Development and preliminary validation of a new task-based objective procedure-specific assessment (OPSA) of inguinal hernia repair procedural safety

Vahagn Nikolian, MD[1], Diego Camacho, MD[2], David Earle, MD[3], Ryan Lehmann, DO[4], Peter Nau, MD[4], Bruce Ramshaw, MD[5], Jonah Stulberg, MD[6]

[1]Division of Gastrointestinal & General Surgery, Oregon Health & Science University; [2]Department of Surgery, Montefiore Medical Center; [3]Department of Surgery, Tufts University School of Medicine; [4]Department of Surgery, Section of Bariatric Surgery, University of Iowa ; [5]CQInsights PBC; [6]Department of Surgery, UT Health Science Center, McGovern School of Medicine

BACKGROUND

- Inguinal hernias are relatively common, with a lifetime prevalence estimated at 27% for men and 3% for women[1].
- Peri-operative complications occur in an estimated 2.9% - 8% of surgical repairs, with post-operative intervention required in 22% (Clavien-Dindo grade ≥3)[2].
- In the era of entrustable professional activities, incorporation of objective assessments of surgical proficiency into surgical training programs is a priority[3].
- As each procedure includes numerous individual tasks that a surgical trainee must complete competently, an objective procedure-specific assessment that includes task-specific assessments may lead to improved training.

OBJECTIVE

- To assess the inter-rater reliability of a novel objective procedure-specific assessment (OPSA) that measures the safe completion of eight specific consecutive tasks required to complete inguinal hernia (IH) repair.

METHODS

Study Design

- Six hernia expert surgeons rated 20 de-identified operative videos of minimally invasive laparoscopic and robotic IH repair of varying difficulty, size, and anatomy using the IH OPSA instrument.
- Raters were blinded to surgeon and to other rater assessments.

Population

- A convenience sample of operative videos selected from one author's (BR) institution and from a library of inguinal hernia videos representing multiple operating surgeons in the United States.
- Of the 20 operative videos, ten were laparoscopic total extraperitoneal (TEP), five were laparoscopic trans-abdominal pre-peritoneal (TAPP), and 5 were robotic TAPP inguinal hernia repairs. The most common inguinal hernia repair defect was indirect (14), followed by direct (5) and femoral (1).

Instrument development

- Initial 8-item instrument was created by two authors (VN and JS).
- Instrument review and feedback was solicited from five hernia surgical experts and incorporated into the instrument.
- The instrument was piloted with five surgeons with specific focus upon getting feedback regarding "unsafe" scoring. Feedback was incorporated into the final instrument's instruction set.
- Final instrument included the following items: 1) inclusion / port placement; 2a/2b) elevation of peritoneal flap (TAPP) or elevation of peritoneal flap (TEP); 3) exposure; 4) reducing the sac; 5) full dissection of the myopectineal orifice; 6) mesh insertion; 7) mesh fixation; and 8) operation flow (Figure 1).
- Scoring for each item: "poor" = 1; "adequate" = 2; "excellent" = 3. For analysis of safe vs. unsafe assessments, the "adequate" and "excellent" response options were combined into the "safe" categorization.

Data and data collection

- Surgical videos were de-identified and uploaded into a cloud-based SaaS platform.
- Each surgeon-rater scored each operative video using the IH OPSA.

Analysis

- Inter-rater reliability was assessed using total agreement and Gwet's AC_2[4].
- Total agreement is the ratio of raters that agreed on case safety. A score of 100% required each of the six raters to score identically as either safe (score of 2 or 3) or unsafe (score of 1).
- Gwet's AC_2 is a measure of test-retest reliability that adjusts for random agreement between observations.

