



A multifactorial evaluation of objective performance indicators and video analysis in the context of case complexity and clinical outcomes in robotic-assisted cholecystectomy

Georges Kaoukabi¹ · Fahri Gokal¹ · Abeselom Fanta² · Xi Liu² · Mallory Shields² · Catherine Stricklin² · Alexander Friedman³ · Omar Yusef Kudsi⁴

Received: 30 May 2023 / Accepted: 31 August 2023 / Published online: 3 October 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Background The increased digitization in robotic surgical procedures today enables surgeons to quantify their movements through data captured directly from the robotic system. These calculations, called objective performance indicators (OPIs), offer unprecedented detail into surgical performance. In this study, we link case- and surgical step-specific OPIs to case complexity, surgical experience and console utilization, and post-operative clinical complications across 87 robotic cholecystectomy (RC) cases.

Methods Videos of RCs performed by a principal surgeon with and without fellows were segmented into eight surgical steps and linked to patients' clinical data. Data for OPI calculations were extracted from an Intuitive Data Recorder and the da Vinci® robotic system. RC cases were each assigned a Nassar and Parkland Grading score and categorized as standard or complex. OPIs were compared across complexity groups, console attributions, and post-surgical complication severities to determine objective relationships across variables.

Results Across cases, differences in camera control and head positioning metrics of the principal surgeon were observed when comparing standard and complex cases. Further, OPI differences across the principal surgeon and the fellow(s) were observed in standard cases and include differences in arm swapping, camera control, and clutching behaviors. Monopolar coagulation energy usage differences were also observed. Select surgical step duration differences were observed across complexities and console attributions, and additional surgical task analyses determine the adhesion removal and liver bed hemostasis steps to be the most impactful steps for case complexity and post-surgical complications, respectively.

Conclusion This is the first study to establish the association between OPIs, case complexities, and clinical complications in RC. We identified OPI differences in intra-operative behaviors and post-surgical complications dependent on surgeon expertise and case complexity, opening the door for more standardized assessments of teaching cases, surgical behaviors and case complexities.

Keywords Robotic cholecystectomy · Clinical outcomes · Objective performance indicators · Surgical data science · Surgical education

More than 90% of the 1.2 million cholecystectomies in the United States annually are performed laparoscopically,

making it the gold standard approach today. Although robotic cholecystectomy (RC) is a routine operation, complexities and post-surgical complications across cases can vary significantly [1, 2]. However, the factors that contribute to these variabilities remain elusive, likely due to the lack of sufficient intra-operative data that could be leveraged to determine how factors like surgical behaviors relate to case complexities and complications. Several scoring systems have been developed to report RC difficulty and complexity, such as the Nassar and Parkland Grading Scales [3, 4]. These simple, 5-point scales rate anatomical features during

✉ Omar Yusef Kudsi
omar.kudsi@tufts.edu

¹ Good Samaritan Medical Center, Brockton, MA, USA

² Applied Research, Intuitive Surgical Inc., Peachtree City, GA, USA

³ Tufts Medical Center, Boston, MA, USA

⁴ Tufts University School of Medicine, Boston, MA, USA

cholecystectomies, such as the appearance of the gallbladder and presence of adhesions, but both are based on subjective visual approximations and provide no objective insight on comprehensive procedural complexity. Beyond this, they require manual assessment and only reflect measures of anatomical characteristics without regard to surgeon actions or critical procedural steps throughout the case. Although surgical performance assessments today are designed to reflect the proficiency of surgical actions [5], they are burdened with the same time-consuming and subjective limitations that complexity scores possess. As such, there is a critical need for assessment of truly objective data captured during RC procedures, which would shed light into specific surgical behaviors and actions and provide metrics for timely readouts of case complexity and difficulty.

Recent digital advancements in surgical robotics enable data from video, surgeon and robotic movements and actions during the procedure to be collected, segmented into functional surgical steps, and calculated into a series of metrics for truly objective surgical evaluation, circumventing subjective, inconsistent, and time-consuming assessments of surgical performance [6, 7]. These metrics, or objective performance indicators (OPIs), have been shown to provide feedback on surgical performance at the surgical step-level [8–12], identify areas of surgical improvement during training, and predict clinical outcomes [13–16]. With the technology ready to support such studies, it is now crucial to define relationships between intra-operative behaviors and compounding factors like case complexity, surgeon experience and behaviors, and post-operative complications. However, no such studies exist today for RC.

In this study, we investigate RC case videos and robotic data streams to establish OPI links, both at the procedural and step-level, to case complexity, surgeon experience, teaching cases, and post-operative complication severities. Elucidation of these relationships paves the way for general surgery to incorporate such findings into new data-driven resources, such as truly objective assessments of surgical performance and RC complexity. These breakthroughs will enable such assessments to be incorporated into surgical training curricula at all levels for standardized and equitable surgical best practice and improved patient outcomes across surgery.

Methods

Patient data

Videos recorded from a prospectively collected database of robotic cholecystectomies performed by one surgeon between 2017 and 2022 were utilized. All cases were performed on the da Vinci® robotic system at a teaching

community institution, where clinical fellows participated in cholecystectomy procedures through a dual console configuration. Pregnancy and age younger than 18 years of age at the time of surgery were excluded from this study. All patients had signed an informed consent form prior to their inclusion in the study.

Pre-, intra- and post-operative patient variables were analyzed and included: patient demographics [age, sex, body mass index (BMI as kg/m²)], American Society of Anesthesiologists classification scores (ASA), cardiovascular and pulmonary comorbidities, smoking status, history of liver or gallbladder disease, and history of previous upper abdominal surgery. Procedure setting (elective or emergent), complicated gallbladders (gangrenous, fistulated, or abscessed), pathology results, operative times, intra-operative complications, estimated blood loss (mL, EBL), hospital length of stay (LOS—defined as the difference between post-operative discharge date and index operation date), emergency department (ED) revisit, hospital readmission within 30 days, and post-operative complications during follow-up visits were also collected. Post-operative complications scores were collected from follow-up notes of the surgeon, patients' medical records, and clinical charts. All complications were categorized according to the Clavien–Dindo classification system [17] and the Comprehensive Complication Index (CCI®, University of Zurich, Zurich, Switzerland), a 1–100 scoring measurement to quantify co-morbidity scores, was utilized as a highly correlated measurement to the Clavien–Dindo post-surgical complication score [18].

Surgeon and fellows robotic experience

The attending who performed the robotic cholecystectomies included in this study had already achieved their learning curve. The fellows had no significant experience in robotic surgery, with all of them assisting in a few number of cases during their residency training.

