysis uses a nonlinear regression model developed by the Surveillance, Epidemiology, and End Results (SEER) of the United States National Cancer Institute to capture changes in morbidity and mortality over time. This change over time is called the annual percent change (APC), and the inflection point at which this APC significantly increases or decreases is called the join point.

**Statistical analyses**

For the baseline characteristics, the Mann-Whitney *U*-test was used to test whether or not two groups were likely to have been derived from the same population. For the evaluation items, such as the operation time, Apgar Score at 1 and 5 min, UmA pH, and surgical blood loss, Student's *t*-test was used to evaluate the differences between the two groups. A multiple comparison test using a one-way analysis of variance was performed to compare the operation time between the two groups. $^{*}p < 0.05$; $^{**}p < 0.01$; N.S, not significant. The JMP® ver. 13.2.1 software program by SAS Institute (Cary, NC, USA), was used for the above statistical analysis, and Microsoft Excel 2016 (Redmond, WA, USA) was used for graph creation. For the time series analysis, XLSTAT ver. 2018.1.49955 by Addinsoft (New York, NY, USA) was used for the fifth-order MA calculation and creation of the ARIMA model with the Box-Jenkins method. Joinpoint Software ver. 3.3 by SEER* Stat of National Institutes of Health (NIH) was used to analyze the inflection points of the time series data.

## Results

During the 5-year period from April 2012 to March 2017, of the 1,055 deliveries via Cesarean section at the hospital, 745 met the inclusion criteria. Of these, 259 cases were delivered by trainees, whereas 486 were by specialists (Fig. 1). The number of trainees and specialists was 18 and 21, respectively.

**The most appropriate indicators for evaluating technical proficiency**

The baseline characteristics did not differ significantly between the two groups (Table 1). Table 2 shows the indications that led to a Cesarean section. There were no differences in the conditions of basic characteristics and surgical indications for the Cesarean section between the specialist group and the trainee group. Among the evaluation results for the four proficiency evaluation indicators, only the operation time differed between the two groups: the I-D, D-S, and I-S times (Table 3). Regarding the I-D time, the mean ± standard deviation time of the trainee group was 9.5 ± 4.1 min, whereas that of the specialist group was 7.6 ± 4.3 min (mean between-group difference, 1.9 min; 95% confidence interval [CI], 1.2 to 2.5 min; $p < 0.01$). Regarding the D-S time, the mean time of the trainee group was 60.7 ± 14.2 min, whereas that of the specialist group was 50.2 ± 14.4 min (10.5 min; 95% CI, 8.3 to 12.6 g; $p < 0.01$). Regarding the I-S time, the mean time of the trainee group was 70.1 ± 16.5 min, whereas that of the specialist group was 57.8 ± 16.6 min (12.3 min; 95% CI, 9.8 to 14.8 min; $p < 0.01$). However, the Apgar Scores at 1 and 5 min, UmA pH, and blood loss did not vary significantly between the trainee and specialist groups: Apgar score 1 min: 8.09 ± 0.99 vs.

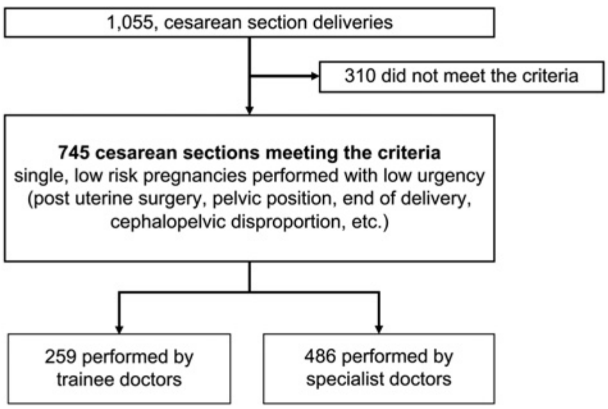


Fig. 1 Patient flow chart for the current analysis.