RESULTS

Table 1. The Inguinal Hernia OPSA

Item	Rating Scale
1. Incision / Port placement	Poor (Unsafe – Position or insertion or removal technique) Adequate (Safe – Functional position [even if awkward] or some port insertion difficulty, but generally safe) Excellent (Expert-level performance – safe, efficient and optimal port position for procedure and anatomy)
2a. Elevation of Peritoneal Flap (TAPP)	Poor (Unsafe – Poor technique and/or excessive bleeding or trauma to flap) Adequate (Safe – Satisfactory technique but made closure more difficult than necessary) Excellent (Expert-level performance – Excellent technique and minimal bleeding with an intact flap for easy closure)
2b. Elevation of Peritoneal Flap (TEP)	Poor (Unsafe – Difficult entrance into pre-peritoneal cavity or difficulty with balloon insertion and inflation) Adequate (Safe – Moderate efficiency in balloon insertion and inflation) Excellent (Expert-level performance – Avoidance of peritoneal injury, careful insertion of balloon and inflation)
3. Exposure	Poor (Unsafe – Poor or inadequate pneumoperitoneum, or camera angle, or retraction with frequent loss of exposure) Adequate (Safe – Adequate establishment and maintenance of pneumoperitoneum, camera angle and retraction with occasional loss of exposure) Excellent (Expert-level performance – Optimizes exposure, efficiently directs retraction and camera while maintaining exposure and pneumoperitoneum)
4. Reducing the sac	Poor (Unsafe – Poor technique requiring greater than expected time) Adequate (Safe – Moderate efficiency in reducing the sac) Excellent (Expert-level performance – Careful and efficient reduction of the sac)
5. Full dissection of the myopectineal orifice	Poor (Unsafe – Poor technique- inadequate exposure of the myopectineal orifice) Adequate (Safe – Moderate efficiency in dissection exposing the full myopectineal orifice) Excellent (Expert-level performance – Careful, efficient, and full exposure in the dissection of the myopectineal orifice)
6. Mesh Insertion	Poor (Unsafe – Excessive time for mesh insertion or poor placement) Adequate (Safe – Moderately efficient mesh insertion and acceptable placement) Excellent (Expert-level performance – Efficient and accurate placement)
7. Mesh fixation	Poor (Unsafe – Poor technique or fixation technique) Adequate (Safe – Acceptable fixation technique) Excellent (Expert-level performance – Ideal fixation technique)
8. Operation Flow	Poor (Unsafe – Frequent lack of forward progression or frequently stopped operating or seemed unsure of next move) Adequate (Safe – Some forward planning and reasonable procedure progression) Excellent (Expert-level performance – Obvious planned course of operation and anticipation of next steps)

Table 2. Total Agreement and Gwet's AC_2 Measures of Inter-Rater Reliability for the Dichotomized (Safe vs Unsafe) Score, Overall (20 Operative Videos), and by Laparoscopic (15 Operative Videos) vs Robotic (5 Operative Videos) Procedures

	Overall		Laparoscopic		Robotic	
	Gwet's AC_2	Total Agreement	Gwet's AC_2	Total Agreement	Gwet's AC_2	Total Agreement
Q1	0.966	0.975	0.954	0.967	1.0	1.0
Q2a	0.822	0.517	0.522	0.767	0.929	0.900
Q2b	0.941	0.617	0.929	0.967	---	---
Q3	0.852	0.892	0.793	0.878	1.00	0.933
Q4	0.893	0.917	0.852	0.911	1.00	0.933
Q5	0.654	0.833	0.662	0.833	0.631	0.833
Q6	0.795	0.892	0.707	0.856	1.00	1.00
Q7	0.715	0.850	0.577	0.800	1.00	1.00
Q8	0.803	0.867	0.719	0.822	1.00	1.00

- Using the total agreement and Gwet's AC_2 measures, inter-rater reliability across the 8 task assessments ranged from 0.52 to 0.98 and from 0.65 to 0.97.
- The lowest rated tasks using total agreement were elevation of the peritoneal flap, regardless of procedure (TAPP: 0.52 and TEP: 0.62).
- Using Gwet's AC_2 the tasks with the lowest scoring reliability were full dissection of the myopectineal orifice (0.65) and mesh fixation (0.72).
- With the exception of Q5) full dissection of the myopectineal orifice, all inter-rater reliability measures were substantially higher in robotic procedures.

Figure 1. Average Rating and 95% Confidence Interval using the Safe (Score 2 or 3) vs Unsafe (Score 1) Rating Scale



- Item 1 (incision / port placement) was almost universally rated as completed safely.
- Full dissection of the myopectineal orifice and mesh fixation had lower average scores, indicating they were rated less safely completed than the other items.

CONCLUSIONS

- The IH OPSA assessment of intraoperative surgical quality has very good intra-rater scoring reliability.
- The IH OPSA is intended to be used in combination with general surgical technique assessment and does not seek to assess all aspects of potential safe intraoperative practice.
- As a task-based assessment, the IH OPSA, coupled with general surgical technique assessments has great potential to enable competency-based training, improve practice, and ensure patient safety.
- The digital recording of VBA-enabled IH task-based assessments have great potential to be fully transitioned to AI-based reviews over case-based assessments.

LIMITATIONS

- Study did not include association of unsafe vs. safe practices with clinical or financial outcomes.
- The small sample size of 20 operative videos limits the assessment of reliability across all levels of procedural complexity, anatomic variation, and procedure type (i.e., robotic, laparoscopic, TAPP, TEP, etc.)

REFERENCES

1. U.S. Department of Health and Human Services. (2019). *Inguinal hernia - niddk*. National Institute of Diabetes and Digestive and Kidney Diseases. From https://www.niddk.nih.gov/health-information/digestive-diseases/inguinal-hernia#common.

2. Weyhe D, Tabriz N, Sahlmann B, Uslar V-N. Risk factors for perioperative complications in inguinal hernia repair – a systematic review. *Innov Surg Sci* 2017;2(2):47-52.

3. Entrustable Professional Activities. American Board of Surgery. (2022). From https://www.absurgery.org/default.jsp?epahome.

4. Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology, 61(1), 29–48. https://doi.org/10.1348/000711006x126600

DISCLOSURES

This study was funded by Caresyntax.
This poster was presented at SAGES MONTREAL 2023.