Video analysis and OPI calculations

Initially, 94 recorded cases were reviewed by an expert surgeon to ensure there were no missing surgical steps during the operation. Seven cases were excluded due to incomplete video recordings to the inability to retrieve systems data. The remaining 87 complete cases were scored on two scales for complexity, the Nassar Scale and the Parkland Grading Scale, with the aim of reducing potential bias by ensuring that the final assessment of each case would be based on a consensus reached between the two scores. Nassar Scale was averaged to score complexity level of three critical steps of each case. Cases with an average Nassar score of 2.5 and above were considered complex. Parkland Grading Scale was used to score the complexity level for the entire case. A

case with a Parkland score of 4 and above were considered complex. Only cases considered standard or complex by both the Nassar and Parkland scoring criteria were assessed in the analyses that required group-wise “standard” and “complex” comparisons. In these tests, cases that were considered complex by one scoring system but standard by the other were excluded ($n = 12$ excluded).

Case-level OPIs were computed with data taken from the surgical videos and both patient- and surgeon-side consoles of the surgical robot for all 87 cases. The OPIs utilized in this study include frequencies and normalized durations of hand controller clutching, arm swaps, camera movements, head movements in and out of the surgical console, clutching behaviors, and energy activation usage of monopolar and bipolar instruments. OPIs specific to either the principal surgeon or the participating fellow (surgeon-specific OPIs) were utilized to determine statistical differences across surgeon experience levels. These surgeon-specific OPIs included (1) arm swap, (2) camera control, (3) head in console, and (4) right and (5) left master clutch. Additionally, OPIs from either single surgeon cases (single console) or teaching cases that utilized both surgical consoles (dual console) are referred to as console-specific OPIs. OPIs for single and dual console cases included energy usage for (1) bipolar cautery, (2) monopolar coagulation, and (3) monopolar cut instruments.

To provide surgical step-specific OPI analyses, surgical videos were segmented into distinct surgical steps. Each surgical step was assigned a task name and corresponding start and stop time by professional surgical annotators. These tasks incorporate patient anatomy and surgical intent and have been standardized by an established annotation card. RC steps utilized in this study include: (1) removal of adhesions surrounding the gallbladder, (2) initial exposure, (3) dissection of Calot’s triangle, (4) ligation/division of cystic artery, (5) ligation/division of cystic duct, (6) dissection of gallbladder off liver bed, (7) specimen removal, and (8) hemostasis of liver bed. Step durations were established for all cases and were utilized for direct comparisons between single and dual console categorizations, as well as standard and complex cases. Step durations were also utilized during the random forest regression analyses for determination of impactful OPIs on both complexity and clinical complication severities.

Relationship establishment across OPIs, complexities, surgeon contributions, and post-surgical complications

OPIs at both the case level and across step-specific segments of the case were utilized in combination with case complexity scores, surgeon experience, console distinctions, and clinical complication severities to establish relevant

relationships across a variety of relevant statistical analyses. For detailed analyses of these relationships, see statistical analysis and results sections.

Statistical analysis

All statistical analyses were performed using SPSS software (Statistical Package for Social Sciences for Windows Version 28) and Scikit-Learn, a free software machine learning library for the Python programming language, Version 1.1.1. Continuous variables were determined as either normally or non-normally distributed with either Kolmogorov–Smirnov or Shapiro–Wilk tests. Categorical variables were represented by frequency [n (%)], while continuous variables were reported as the mean \pm the standard deviation (SD) for normal distributions or the median with interquartile range (IQR) for non-normal distributions. Categorical variables were analyzed using Pearson χ^2 or Fisher’s Exact Test. Continuous variables were binned by complexity and post-operative complication severity and were compared group-wise with independent-sample t -tests with Bonferroni’s correction. To establish OPI impacts on complexity and complication severities, random forest regression analyses between OPIs and complexity and post-operative severity were performed. Both OPIs and step durations were utilized in random forest regression models to establish rankings of the highest impact variables on case complexity and on complication severities. A p -value of 0.05 or less was considered statistically significant.

Results

Patient demographics

Eighty-seven RC cases with a complete video recording were included in the study. Twenty-three of these procedures were identified as teaching cases. Within this cohort, the median (IQR) age was 56 (42–67) with a median (IQR) body mass index of 30 (26–34) kg/m². Patient demographics and medical history are reported in Table 1.

Intra-operative and pathology details

Complicated gallbladders (gangrenous, fistulated, or abscessed) were observed in 15 (17.2%) cases. Decompression of the distended gallbladder was required in 6 (6.9%) cases. Seventeen (19.5%) patients underwent a lysis of adhesions due to prior surgeries or pericholecystic inflammation. Three (3.4%) patients experienced an intra-operative complication: one experienced bleeding due to a slipped cystic artery clip, one experienced liver injury, and one experienced a serosal small bowel injury during lysis of dense

Table 1 Patient demographics and comorbidities

	Cases (<i>n</i> = 87)
Age, year, median (IQR)	56 (42–67)
Sex, male, <i>n</i> (%)	26 (29.9)
BMI, kg/m ² , median (IQR)	30 (26–34)
ASA score, median (IQR)	3 (2–3)
Cardiovascular comorbidities, <i>n</i> (%)	47 (54)
Pulmonary comorbidities, <i>n</i> (%)	15 (17.2)
Smoking, <i>n</i> (%)	14 (16.1)
Diabetes mellitus, <i>n</i> (%)	14 (16.1)
Immunosuppression, <i>n</i> (%)	1 (1.1)
Medications affect coagulation, <i>n</i> (%)	12 (13.8)
Liver disease (hepatitis, fatty liver, cirrhosis), <i>n</i> (%)	9 (10.3)
History of abdominal surgery, <i>n</i> (%)	45 (51.7)
Involving upper abdomen, <i>n</i> (%)	17 (19.5)

IQR interquartile range, BMI body mass index, ASA American Society of Anesthesiologists

Table 2 Intra-operative variables and pathology results

	Cases (<i>n</i> = 87)
Robot model	
DaVinci Si, <i>n</i> (%)	28
DaVinci Xi, <i>n</i> (%)	59
Concomitant procedure, yes, <i>n</i> (%)	4 (4.5)
Teaching cases, <i>n</i> (%)	23 (26.4)
Port setup	
Single port, <i>n</i> (%)	22 (25.3)
Multiport, <i>n</i> (%)	65 (74.7)
Console time, min, median (IQR)	29 (15–47)
Skin-to-skin time, min, median (IQR)	42 (26–64)
Estimated blood lost, mL, median (IQR)	5 (5–5)

IQR interquartile range

adhesions. There were no conversions to the open approach. Additional operative details and pathology results are summarized in Table 2.

