

REVIEW ARTICLE

## Skill accreditation system for laparoscopic gastroenterologic surgeons in Japan

TOSHIYUKI MORI<sup>1,2</sup>, TAIZO KIMURA<sup>1,3</sup>, MASAKI KITAJIMA<sup>4</sup>

<sup>1</sup>ESSQS Committee of the Japan Society for Endoscopic Surgery, <sup>2</sup>Department of Surgery, Kyorin University, Tokyo, Japan, <sup>3</sup>Department of Surgery, Fujinomiya City General Hospital, Shizuoka, Japan, and <sup>4</sup>Department of Surgery, International University of Health and Welfare, Tochigi, Japan

### Abstract

The Japan Society for Endoscopic Surgery (JSES) has established an Endoscopic Surgical Skill Qualification System and started examination in 2004. Non-edited videotapes were assessed by two judges in a double-blinded fashion with strict criteria. Two kinds of criteria, namely common and procedure-specific, were prepared. The common criteria were designed to evaluate set-ups, autonomy of the operator, display of the surgical field, recognition of surgical anatomy, co-operation of the surgical team. The procedure-specific criteria were made to assess the operation in a step-by-step fashion. In total, out of 1114 surgeons who were assessed by this qualification system over a period of four years, 537 (48.2%) have been accredited. The qualification rate in each surgical field has remained at the same level of 40 to 50% to date. Inter-rater agreement of two judges was low at 0.31 in the first year, but improved with revision of the criteria and consensus meetings. Surgeons assessed by this system as qualified experienced less frequent complications when compared to those who failed. This system has impacted on the improvement and standardization of laparoscopic surgery in Japan.

**Key words:** Skill assessment, laparoscopic surgery, complication

### Introduction

Laparoscopic surgery is widely accepted as a less invasive alternative to classical surgery. Nevertheless it also imposes serious problems. Surgical knowledge learned in open surgery is not intuitively applicable to laparoscopic surgery. Visual perception is limited and surgical exertion is awkward for long and leveraged instruments. These factors can result in complications including fatal ones. The higher rate of complications was initially attributed to the learning curve problem, but in reality it was not. In Japan there have also been reports on considerable complications after laparoscopic operations. In some of those cases patients died and picked up as lawsuit cases. To address this problem, the Japan Society for Endoscopic Surgery (JSES) has established the Committee of the Endoscopic Surgical Skill Qualification System in 2001 which consisted of surgeons from various fields. The actual accreditation system in gastrointestinal

(GI) and general surgery started in 2004. A skill qualification and accreditation system run by an academic body like this has not been established in any other country. In this paper, we will discuss the methods and results of this endoscopic surgery accreditation system in the field of digestive surgery, with specific focus on the inter-rater agreement of this system to test reproducibility and on the complication rate of the applicants to assess clinical relevance.

### Material and methods

#### Selection of judges

To select judges, a committee was established that consisted of eight JSES directors and three accreditation system committee members, all of them from the GI surgery field. In the beginning, 25 surgeons who actively present their surgical skills in meetings

were selected. Non-edited videotapes were assessed based on the hypothesis that the steps of perceptual motor skill in laparoscopic surgery could all be evaluated by reviewing these videotapes with strict criteria. The videotapes of these surgeons were mutually evaluated and they were accredited as initial judges. As a result, they were all assessed to have a level of technique that was adequate for becoming judges. Judges (a total of 49) were then replaced or added by selection from among surgeons who had passed the accreditation examination with a high score. The judges were then divided into six groups according to their subspecialty: Esophagus, biliary tract, stomach, colon, spleen, and others. The names of the judges were kept confidential.

### Requirements and specifications

The following requirements must be fulfilled by applicants for eligibility:

- He/she must be a board certified surgeon;
- he/she must have attended the JSES scientific meeting (three points), educational seminars (four points) or workshops using animals (three points), with a total of more than 12 points;
- he/she must have conducted simple surgical procedures (e.g. cholecystectomy) in at least 50 patients, or complex procedures (e.g. colectomy) in 20 patients or more;
- he/she must be recommended by two instructors;
- he/she must have presented at least three reports on endoscopic surgery at scientific meetings.