Table 1 Baseline characteristics of the cases under the trainee and specialist groups

| Characteristic | Trainee group (n = 259) | Specialist group (n = 486) |
|---|---|---|
| No. of surgeons | 18 | 21 |
| Mean number of Cesarean sections performed per surgeon | 14.4 | 23.1 |
| Minimum | 2 | 5 |
| Maximum | 35 | 89 |
| Maternal age | | |
| Mean | 34.9 ± 5.0 | 34.8 ± 4.9 |
| Range | 22–46 | 17–45 |
| <30 years | 48 (14.8) | 41 (15.6) |
| 30 to <40 years | 59 (68.1) | 167 (64.5) |
| ≥40 years | 93 (17.0) | 51 (19.7) |
| Gestational age (weeks) | | |
| Mean | 37.9 ± 1.0 * | 37.9 ± 1.0 * |
| <37 | 0 | 0 |
| 37 to <40 | 235 (90.7) | 427 (90.3) |
| ≥40 | 24 (9.2) | 47 (9.7) |
| Weight of newborn (g) | | |
| Mean | 2,922 ± 387 | 2,916 ± 417 |
| Range | 1,941–4,648 | 1,536–4,468 |
| <1,500 | 0 | 0 |
| 1,500 to <2,000 | 1 (3.9) | 4 (0.8) |
| 2,000 to <2,500 | 30 (11.6) | 54 (11.1) |
| 2,500 to <3,000 | 123 (47.5) | 232 (47.7) |
| 3,000 to <3,500 | 91 (35.1) | 155 (31.9) |
| ≥3,500 | 14 (5.4) | 41 (8.5) |

Values are expressed as mean ± standard deviation or number (percentage, %). Mann-Whitney *U*-tests were used to assess whether the two groups were derived from the same population. No significant differences ($p<0.05$) in any of the baseline characteristics were found between the two groups.

Table 2 Indications for Cesarean section

| | Trainee group (n = 259) | Specialist group (n = 486) |
|---|---|---|
| Previous Cesarean section | 110 (42.5) | 213 (43.8) |
| Fetal abnormal position † | 56 (21.6) | 106 (21.8) |
| Pregnancy after uterine surgery ‡ | 34 (13.1) | 56 (11.5) |
| Prolonged and arrested labor | 26 (10) | 58 (11.9) |
| Maternal complications § | 17 (6.6) | 32 (6.6) |
| Cephalopelvic disproportion (CPD) | 9 (3.5) | 11 (2.3) |
| Fetal malformation ¶ | 4 (1.5) | 9 (1.9) |
| Chorionic meningitis | 3 (1.2) | 1 (0.2) |

Values are expressed as numbers (percentage, %). We conducted a Chi-square test, and the results did not reveal significant differences among conditions ($\chi^2$ (7) = 4.84, p = 0.67).

† Most of the fetal position abnormalities were breech presentation (319 cases). Lateral position (three cases) and face position (one case) were also indicated.

‡ All uterine surgeries leading to Cesarean section were either laparoscopic or abdominal enucleatic myomectomy (90 cases).

§ Major maternal complications included pregnancy with uterine myoma and hypertensive disorders of pregnancy, as well as pregnancy after renal transplantation, epilepsy, sexually transmitted diseases, heart disease, and psychiatric disease.

¶ Among the reasons for planned Cesarean section for fetal anomalies, fetal growth restriction (five cases) and spina bifida (five cases) were the most common, followed by fetal anemia because of Rh-incompatible pregnancy (one case), diaphragmatic hernia (one case), and multiple malformations (one case).

Table 3 Evaluation results for the proficiency evaluation indicators

| | | Trainee group (n = 259) | Specialist group (n = 486) | Mean between-group difference (95% CI) | p value |
|---|---|---|---|---|---|
| Operation time (min) | I-D time | 9.5 (± 4.1) | 7.6 (± 4.3) | 1.9 (1.2 to 2.5) | < 0.01 |
| | D-S time | 60.7 (± 14.2) | 50.2 (± 14.4) | 10.5 (8.3 to 12.6) | < 0.01 |
| | I-S time | 70.1 (± 16.5) | 57.8 (± 16.6) | 12.3 (9.8 to 14.8) | < 0.01 |
| Apgar score (points) | 1 minute | 8.1 (± 1.0) | 8.2 (± 1.0) | − 0.1 (− 0.1 to 0.3) | 0.18 |
| | 5 minutes | 9.0 (± 0.4) | 8.9 (± 0.7) | 0.1 (− 0.1 to 0.1) | 0.39 |
| UmA pH | | 7.31 (± 0.06) | 7.31 (± 0.05) | 0.00 (− 0.01 to 0.01) | 0.40 |
| Blood loss (mL) | | 998.2 (± 500.8) | 997.3 (± 521.7) | 0.9 (− 77.3 to 79.1) | 0.98 |

Analyzed using Student's *t*-test. The value indicates the mean (± standard deviation).