Post-operative complications and details

Of the 87 analyzed cases, 50 (57.5%) patients were discharged on the same day of their surgery, with a 2-day median hospital LOS (0–3 day range). Fourteen (16.1%) patients were lost to follow-up due to the absence of any post-operative office or telehealth visits. The average follow-up time was 15.6 (max: 55) months for the entire cohort. In total, 13 complications (17.8%) were observed, of which 7 were Clavien–Dindo grade-I and II, indicating more minor complications than grades III–V. In these 13 cases, 9 (10.3%) patients revisited the emergency

Table 3 Post-operative outcomes

	Cases (<i>n</i> = 73)
Overall complications	13 (17.8)
Surgery related complication, <i>n</i> (%)	4 (5.5)
Clavien–Dindo highest grades	
Grade-1, <i>n</i> (%)	4 (5.5)
Grade-2, <i>n</i> (%)	3 (4.1)
Grade-3A, <i>n</i> (%)	2 (2.7)
Grade-3B, <i>n</i> (%)	1 (1.4)
Grade-4A, <i>n</i> (%)	2 (2.7)
Grade-4B, <i>n</i> (%)	1 (1.4)
Comprehensive complication index, mean (min–max)	5.0 (0–100)

Table 4 Standard, complex, and mutual agreement sample sizes categorized by Nassar and Parkland complexity grades

	Nassar	Parkland	Mutual agreement
Standard	56	59	51
Complex	31	28	24
Total	87	87	75

department within 30 days following their surgery. Of these, 4 (4.6%) patients were readmitted to the hospital. Post-operative complications and their severity are represented in Table 3.

Complexity scoring and its relationship to post-operative complication severity

All cases were categorized as standard or complex with both the Nassar and Parkland scoring criteria, but only those cases in which both scoring criteria agreed on the complexity categorization were included in complexity analyses (Table 4). Of these, 75 cases were utilized as either standard or complex.

To determine the relationship between post-operative Clavien–Dindo complication scores and case complexity, cases that reported any post-operative complications (Table 3) with a consistent complexity categorization across both Nassar and Parkland scores (Table 4) were analyzed (*n* = 75). Cases with a reported CCI co-morbidity score were also compared in both standard and complex cases, as case complexity and difficulty are highly correlated with patient co-morbidity [19]. As anticipated, across both the Clavien–Dindo and CCI scales, both scores were significantly higher in complex cases compared to standard cases (Fig. 1, $0.001 \leq p \leq 0.01$), indicating a greater chance of post-operative complications during a more complex RC case.

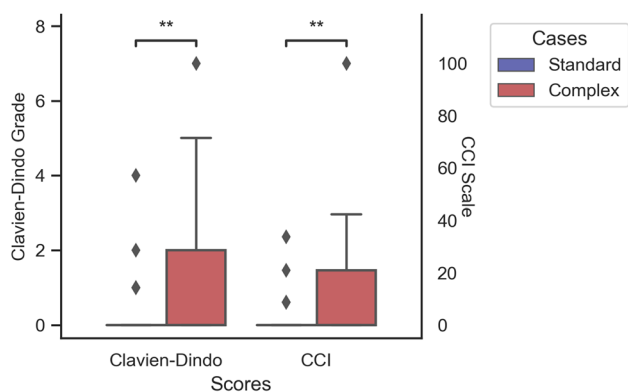


Fig. 1 Post-operative complication and co-morbidity scores vs. complexity. Complex RC cases exhibit higher post-surgical complication and co-morbidity scores than standard RC cases. Pair-wise comparisons of both Clavien–Dindo complexity and CCI co-morbidity scores across standard and complex RC cases. *Indicates $0.001 \leq p \leq 0.01$

OPI differences observed in both surgeon-specific and complexity comparisons

To determine differences in surgical experience and case complexities, direct comparisons of arm swap durations, camera control durations and frequencies, percentages of time surgeons kept their head in the surgical console, and clutching frequency and duration were performed across all groups (Fig. 2; Table 5). Significant differences were found between the principal surgeon and the clinical fellows. In standard cases, arm swapping from the principal surgeon was performed at a significantly shorter duration than the fellows (Fig. 2a), indicating a better utilization of the robotic arms. A similar behavior was observed for the camera control (Fig. 2b) and master clutching behaviors (Fig. 2e), in which the principal surgeon exhibited decreased camera control and clutching durations and increased frequencies, consistent with short, frequent movements upon mastery. In complex cases, no statistically significant differences were observed between the principal surgeon and fellows.

There were differences observed in comparisons across complexity levels, as well. A decrease in the relative time the principal surgeon's head was spent in the surgical console was observed in complex cases when compared to standard cases (Fig. 2c), and an increased camera control frequency from the principal surgeon during complex cases was reported, as well (Fig. 2b). No differences were observed in actions of the fellows across case complexity levels (Fig. 2).

Energy use differences in single console non-teaching and dual console teaching cases

To determine changes in energy use across cases that are performed individually with a single console or as a dual

console teaching case energy, OPIs across three energy types were analyzed: (1) monopolar coagulation energy usage, (2) monopolar cut energy usage, and (3) bipolar cautery energy usage (Fig. 3; Table 6). In standard cases, monopolar coagulation energy activation durations were longer in single console cases compared to teaching dual cases control, which could indicate a higher degree of confidence witnessed with an experienced surgeon. A longer median duration of energy activation was also noted in complex single console cases compared to standard ones for the principal surgeon only (Fig. 3), indicating that once mastered, higher complexity cases will require greater energy expenditure levels to ensure a proper lysis of adhesion and anatomical dissection. Energy OPIs across monopolar cut energy usage and bipolar energy usages were also analyzed, but no statistical differences were observed (Table 6).

Surgical step-specific analysis provides insight into case complexity

Step-specific durations reveal differences across standard and complex RC cases

To investigate surgical step-specific metrics and their impact on case complexities and console attributions, surgical step durations were compared across complexities in both single and dual console cases. Upon analysis, five surgical steps resulted in duration differences for either complexity, single or dual console contributions, or both (Fig. 4). As shown in Fig. 4, the removal of adhesions around the gallbladder step took significantly longer for both single and dual console cases in complex procedures. The dissection of Calot's triangle and of the gallbladder off the liver bed steps displayed a similar pattern, increased significantly with complexity in single consoles cases, and also showed a significant increase in duration in standard, dual console cases when compared to only a single console case. The cystic artery ligation task and the specimen removal task both displayed longer durations in dual console standard cases only. Initial exposure, cystic duct ligation, and hemostasis of liver bed tasks displayed no significant changes. Together, this provides feasibility for surgical step-specific metrics, specifically step duration, as specific and truly objective metrics to uncover novel insights regarding console attribution and case complexities.