The paperwork included a CV, a list of cases on which the applicant has operated (including complications), and recommendation letters. Submission of non-edited videotape(s) was required. If the submitted video did not include suturing and knot-tying, an additional video that included suturing and knot-tying was also required. In the first year, videotapes of a broad spectrum of procedures were accepted. But it turned out to be difficult to objectively assess procedures without strict criteria. Therefore, from the second year on, procedures have been limited, as shown in Table I.

### Criteria

Two kinds of criteria were prepared for assessment: Common and procedure-specific criteria.

The common criteria were designed with the intention to evaluate set-ups, autonomy of the operator, display of the surgical field, recognition of surgical anatomy, co-operation of the surgical team (Table II).

Table I. Procedures to be submitted.

- 
- Esophagectomy
  - Nissen's operation
  - Heller's operation
  - Cholecystectomy
  - Common bile duct clearance
  - Distal gastrectomy
  - Splenectomy
  - Adrenalectomy
  - Nephrectomy
  - Sigmoidectomy
  - Mastectomy
  - Thyroidectomy
  - Inguinal hernia repair
- 

Table II. Common criteria (points allotted).

#### Category 1: Progress of the operation

Smooth conduct of the operation (4 pts)

Autonomy of the operator (4 pts)

Leadership ability (4 pts)

Cooperation with assistants (4 pts)

#### Category 2: Display of the operating field

Proper positioning of the access ports (3 pts)

Display of the operating field in the center of the monitor (3 pts)

Clear display of the target organ (3 pts)

Proper use of the retractor (3 pts)

Proper use of non-dominant forceps (3 pts)

#### Category 3: Operative techniques

Proper selection and appropriate use of forceps (dominant side) (3 pts)

Proper methods of traction and tissue handling (3 pts)

Appropriate and smooth use of the correct type of energy (3 pts)

Correct layer of tissue dissection (5pts)

Correct identification and proper coagulation or clipping of blood vessels (5 pts)

#### Category 4: Suturing and knot-tying

Suturing (5 pts)

Knot-tying (5 pts)

---

It should be stressed that these criteria are not intended to evaluate the manual dexterity of the applicants. Laparoscopic suturing and knot-tying were also evaluated to assess the two-hand coordination. Sixty points were allotted for the common criteria.

The procedure-specific criteria were made to assess the operation in a step-by-step fashion (Table III; Gastrectomy). Forty points were allotted for the procedure-specific criteria. At the early stage, a broad

Table III. Procedure-specific criteria (distal gastrectomy).

---

Two points are allotted for each item.

1. Are the ports positioned appropriately?
  2. Is the operating field secured by appropriate exclusion of the liver, etc.?
  3. Are appropriate grasping forceps used that can prevent injury to the stomach, duodenum, and small and large intestines?
  4. Are the tissues grasped with appropriate force?
  5. Is the gastric wall or intestinal wall grasped at full thickness (all layers)?
  6. Is the correct site pulled in the proper direction?
  7. No serosal damage?
  8. No bleeding caused by rough use of grasping forceps?
  9. Is the gastrocolic ligament divided appropriately?
  10. Are any measures taken to prevent injury to the large intestine?
  11. Is the left gastric vein transected appropriately?
  12. Is the left gastric artery transected appropriately?
  13. Are any measures taken to prevent injury to the pancreas?
  14. Is the No. 1 lymph node dissected appropriately?
  15. Is the No. 3 lymph node dissected appropriately?
  16. Is the extent of lymph node dissection sufficient?
  17. Is gastro-duodenal anastomosis completed without error?
  18. Is blood flow adequate at the anastomotic site?
  19. No excessive tension at the anastomosis?
  20. Is the anastomosis constructed with a good shape?
- 

spectrum of procedures was accepted; however, procedures are now limited for criteria preparation. Criteria have been revised several times for more accurate and reproducible evaluation.

### Accreditation process

For evaluation in a double-blinded fashion, two judges were assigned to the videotape submitted by an applicant. Judges were informed only by the number on the videotapes so that they did not notice the applicant's name or institution or the other judge who evaluated the same videotape.

The qualification line was set at 70 points. In cases where the results of two judges were the same, the decision was automatically made accordingly. When the results of two judges were different, the videotape was subjected to a third judge or to group discussion.