8.20 ± 0.97 points (0.1 points; 95% CI, −0.1 to 0.3 points; p = 0.18), Apgar score 5 min: 9.0 ± 0.4 vs. 8.9 ± 0.7 points (0.1 points; 95% CI, −0.1 to 0.1 points; p = 0.39), UmA pH: 7.31 ± 0.06 vs. 7.31 ± 0.05 (0.00; 95% CI, −0.01 to 0.01; p = 0.40), and blood loss: 998.2 ± 500.8 vs. 997.3 ± 521.7 mL (0.9 mL; 95% CI, −77.3 to 79.1 mL; p = 0.98).

**Comparing the operation time between specialists and trainees**

Given the significant differences in the I-D, D-S, and I-S times, we compared the mean times of every five case groupings between the trainees and specialists (Fig. 2). Fig. 2 (a-1, b-1, c-1) shows a box plot of the medians and quartiles of the I-D, D-S, and I-S times for both groups. The I-D time differed significantly between the trainees and specialists for the first 10 surgeries. However, the operation time became shorter after the 11th procedure, and no significant differences were observed thereafter (Fig. 2a-1). The D-S time (Fig. 2b-1) and I-S time (Fig. 2c-1) did not differ significantly between the groups after the 16th procedure. Fig. 2 (a-2, b-2, c-2) shows plot curves showing the mean operation time for each case grouping, as well as a curve smoothed using the fifth-order MA method. The I-D time tended to decrease gradually from the first case (a-2), whereas the D-S time (b-2) increased or flattened until about the ninth case, showed a sharp reduction in time at around the 10th case, and plateaued after 16 cases. The I-S