Regression analysis identifies relative OPI impacts on case complexity

To establish a more comprehensive characterization of the relationships across OPIs and case complexity, a random forest regression analysis was performed. This analysis combines both surgeon- and console-specific OPIs with surgical step

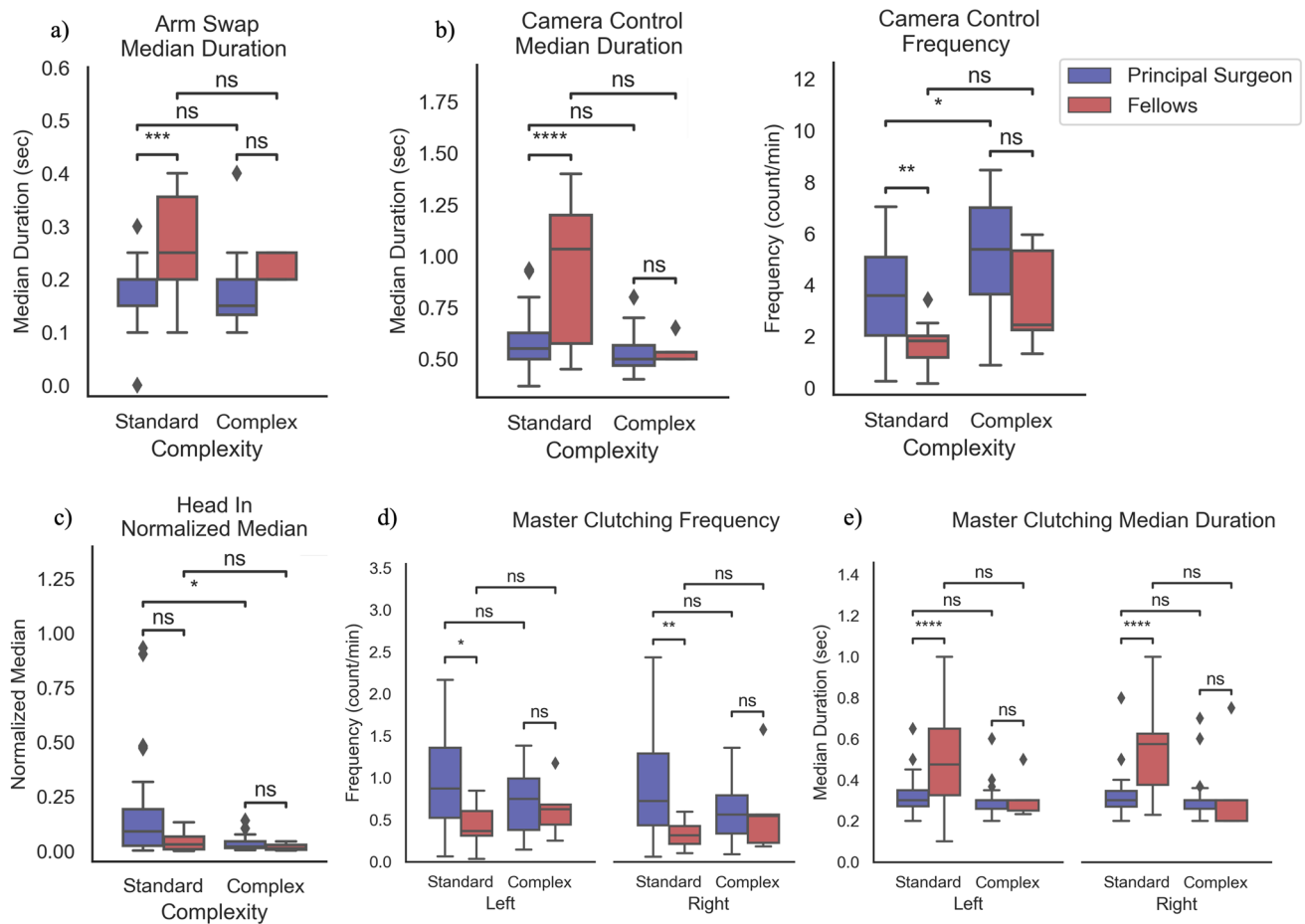


Fig. 2 OPI comparisons categorized by surgeon experience and case complexity. **a** Arm swap median values normalized by case duration exhibit significant differences in standard cases when the principal surgeon and fellows are compared, but no other differences are observed, **b** camera control median duration and frequency exhibit significant differences in standard cases when the principal surgeon and fellows are compared, as well as in the principal surgeon's camera control across standard and complex cases, **c** head-in median

duration normalized by case duration exhibits higher durations during standard cases for the principal surgeon compared to complex cases, **d** master clutch frequency exhibits higher clutching rates between the principal surgeon and the fellows in both standard and complex cases, and **e** master clutch median duration shows longer durations for the principal surgeon compared to the fellows in both standard and complex cases. *ns* not significant, *indicates $0.01 < p \leq 0.05$, **indicates $0.001 < p \leq 0.01$, ****indicates $p \leq 0.0001$

durations to provide a comprehensive, unbiased analysis of potential features of importance. These OPIs were ranked and their relative impacts on complexity scores across both Nassar and Parkland grades were computed. This analysis reveals that the duration of the removal of adhesions step exhibits the highest indication for complexity across all variables utilized (Fig. 5). Other OPIs of importance include case duration, Calot's triangle dissection duration, dissection of gallbladder off liver bed, and arm swap frequency of the principal surgeon. This analysis enabled the detection of critical surgical steps in higher complexity cases.

Regression analysis across all OPIs identifies intra-operative surgical indicators of complication severities

Objective identification of the relationships between OPIs and postsurgical complications is critical to shed light into intra-operative variables that could contribute to better outcomes. As such, weighted rankings of OPIs were established with a random forest regression analysis. These OPI rankings reflect their impact on post-surgical complication scores and co-morbidity scores. Results indicate that the total case duration was the greatest indicator of the Clavien–Dindo

Table 5 Grouped comparison *p*-values across surgical complexity and surgeon experience

OPIs	Standard vs. complex		Principal surgeon vs. fellow	
	Principal surgeon	Fellow	Standard	Complex
Arm swap				
Frequency	1.000	0.397	0.734	1.000
Total duration	1.000	1.000	1.000	1.000
Median duration	1.000	1.000	<0.001	0.619
Camera control				
Frequency	0.022	0.082	0.006	0.712
Total duration	0.577	1.000	0.138	0.865
Median duration	0.137	0.136	<0.001	1.000
Head in				
Frequency	0.809	0.605	0.067	1.000
Total duration	1.000	1.000	1.000	0.388
Median duration	0.027	0.728	0.096	0.845
Left Master clutch				
Frequency	0.351	0.756	0.012	1.000
Total duration	0.270	1.000	0.532	1.000
Median duration	1.000	0.515	<0.001	1.000
Right Master clutch				
Frequency	0.289	0.492	0.005	1.000
Total duration	0.143	1.000	0.326	1.000
Median duration	1.000	0.523	<0.001	1.000

Values in bold indicate a *p*-value < 0.05
OPIs objective performance indicators

complication score, whereas hemostasis of liver bed showed the highest impact on highly correlated CCI co-morbidity scale (Fig. 6). Other OPIs of interest include principal surgeon camera control and arm swap frequencies, the duration of the removal of adhesions step, the specimen removal step, and Calot’s triangle dissection duration.