### Analysis

In addition to the qualification rate in each field, inter-rater agreement between two independent judges was

also calculated using weighted kappa value (Cohen 1961), to assess the reproducibility of this system. In order to assess the validity of the system and its relevance for clinical practice, complication rates were compared for qualified and non-qualified surgeons. As a test for the significance of differences, the Mann-Whitney U test was used.

## Results

### Qualification rate (Table IV)

In 2004, 422 surgeons applied to this accreditation system and 212 surgeons qualified, yielding a qualification rate of 50%. The qualification rate in the biliary group was as high as 65% and that in the esophagus group was as low as 28%. Thereafter, 269, 217, and 203 surgeons applied in 2005, 2006, and 2007, respectively. The overall qualification rate has consistently been between 40 to 50%. As mentioned already, criteria in subspecialty fields have been revised several times. Furthermore, there has been no cross-talk between subspecialty groups or any adjustment of the qualification rate. Nonetheless, the qualification rate in each field tends to plateau at the same level. In total, out of the 1,114 surgeons who applied to this qualification system, 537 (48.2%) have been accredited.

### Inter-rater agreement (Table V)

Inter-rater agreement in 2004 was 0.31 in total. The colon group displayed the most reproducible judge, and weighted kappa value was as low as 0.18 in the biliary group. As procedure-specific criteria were revised, and consensus meetings were held, inter-rater agreement improved to 0.4 in 2005, and the most remarkable improvement was observed in the biliary group. Despite continuing revision of the criteria, limitation of the procedure, and consensus meetings, inter-rater agreement remained 0.36 and 0.38 in total, in 2006 and 2007, respectively.

### Complications

Although inter-rater agreement is not satisfactory, it is notable that the overall incidence of complications was significantly lower in patients treated by applicants who acquired accreditation (4.8%) than in patients treated by applicants who failed it (6.5%) in 2004. Nevertheless, there were no significant differences in complication rates in relation to the organs. Because a broad spectrum of the procedures

Table IV. Qualification Rate.

|                | 2004              |               |              | 2005              |               |              |
|----------------|-------------------|---------------|--------------|-------------------|---------------|--------------|
|                | No. of Applicants | No. of Qulif. | Qualif. Rate | No. of Applicants | No. of Qulif. | Qualif. rate |
| Biliary tract  | 170               | 110           | 65%          | 137               | 62            | 45%          |
| Esophagus      | 32                | 9             | 28%          | 17                | 10            | 59%          |
| Stomach        | 81                | 37            | 46%          | 37                | 21            | 57%          |
| Colon          | 103               | 38            | 37%          | 61                | 24            | 39%          |
| Spleen         | 18                | 9             | 50%          | 5                 | 2             | 40%          |
| Endocrine, etc | 18                | 9             | 50%          | 12                | 9             | 75%          |
| Total          | 422               | 212           | 50%          | 269               | 128           | 48%          |
|                | 2006              |               |              | 2007              |               |              |
|                | No. of Applicants | No. of Qulif. | Qualif. rate | No. of Applicants | No. of Qulif. | Qualif. rate |
| Biliary tract  | 101               | 32            | 32%          | 82                | 32            | 39%          |
| Esophagus      | 9                 | 3             | 33%          | 9                 | 6             | 67%          |
| Stomach        | 35                | 17            | 49%          | 47                | 21            | 45%          |
| Colon          | 60                | 25            | 40%          | 58                | 27            | 47%          |
| Spleen         | 2                 | 1             | 50%          | 1                 | 1             | 100%         |
| Endocrine, etc | 9                 | 5             | 56%          | 6                 | 3             | 50%          |
| Total          | 217               | 83            | 47%          | 203               | 90            | 44%          |

was accepted in the first year, there was a wide range of complexity of operations. This may have resulted in a blurred result of complication rates in each group. From 2005, the procedures in the gastric and colon groups were limited to distal gastrectomy and sigmoidectomy for malignancy, respectively. The results of the complication study were notable in both groups, with a significant difference between qualified and non-qualified applicants. In total, from 2004 to 2007, the complication rates for qualified surgeons compared significantly better than those for non-qualified surgeons,  $p < .03$  in the gastric group and  $P < .005$  in the colon group, respectively (Table VI). On the other hand, when Nissen, Heller-Dor procedures and esophagectomy for cancer were still accepted in the esophagus group, this accreditation system for the esophagus did not display the ability to detect higher rates of complications. In the biliary group, videotapes that recorded cholecystectomy for the non-inflamed gallbladder were accepted. It turned

out very difficult to evaluate them with regard to possible complications.