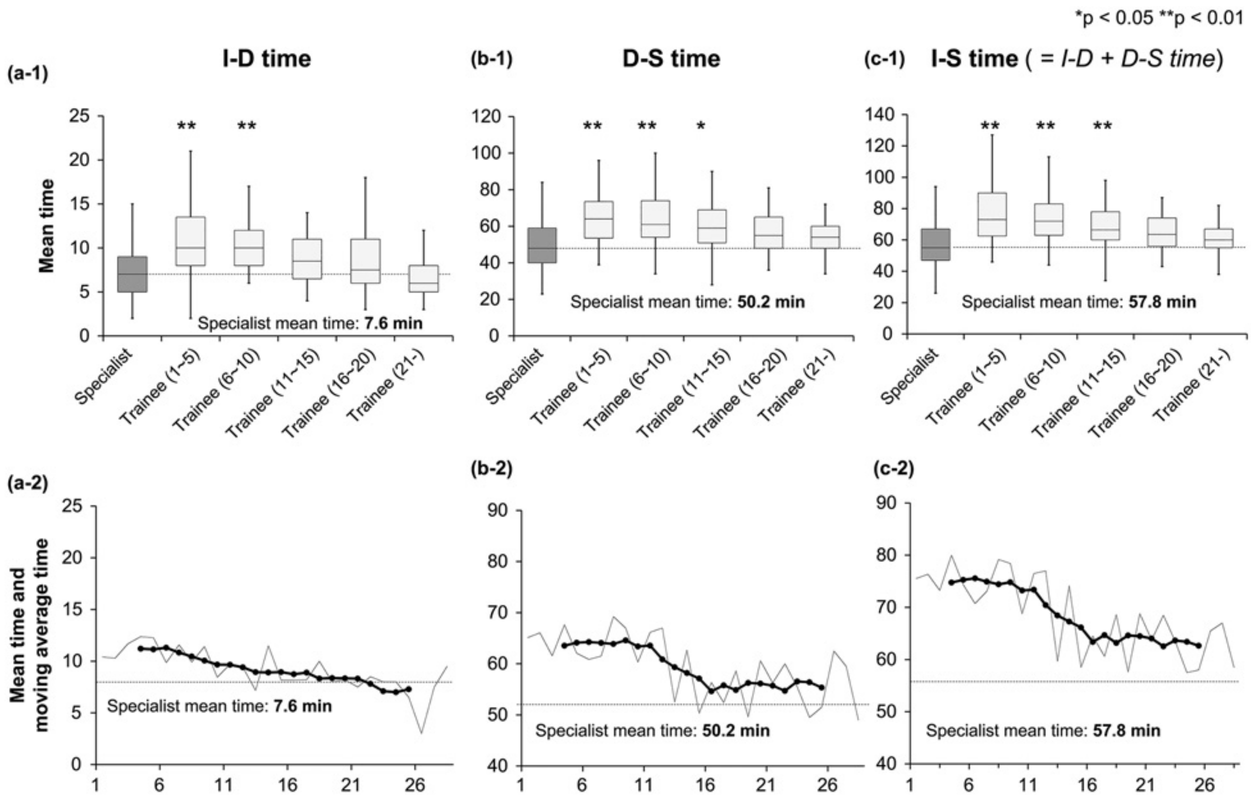


Fig. 2 Learning curve and moving average time.
When surgical experience exceeded 11 cases, no significant difference was found between the groups in the I-D time. For the D-S and I-S times, the difference disappeared after the 16th case. The results using the fifth-order moving average method are shown in the bottom graphs (a-2 to c-2).

time (c-2) showed a similar plot as the D-S time.

**Time series analysis and the inflection point in the learning curve by trainees**

Fig. 3 (a-3, b-3, c-3) shows the points of the best-fitted ARIMA model and the regression curve obtained via join point analysis. The black arrow (Fig. 3a-4, b-4, c-4) indicates a join point, which is a statistically significant inflection point. A peak in the I-D time was observed in the eighth case, which decreased sharply, thereafter decreasing monotonously. Inflection points were also identified for both the D-S and I-S time at the 13th and 18th procedures. In the 13th case, the learning curve entered a steep decline, and no marked increase or decrease was observed (a plateau) after the 18th inflection point.

## Discussion

The results indicate that operation time may be used as an indicator to evaluate the technical proficiency of trainees in the Cesarean section and that experience of at least 16 procedures may indicate proficiency equivalent to that of specialists in terms of operation time. The operation duration showed similar time series changes in the I-D, D-S, and I-S times, depending on the number of surgical experiences. Since the Apgar score, UmA pH, and blood loss did not differ significantly between the two groups, they were thus considered inappropriate as indicators of technical proficiency. Previous reports have often indicated complications and duration of hospitalization after surgery as proficiency indicators.[1–5] Certainly, these indicators are important in assessing the quality of surgery; however, using them as objective indicators for evaluating trainees' technical proficiency in Cesarean section seems difficult. A difference between the trainees and the specialist in terms of the incidence of complications and the length of hospital stay is disadvantageous to the patient assigned to the internship operation. Therefore, it is considered that these results should not be treated as indicators from the viewpoint of medical ethics.

We also found no significant differences in the operation times between the trainee and specialist groups after 15 cases, which suggests that at least 16-20 cases must be experienced for a trainee doctor to become proficient in Ce-