Discussion

Although considered a routine procedure, robotic cholecystectomy can be difficult and pose a high risk of complications dependent on patient history [20] and case complexity, but the impact of surgical behaviors that can affect these outcomes remains elusive. This study establishes associations between surgical behavior based OPIs, participation of clinical fellows, case complexities, and post-surgical complication severities in RC. Through a thorough and holistic analysis of nearly 90 RC cases performed by a single surgeon, we report specific differences in OPIs between an experienced surgeon and clinical fellows across case complexities. Additionally, we identify specific surgical tasks and their significance across complexities. Critically, we provide the first weighted ranking of OPIs and their impacts on both case complexity and post-surgical complication scores.

Across comparisons of arm swapping, camera movement, head in console, and clutching behavior OPIs, we report many significant differences in standard cases between experienced surgeons and learning clinical fellows across these

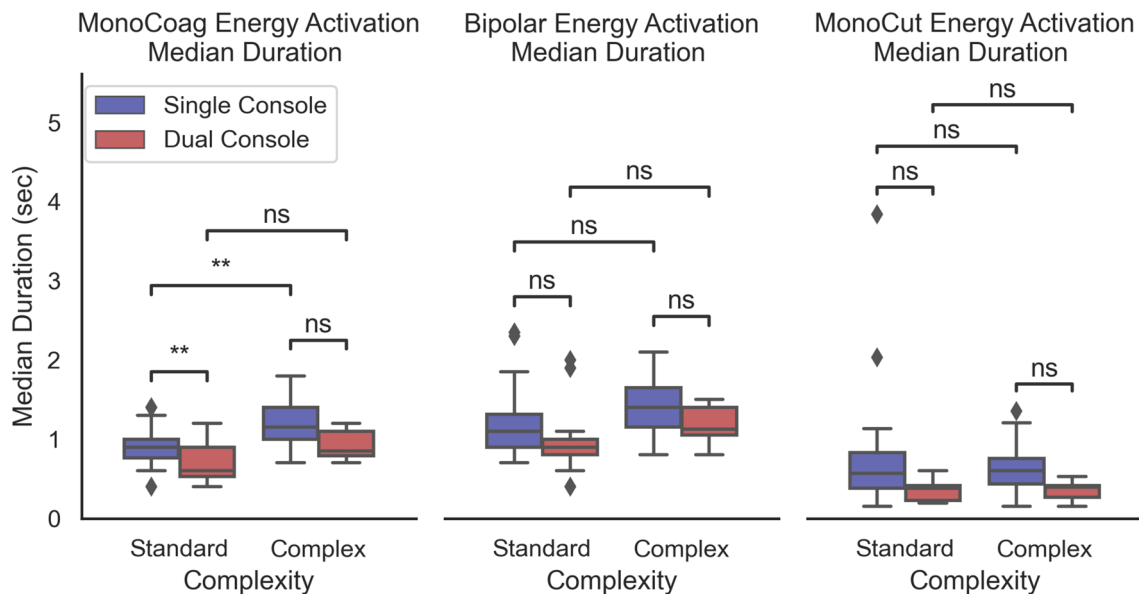


Fig. 3 Monopolar coagulation energy activation differences observed across complexities and teaching designations. Figure depicts the median duration of monopolar activation energy differences in single and dual console cases in both standard and complex cases. Energy

use was significantly less in dual console than single console cases in standard cases only, where single console cases also exhibited an increase in monopolar energy use duration in complex cases compared to standard cases. **Indicates 0.001 < *p* ≤ 0.01

Table 6 Statistical analysis results across complexities and console participation in energy OPIs

OPIs	Standard vs. complex		Single console vs. Dual console	
	Single console	Dual console	Standard	Complex
Monopolar coagulation energy				
Frequency	0.328	0.244	1.000	1.000
Total duration	1.000	1.000	0.844	0.652
Median duration	0.001	0.085	0.010	0.097
Bipolar cautery energy				
Frequency	0.567	1.000	0.299	1.000
Total duration	0.222	1.000	1.000	1.000
Median duration	0.413	0.815	0.269	0.539
Monopolar cut energy				
Frequency	1.000	0.836	1.000	0.583
Total duration	1.000	1.000	0.608	0.453
Median duration	1.000	1.000	1.000	0.191

Values in bold indicate a p -value < 0.05

OPIs objective performance indicators

OPIs, as well as differences across complexity levels for the principal surgeon. However, no changes across groups were reported in complex cases or for fellows across complexities. This same pattern is observed in monopolar energy activation usage across individual surgeon cases and teaching cases provided in this study. Although each result can be interpreted independently, experienced surgeons exhibited a better utilization of the arms and endoscope possibly due to their greater understanding of the robotic system, the anatomy of the patient, and the surgical procedure. They are able to anticipate where the instruments need to be positioned and how to move them efficiently, and thus their comportment may vary minimally across standard cases compared to novice clinical fellows. Oppositely, complex cases may reveal more nuanced behaviors for experienced surgeons potentially due to the operator deviating from their routine chronology in order to accommodate for unexpected occurrences or challenging tasks.

To achieve a profound understanding of RC, the procedures were divided into eight distinct surgical steps defined in a manner capable of delimiting the clinical step duration. In our study, we establish for the first time an association between RC step-specific duration, complexity and clinical outcomes. Removal of adhesions and hemostasis of the liver bed were highly correlated with increased complexity and worse clinical outcomes, respectively. Adhesions are a common finding in patients with a surgical history [21–23] and may be associated with organ dysfunction and often reoperation [24, 25]. Consequently, surgeons may be faced with an occasionally lengthy adhesiolysis to minimize the difficulty of a proper anatomical dissection. Our analysis determined lysis of adhesions to be a lengthier task with the increasing complexity of the

cases. Moreover, the presence of adhesions in complex cases may impede subsequent surgical steps such as the dissection of the gallbladder off the liver bed and of Calot's triangle. The latter step was particularly longer in complex cases and has been previously demonstrated to be a predictor of laparoscopic cholecystectomy difficulty [26]. Finally, bleeding during a cholecystectomy can arise at any time, and has been reported to occur in up to 10% of all cases [27]. It can range from minor bleeding to life-threatening injury to major vessels and therefore needs proper and prompt control.

Through a more granular approach to surgical assessment, our step-specific analysis, especially in term of duration, may enable surgeons to focus on particular areas that require increased attention. Moreover, by breaking down the surgical steps and analyzing each one separately, researchers can gain a more detailed understanding of the surgical process and identify areas for improvement. Beyond the establishment of OPI relationships to case complexity and patient outcomes, this study also highlights the inherent potential of OPIs to provide comprehensive evaluations of RC surgeries by contextualizing surgical performance in relation to levels of experience, case complexity and patient outcomes. Therefore, the implementation of this cutting-edge technology that is capable of analyzing surgeons' behaviors during procedures, should be pursued across institutions and throughout training programs [28]. Moreover, the American Board of Surgery is implementing new curricula across all surgery residency programs that bases assessments on competency with Entrusted Professional Activities (EPAs) [29, 30]. Combining EPAs with the appropriate digital tools will provide regular, time-efficient, feedback-oriented, and workplace-based assessments during routine clinical tasks.