## Discussion

First, the basic purpose of skill assessment is to develop a tool for the reliable and reproducible evaluation of the surgical techniques of trainees, thus to help establish a structured training program of operation. There have been a number of reports about technical assessment methods, including the use of videos or watching live surgery. Winckel et al. (1) introduced a structured technical skills assessment form (STSAF), which employed both procedure-specific checklists and global rating (global assessment). Martin et al. (2) also introduced an objective structured assessment of technical skill (OSATS), which included specific checklists and a global rating, as well as a pass/fail judgment. For assessment of techniques in laparoscopic cholecystectomy, Eubanks et al. (3) introduced an objective scoring system, which rates surgical techniques by adding points for a pass and subtracting points for a *fail* in each step. In 2005, Vassiliou et al. (4) introduced the global operative assessment of laparoscopic skills (GOALS), which includes both global assessment and a 10-item checklist, taking into account a visual analogue scale for surgical difficulty. These assessment methods are designed for surgical trainers when assessing residents and young surgeons, and high inter-rater

Table V. Inter-rater agreement ( $\kappa$ w).

|               | 2004 | 2005 | 2006 | 2007 |
|---------------|------|------|------|------|
| Biliary tract | 0.18 | 0.35 | 0.29 | 0.2  |
| Esophagus     | 0.36 | 0.28 | 0.3  | 0.62 |
| Stomach       | 0.37 | 0.59 | 0.34 | 0.28 |
| Colon         | 0.4  | 0.37 | 0.4  | 0.37 |
| Total         | 0.31 | 0.4  | 0.36 | 0.38 |

Table VI. Complication rate in qualified and non-qualified applicants.

|               | Complication rate (%) mean±SE 2004-2007 |                              |                          | p value |
|---------------|---|------------------------------|--------------------------|---------|
|               | No of Applicants Accredited/Failed      | Complication rate Accredited | Complication rate Failed |         |
| Biliary tract | 555<br>(274/281)                        | 3.2±5.9                      | 3.7±7.7                  | 0.871   |
| Esophagus     | 70<br>(34/36)                           | 11.0±11.8                    | 9.7±11.6                 | 0.6538  |
| Stomach       | 262<br>(127/135)                        | 4.7±5.9                      | 7.6±1.0                  | 0.0284  |
| Colon         | 350<br>(155/195)                        | 4.4±6.4                      | 6.6±7.9                  | 0.0048  |
| Total         | 1114<br>(537/577)                       | 4.3±6.8                      | 5.6±8.4                  | 0.0096  |

agreement between the trainers and good reproducibility of their assessments has been reported.

In addition to the above-mentioned purposes, we aimed to assess the safety and efficacy of operation. The most significant difference in the aim of the system may be that our system was designed to evaluate the skill of practicing surgeons rather than that of trainees. As described previously, our colleagues were sued for having inappropriate skills to perform total prostatectomy, and even sentenced to jail. It is certainly difficult for local privileging committees to assess the skills of surgeons who perform laparoscopic operations, because they are all board certified surgeons in their specialty, and defective information is just skill level in laparoscopic surgery.

We decided to assess non-edited videotapes in a belief that the steps of perceptual motor skill in laparoscopic surgery could all be evaluated by reviewing these videotapes with strict criteria. Judges can see all the images that surgeons comprehend during the operation, and are thus able to evaluate not only surgical techniques, but also misconceptions or risky maneuvers that may lead to complications.

At the early stage, common criteria 'Category 1: Progress of the operation' did not include the items to check autonomy and leadership ability of the operator. It turned out that, in not a few instances, the super-assistant helped too much, obscuring the leadership of the operator. Then we decided to demand that the operator be a conductor of the operation and a leader of surgical team. Other items in the common criteria catalogue have received only minor revision throughout the study period. Again, highlight is placed on smooth conduction of the procedure, proper recognition of surgical anatomy, and appropriate selection and use of surgical instruments. Manual dexterity is of less

concern except for suturing and knot-tying items to evaluate two-hand coordination.

The procedure-specific criteria were designed to assess the procedure in a step by step fashion. In the course of criteria revision, ambiguous description of steps has been corrected. As a result, each procedure has been much standardized.