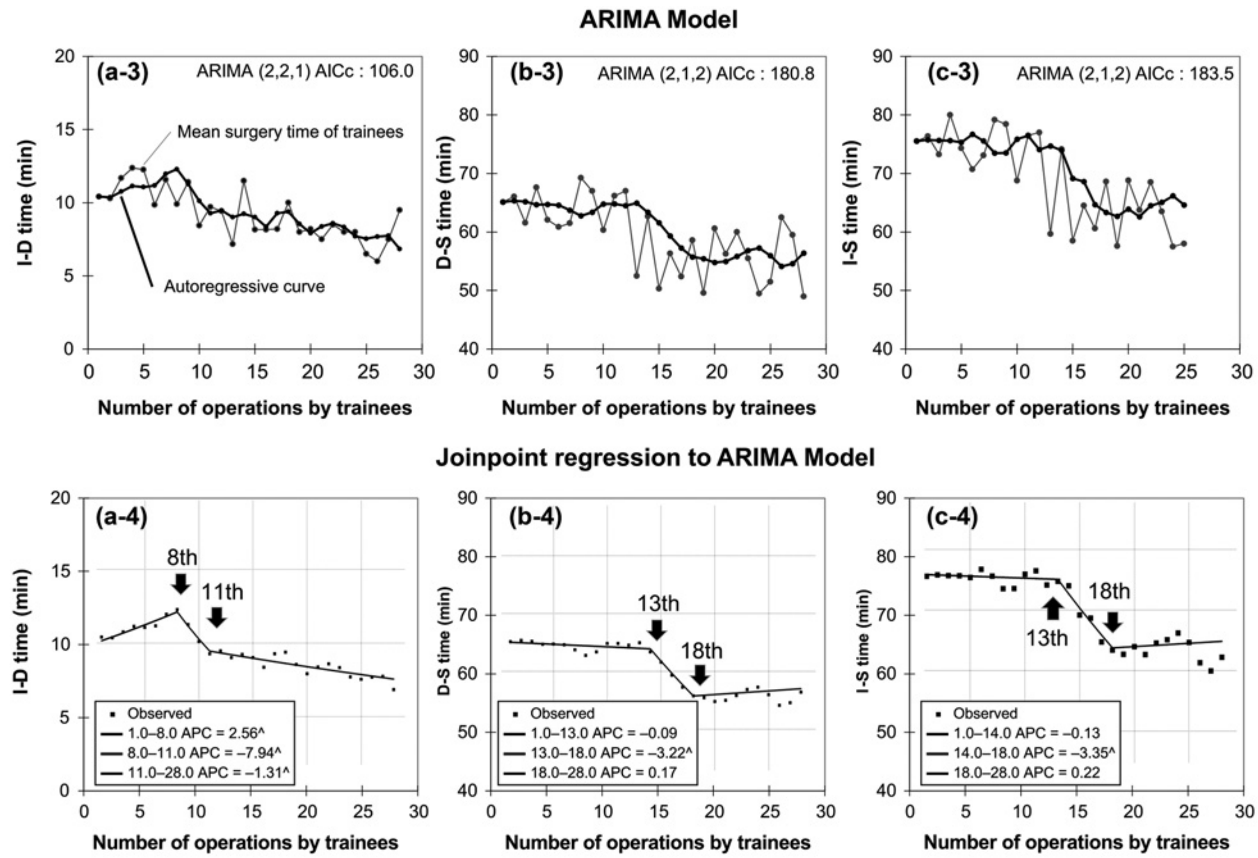


Fig. 3 Results of the time series analysis and inflection points.
Regarding the time series analysis model used, ARIMA (2, 2, 1) (AICc: 106.0) was adapted for the I-D time, and the models for the D-S and I-S times conformed to ARIMA (2, 1, 2) (AICc: 180.8) and ARIMA (2, 1, 2) (AICc: 183.5), respectively. The bottom graphs (a-4 to c-4) show the position of the inflection points (black arrows). Regarding the I-D time, statistically significant inflection points were observed in the eighth and 11th cases, whereas both the D-S and I-S times had significant inflection points in the 13th and 18th cases.

sarean section. This is just the minimum number of cases for proficiency, not the number of cases evaluated as "sufficiently mastered." Müller et al.[2] also reported no significant difference after 16 procedures and further noted a steep change beyond 20 procedures. Fok et al.[3] and Soergel et al.[4] noted a plateau in the learning curve after 10-15 procedures, whereas Waalewijn et al.[5] found a steep reduction in operation time after 15 procedures. The results of our study were thus consistent with previous reports.

The results of the ARIMA model and join point analysis explain these results more clearly. The I-D time tended to be longer initially, but a peak was observed around the eighth case and thereafter decreased steadily. This change is because the I-D time involves the same operation procedure being performed in all cases, with skin, fascia, peritoneum, and myometrial incisions. This suggests that performing the surgical procedure around eight times helps polish the technique. By contrast, the learning curves for the D-S and I-S times entered a steep decline phase at the 13th case, and no marked increase or decrease was subsequently observed (plateau) after the inflection point at the 18th case. This trend seems to be due to the need for suturing techniques, as the D-S time increases with the number of suturing operations required, such as muscular stitching, fascia, peritoneum, and subcutaneous suturing following fetal delivery. These findings suggest that experience with at least 18 procedures is needed to refine suturing techniques.