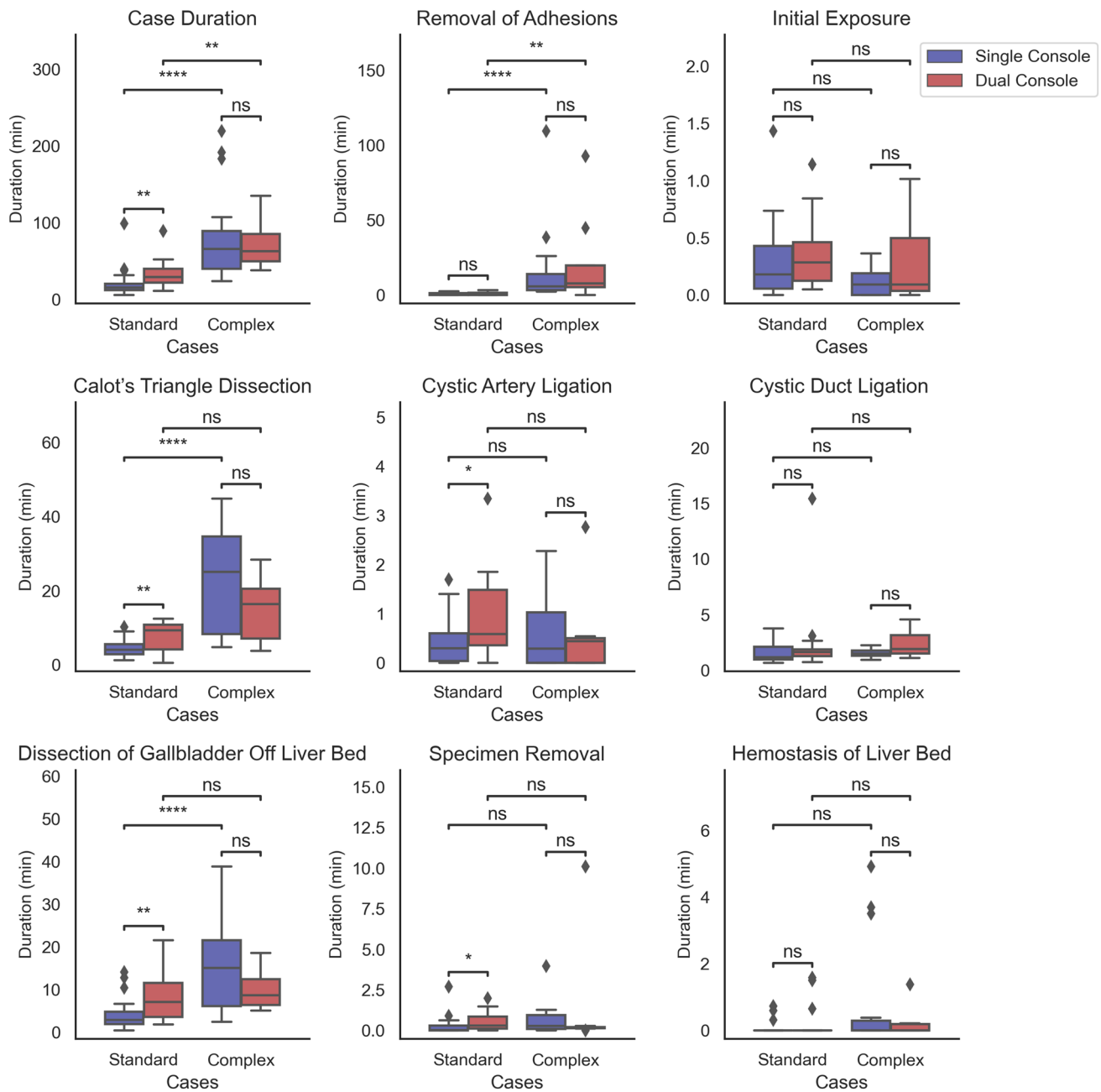


Fig. 4 Case and surgical step durations across complexities and console participations. Each panel depicts group-wise duration comparisons of either the full case duration (top left panel) or each surgical step duration across complexities and surgical console participations (single console case or dual console teaching case). Statistically significant differences are observed for the full case duration, removal of

adhesions step, the dissection of Calot's triangle step, the cystic artery ligation step, the dissection of the gallbladder off the liver bed step, and the specimen removal step. No statistically significant differences were found across the remaining steps, indicated ns. *Indicates $0.01 < p \leq 0.05$, **indicates $0.001 < p \leq 0.01$, ****indicates $p \leq 0.0001$

A transition from subjective scoring metrics to machine learning (ML) enabled evaluation of OPIs in surgery is critical to enhance performance assessment in surgery and correlate it with clinical outcomes. Toward these efforts, several investigators have already started to evaluate the value of ML in OPIs [31]. Expert panels envisage AI to

aid in anatomical recognition directly during the surgical procedures and to provide extensive performance feedback for surgeons directly after any procedures [32]. However, the increasing popularity of ML and deep learning architectures requires a substantial amount of intra-operative data and proper digitization of operating rooms. Addressing these