Inter-rater agreement between the two judges was not very high. The overall kappa value for the 2004 ratings was low at 0.31, and the reliability of the judgments was questioned. Fairness was ensured to some extent by either a third judge or a group decision when the two raters did not agree. However, an increase in inter-rater agreements appears to be essential for this system to be valuable. After the completion of examinations in 2004, the judging committee discussed the reasons for the low inter-rater agreement. Possible reasons included the following:

- Surgical procedures differed between institutions (especially cholecystectomy);
- specifications for surgical techniques were vague;
- procedure-specific criteria were imprecise.

As countermeasures

- consensus meetings were held for judges to discuss differences in surgical techniques and the permissible range—especially for cholecystectomy;
- the following specifications were added to reduce differences in the difficulty of surgery: "A video of hand-assisted laparoscopic surgery is not to be accepted;" "Distal gastrectomy with lymph node dissection for gastric cancer is only to be accepted in gastric surgery;" and "Sigmoidectomy with lymph node dissection for colon cancer and intra-abdominal anastomosis in colonic surgery;"

- the procedure-specific criteria were modified to more detailed criteria for various surgical procedures.

Such modifications were related to placing stricter limitations on surgical techniques. Table III shows the criteria for judgment used in 2005. When these measures were taken, the overall kappa value for 2005 increased to an acceptable level of 0.40. However, this is still not satisfactory. It is necessary to further improve inter-rater agreement by holding consensus meetings and by improving the judgment criteria.

To investigate the clinical relevance of accreditation, we studied the lists of patients submitted by the applicants and investigated differences in the incidence of surgical complications between patients treated by applicants who passed the examination and patients treated by applicants who failed it. In total, the incidence of complications was significantly lower in patients treated by applicants who passed ( $p < .01$ ). The stomach and colon groups displayed a sufficient ability to stratify the risk of complication. On the other hand, the biliary and esophagus groups failed to display such abilities. The reasons for this difference may be attributed to the fact that relatively simple cholecystectomy is not technically demanding enough to be used for this purpose. In the esophagus group, three kinds of procedures, esophagectomy, Nissen, and Heller-Dor, were accepted, resulting in a blurred outcome. Considering that the definition of complications was vague and that complications were reported voluntarily by the applicants (verification was not conducted), these results are not reliable enough, although they appear to suggest the validity of accreditation. We think the criteria need to be repeatedly revised for a better reproducibility, and must have clinical relevance or educational importance.

The Committee of the Endoscopic Surgical Skill Qualification System consisted of members from GI surgery, urologic surgery, respiratory surgery, orthopedic surgery, and pediatric surgery. The committee first discussed the basic concepts of an accreditation system in 2001 and reached an agreement on the following four points:

- An applicant must be a specialist in a certain field and have sufficient experience in endoscopic surgery;

- technical assessment is performed by viewing an unedited video of surgery performed by the applicant;
- surgeons whose technique is sufficiently good for them to act as instructors should be accredited;
- a Judging Committee should be established in each field.

In parallel to the accreditation system in the GI field, the first accreditation examination for endoscopic surgeons was held for obstetric/gynecologic surgeons in 2003, while accreditation started in 2004 for urologic (5) and orthopedic surgeons.

It is notable that qualification rate and inter-rater agreement are essentially alike in systems, although details in systems are different.

## Conclusions

The endoscopic surgery accreditation system that we describe here is the first in the world. We believe that laparoscopic surgical skill can be assessed by this system. Of course, the system still needs improvement, but it has already prompted more educational seminars and lectures to be held in Japan. Considering that it has already contributed much to improvement and standardization, this system may enhance the surgical skills of endoscopic surgeons and thus decrease adverse outcome of video-endoscopic surgery.

## References

1. Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form (STSAF). *Am J Surg.* 1994;167:423-7.
2. Martin JA, Begehr G, Reznick R. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84:273-8.
3. Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S, et al. An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg.* 1999;189:566-74.
4. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg.* 2005;190:107-13.
5. Matsuda T, Ono Y, Terachi T, Naito S, Baba S, Miki T, et al. The endoscopic surgical skill qualification system in urological laparoscopy: A novel system in Japan. *J Urol.* 2006;176:2168-72.