These three inflection points strongly indicate that the Cesarean section learning curve is a sigmoidal growth curve. We may thus assume four stages to the learning process: an initial "imitation period" showing a steady increase, followed by "a trial-and-error period," a "steep improvement period" in which the time steadily decreases, and, finally, a "plateau period" at which the techniques and skills stabilize after the 16th-20th procedure. This four-stage learning process is referred to at our hospital as the "Cesarean Section Learning Model." Based on our findings, at least 20 cases of surgical experience are thus required for obstetrics and gynecology trainees to become proficient in the Cesarean section and stabilize their skills.

**Limitations of the study**

This study has some limitations. This was a single-institution study, which necessitates other similar studies to be conducted in other institutions and obtain similar conclusions. Additionally, although comparisons were made between the trainee and specialist groups, some bias may have occurred in the selection of patients who were assigned to the specialist group that the trainees could have also handled, as suggested by the lack of significant differences between the two groups in the Apgar score, UmA pH, and blood loss. Many studies have reported the risk of Cesarean delivery among maternal overweight or obese compared with normal weight women.[28–30] BMI before pregnancy and weight gain during pregnancy may affect the degree of difficulty of Cesarean section surgery. However, a retrospective review of electronic medical records revealed that the data on maternal height and weight were often missing or absent in electronic medical data. Pregnant women with high BMI may be included in the specialist group due to the possibility of difficulty of the operation, which may result in longer operating times for specialist groups. In previous studies overseas, postoperative blood data (WBC, CRP, Hb, etc.), postoperative surgical site infection, scar formation, etc. were taken up as technical evaluations for the postoperative complication assessments.[1–5] One of the postoperative evaluation items is the length of hospital stay, but the medical system in Japan is based on the public insurance system, and in many cases, the number of days of hospitalization for Cesarean section is determined in advance (mostly 7 or 8 days after surgery). For this reason, assessing the length of hospital stay is difficult. As for the blood data, many of the cases in this study were scheduled in advance, and the preoperative blood test was performed at any time within 3 months. Additionally, since all postoperative blood sampling tests were not performed on a fixed schedule, comparable data could not be obtained. These limitations are because of a retrospective study. At our hospital, we have selected Cesarean section cases that can safely be performed under the guidance of specialists, even by inexperienced trainees in order not to occur postoperative complications. For ethical considerations, postoperative complication assessment was not performed in this study. Establishing a proficiency model for the Cesarean section by conducting similar quantitative evaluations at multiple facilities will aid in determining standardized criteria for assessing the qualifications of obstetrics and gynecology trainees in Japan.

**Conclusion**

Our study showed that operation time could be used to objectively evaluate technical proficiency in the Cesarean section among trainees. Analysis of the operation time suggests that at least 20 cases of surgical experience under the guidance of an instructing specialist doctor are required before a trainee doctor obtains stable skills. The time series analysis of the learning curve suggests four stages to the learning process: the "imitation period," "trial-and-error period," "improvement period," and "plateau period." We believe that our study will prove extremely useful for determining standardized criteria for proficiency in the Cesarean section among Japanese obstetrics and gynecology trainees.

**Acknowledgement/Funding source:** The authors would like to thank Enago (www.enago.jp) for the English language review.

**Authors' contribution:** T.R., E.H., designed the study; T.R., E.H., and K.M performed the experiments and analyzed the data; M.M., M.N., K.M., and T.H. provided critical reagents; M.N. and T.H. supervised the experiments; T.R. wrote the manuscript.

**Ethics statement:** The study protocols involving human participants were reviewed and approved by the Ethics Committee of Toho University Omori Medical Center (M17226).

**Conflicts of interest:** None declared.

## References

1) Madsen K, Grønbeck L, Rifbjerg Larsen C, Østergaard J, Bergholt T, Langhoff-Roos J, et al. Educational strategies in performing cesarean section. Acta Obstet Gynecol Scand. 2013; 92: 256-63.
2) Müller I, Zimmermann R. Die Lernkurve im Zusammenhang mit der Sectio caesarea [The learning curve in the context of the cesarean section]. Gynakol Geburtshilfliche Rundsch. 2003; 43: 238-44.
3) Fok WY, Chan LY, Chung TK. The effect of learning curve on the outcome of caesarean section. BJOG. 2006; 113: 1259-63.
4) Soergel P, Jensen T, Makowski L, von Kaisenberg C, Hillemanns P. Characterisation of the learning curve of caesarean section. Arch Gynecol Obstet. 2012; 286: 29-33.
5) Waalewijn BP, van Duinen A, Koroma AP, Rijken MJ, Elhassein