Weighted OPI Impact on Complexity

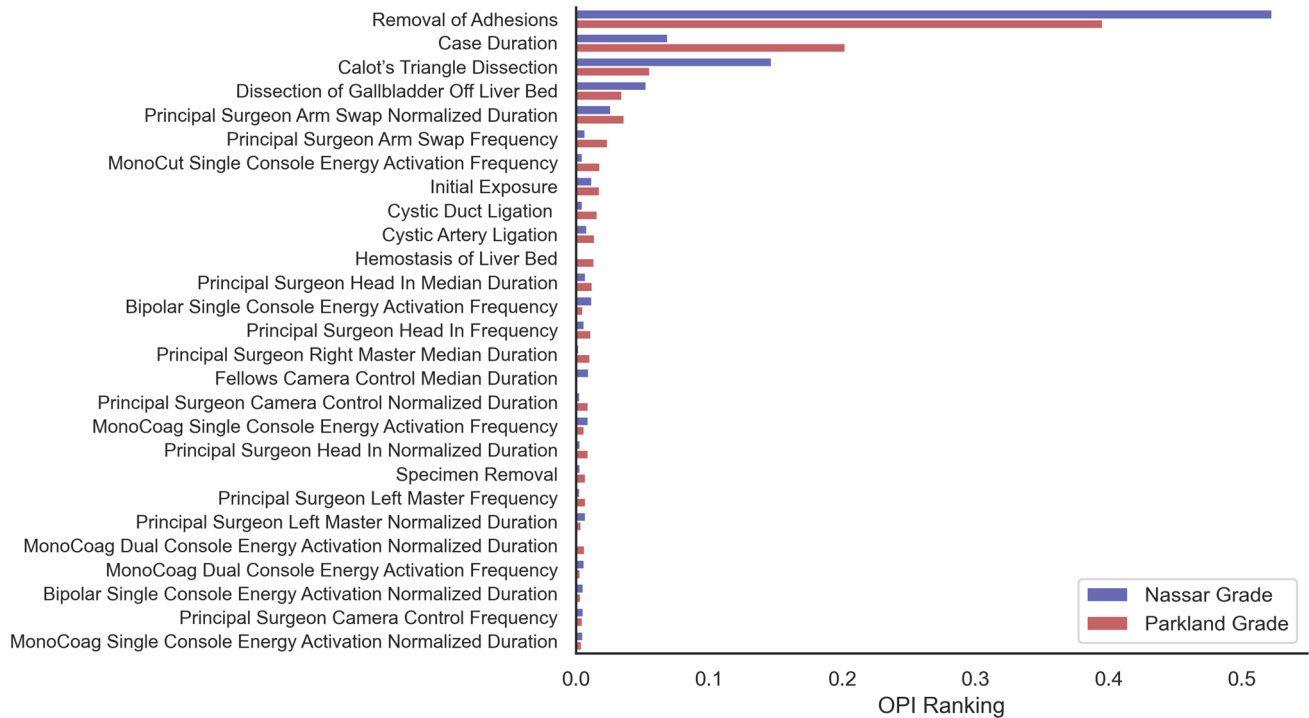


Fig. 5 Complexity regression of OPI rankings of most impactful indicators in case complexity. Figure depicts the OPI rankings averaged over 100 random forest estimators. All estimators indicate that removal of adhesions and overall case duration to be the most impor-

tant OPIs to indicate case complexity. Fit statistics: Nassar grade coefficient of correlation=0.96, p -value=0.01; Parkland grade coefficient of correlation=0.94, p -value=0.01

challenges and establishing collaborations between different research groups will resolve issues in ML for surgical performance assessment and advance the emerging surgical data science field.

Although our study has several limitations, it provides many opportunities for future studies. Its retrospective nature limits the data available for evaluation, but the ability to calculate OPIs retrospectively could enrich other video datasets for similar projects in other specialties. Additionally, the generalizability of our findings to other providers may be limited as we conducted an analysis on a small sample size of nearly 90 surgical videos paired to patient demographics and outcomes. However, by focusing on the performance of one principal surgeon, we have potentially eliminated the added layer of variability typically generated by the presence of multiple surgeons, thus rendering the analysis of our cohort sensitive to relevant findings. With the hope of establishing a link between OPIs and clinical outcomes, the fit statistics for patient outcome were lower compared to that of complexity, potentially due to the limited sample size of cases that resulted in severe post-operative complications, as most of the recorded complications were minor. Lastly,

as mentioned above, the step-specific analyses were limited to step duration, but their significance in subsequent tests of complexity and complication severities highlight how fundamental studies that investigate a wider breadth of OPIs within surgical steps will be in future studies. With these limitations in mind, this study provides a critical foundation for future studies of OPIs, and our team is primed to pursue additional projects with the compiled and paired surgical, clinical, and patient data established.

Together, this study provides the building blocks for specific, consistent, efficient, and personalized objective feedback for surgeons performing RC procedures at all training levels, tying RC case complexity, surgeon experience, and complications to cases and specific steps of cases. As we embrace the digitization of surgery, this work paves the way for future breakthroughs in machine learning to build upon our findings to one day predict case complexity and surgeon experience in robotic cholecystectomy. Ultimately, our findings and future studies will enable powerful OPI-based digital tools for scalable feedback to surgeons to optimize surgical practice and lead to better outcomes for patients.

OPI Ranking with Post-Operative Complications and Co-Morbidity Scores

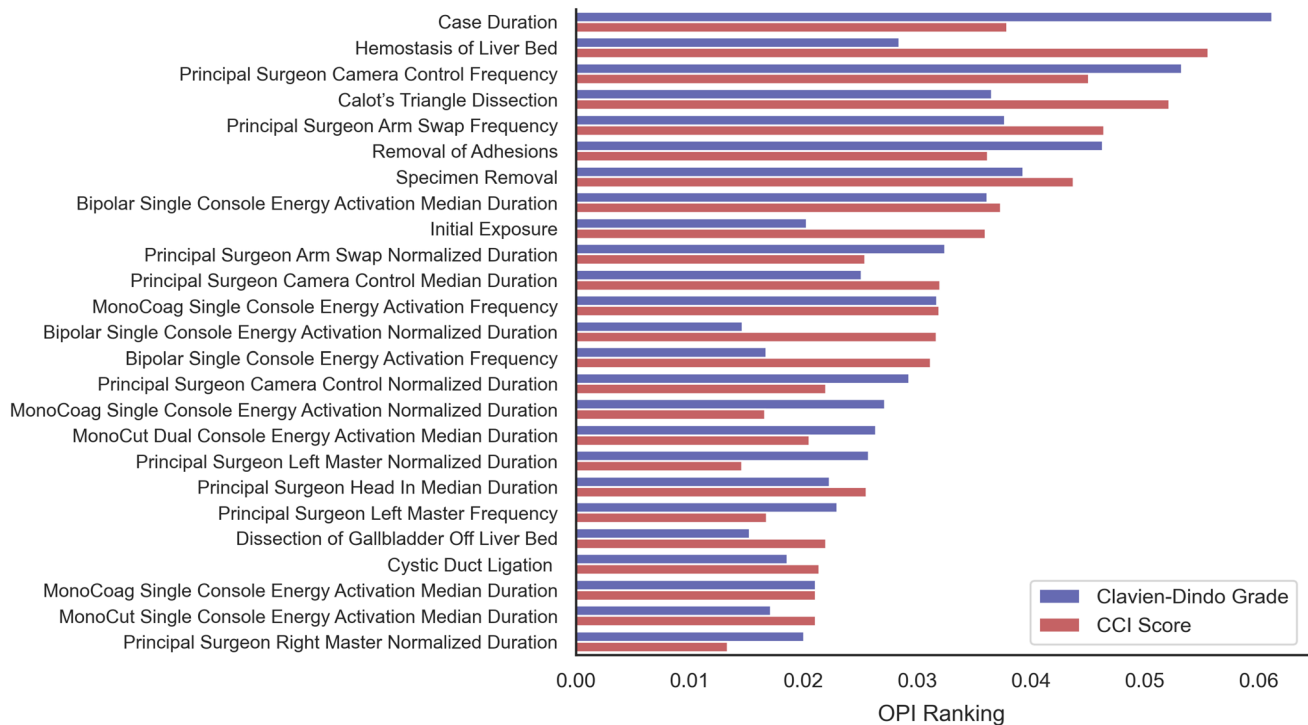


Fig. 6 Rankings of OPIs and their impact on patient complication severities. OPI rankings averaged over 100 random forest estimators indicate that overall case duration and the camera control behavior of the principal surgeon were the most impactful OPIs for post-operative

complication severity. Fit statistics: Clavien–Dindo Grade coefficient of correlation = 0.852, $p = 0.099$; CCI Score coefficient of correlation = 0.847, $p = 0.089$

Acknowledgements The authors acknowledge Michelle Liu and Busisiwe Mlambo, MD for their help in the surgical step annotations and video breakdown.