M, Bolkan HA. Learning curve characteristics for caesarean section among associate clinicians: a prospective study from Sierra Leone. World J Surg. 2017; 41: 2998-3005.
6) Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, et al. Beyond "see one, do one, teach one": toward a different training paradigm. Qual Saf Health Care. 2009; 18: 63-8.
7) Sheehan H, Gray T, Farrell T. The potential value of surrogate performance markers at caesarean section for the assessment of surgical competence. Eur J Obstet Gynecol Reprod Biol. 2018; 231: 30-4.
8) Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997; 84: 273-8.
9) Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg. 1997; 173: 226-30.
10) Goff B, Mandel L, Lentz G, VanBlaricom A, Oelschlager AM, Lee D, et al. Assessment of resident surgical skills: is testing feasible? Am J Obstet Gynecol. 2005; 192: 1331-8.
11) Landau A, Reid W, Watson A, McKenzie C. Objective structured assessment of technical skill in assessing technical competence to carry out caesarean section with increasing seniority. Best Pract Res Clin Obstet Gynaecol. 2013; 27: 197-207.
12) Oestergaard J, Larsen CR, Maagaard M, Grantcharov T, Ottesen B, Sorensen JL. Can both residents and chief physicians assess surgical skills? Surg Endosc. 2012; 26: 2054-60.
13) Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. BJOG. 2008; 115: 908-16.
14) van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010; 97: 972-87.
15) Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ. 2003; 327: 1032-7.
16) Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. J Surg Educ. 2008; 65: 77-83.
17) Paisley AM, Baldwin PJ, Paterson-Brown S. Validity of surgical simulation for the assessment of operative skill. Br J Surg. 2001; 88: 1525-32.
18) Desender LM, Van Herzeele I, Aggarwal R, Vermassen FE, Cheshire NJ. Training with simulation versus operative room attendance. J Cardiovasc Surg (Torino). 2011; 52: 17-37.
19) Palter VN, Grantcharov TP. Simulation in surgical education. CMAJ. 2010; 182: 1191-6.
20) Moulton CA, Dubrowski A, MacRae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. Ann Surg. 2006; 244: 400-9.
21) Moulton CA, Regehr G, Lingard L, Merritt C, MacRae H. Operating from the other side of the table: control dynamics and the surgeon educator. J Am Coll Surg. 2010; 210: 79-86.
22) van Det MJ, Meijerink WJ, Hoff C, Middel LJ, Koopal SA, Pierie JP. The learning effect of intraoperative video-enhanced surgical procedure training. Surg Endosc. 2011; 25: 2261-7.
23) Xeroulis GJ, Park J, Moulton CA, Reznick RK, Leblanc V, Dubrowski A. Teaching suturing and knot-tying skills to medical students: a randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. Surgery. 2007; 141: 442-9.
24) Driscoll PJ, Paisley AM, Paterson-Brown S. Video assessment of basic surgical trainees' operative skills. Am J Surg. 2008; 196: 265-72.
25) Byrne AJ, Pugsley L, Hashem MA. Review of comparative studies of clinical skills training. Med Teach. 2008; 30: 764-7.
26) Sharma B, Mishra A, Aggarwal R, Grantcharov TP. Non-technical skills assessment in surgery. Surg Oncol. 2011; 20: 169-77.
27) Undre S, Koutantji M, Sevdalis N, Gautama S, Selvapatt N, Williams S, et al. Multidisciplinary crisis simulations: the way forward for training surgical teams. World J Surg. 2007; 31: 1843-53.
28) Machado LS. Cesarean section in morbidly obese parturients: practical implications and complications. N Am J Med Sci. 2012; 4: 13-8.
29) Berendzen JA, Howard BC. Association between cesarean delivery rate and body mass index. Tenn Med. 2013; 106: 35-7.
30) Getahun D, Kaminsky LM, Elsasser DA, Kirby RS, Ananth CV, Vintzileos AM. Changes in prepregnancy body mass index between pregnancies and risk of primary cesarean delivery. Am J Obstet Gynecol. 2007; 197: 376.e1-7.