Funding No funding was received for this study.

Declarations

Disclosures Drs. Georges Kaoukabani, Alexander Friedman, and Fahri Gokcal have no conflicts of interest or financial ties to disclose. Abeslalom Fanta, Xi Liu, and Mallory Shields are employees at Intuitive Surgical and do not report any other conflict of interest of financial ties. Dr. Omar Yusef Kudsi has received a teaching course and/or consultancy fees from Intuitive Surgical, Bard-Davol, and W.L. Gore outside the submitted work.

Ethical approval The database used for this study approved by the Institutional Review Board.

Human and animal rights All procedures performed in studies involving human participants were under the ethical standards of the Institutional and/or National Research Committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Vivek MA, Augustine AJ, Rao R (2014) A comprehensive predictive scoring method for difficult laparoscopic cholecystectomy. *J Minim Access Surg* 10(2):62–67. <https://doi.org/10.4103/0972-9941.129947>
- Griffiths EA, Hodson J, Vohra RS et al (2019) Utilisation of an operative difficulty grading scale for laparoscopic cholecystectomy. *Surg Endosc* 33(1):110–121. <https://doi.org/10.1007/s00464-018-6281-2>
- Madni TD, Leshikar DE, Minshall CT et al (2018) The Parkland grading scale for cholecystitis. *Am J Surg* 215(4):625–630. <https://doi.org/10.1016/j.amjsurg.2017.05.017>
- Nassar AHM, Hodson J, Ng HJ et al (2020) Predicting the difficult laparoscopic cholecystectomy: development and validation of a pre-operative risk score using an objective operative difficulty grading system. *Surg Endosc* 34(10):4549–4561. <https://doi.org/10.1007/s00464-019-07244-5>
- Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ (2012) Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol* 187(1):247–252. <https://doi.org/10.1016/j.juro.2011.09.032>
- Aghazadeh MA, Jayaratna IS, Hung AJ et al (2015) External validation of Global Evaluative Assessment of Robotic Skills (GEARS). *Surg Endosc* 29(11):3261–3266. <https://doi.org/10.1007/s00464-015-4070-8>
- Martin JA, Regehr G, Reznick R et al (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. *Br*

- J Surg 84(2):273–278. <https://doi.org/10.1046/j.1365-2168.1997.02502.x>
8. Fard MJ, Ameri S, Darin Ellis R, Chinnam RB, Pandya AK, Klein MD (2018) Automated robot-assisted surgical skill evaluation: predictive analytics approach. *Int J Med Robot*. <https://doi.org/10.1002/rcs.1850>
 9. Jarc AM, Curet MJ (2017) Viewpoint matters: objective performance metrics for surgeon endoscope control during robot-assisted surgery. *Surg Endosc* 31(3):1192–1202. <https://doi.org/10.1007/s00464-016-5090-8>
 10. Lyman WB, Passeri M, Murphy K et al (2018) Novel objective approach to evaluate novice robotic surgeons using a combination of kinematics and stepwise cumulative sum analyses. *J Am Coll Surg* 227(4):S223–S224
 11. Hung AJ, Oh PJ, Chen J et al (2019) Experts vs super-experts: differences in automated performance metrics and clinical outcomes for robot-assisted radical prostatectomy. *BJU Int* 123(5):861–868. <https://doi.org/10.1111/bju.14599>
 12. Lyman WB, Passeri MJ, Murphy K et al (2021) An objective approach to evaluate novice robotic surgeons using a combination of kinematics and stepwise cumulative sum (CUSUM) analyses. *Surg Endosc* 35(6):2765–2772. <https://doi.org/10.1007/s00464-020-07708-z>
 13. Ma R, Lee RS, Nguyen JH et al (2022) Tailored feedback based on clinically relevant performance metrics expedites the acquisition of robotic suturing skills—an unblinded pilot randomized controlled trial. *J Urol* 208(2):414–424. <https://doi.org/10.1097/JU.0000000000002691>
 14. Hung AJ, Chen J, Jarc A, Hatcher D, Djaladat H, Gill IS (2018) Development and validation of objective performance metrics for robot-assisted radical prostatectomy: a pilot study. *J Urol* 199(1):296–304. <https://doi.org/10.1016/j.juro.2017.07.081>
 15. Hung AJ, Chen J, Che Z et al (2018) Utilizing machine learning and automated performance metrics to evaluate robot-assisted radical prostatectomy performance and predict outcomes. *J Endourol* 32(5):438–444. <https://doi.org/10.1089/end.2018.0035>
 16. Brown KC, Bhattacharyya KD, Kulason S, Zia A, Jarc A (2020) How to bring surgery to the next level: interpretable skills assessment in robotic-assisted surgery. *Visc Med* 36(6):463–470. <https://doi.org/10.1159/000512437>
 17. Dindo D, Demartines N, Clavien PA (2004) Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 240(2):205–213. <https://doi.org/10.1097/01.sla.0000133083.54934.ae>
 18. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien PA (2013) The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg* 258(1):1–7. <https://doi.org/10.1097/SLA.0b013e318296c732>
 19. Mehta HB, Dimou F, Adhikari D et al (2016) Comparison of comorbidity scores in predicting surgical outcomes. *Med Care* 54(2):180–187. <https://doi.org/10.1097/mlr.0000000000000465>
 20. Salky BA, Edye MB (1998) The difficult cholecystectomy: problems related to concomitant diseases. *Semin Laparosc Surg* 5(2):107–114. <https://doi.org/10.1177/155335069800500205>
 21. Lauder CI, Garcea G, Strickland A, Maddern GJ (2010) Abdominal adhesion prevention: still a sticky subject? *Dig Surg* 27(5):347–358. <https://doi.org/10.1159/000314805>
 22. Weibel MA, Majno G (1973) Peritoneal adhesions and their relation to abdominal surgery. A postmortem study. *Am J Surg* 126(3):345–353. [https://doi.org/10.1016/s0002-9610\(73\)80123-0](https://doi.org/10.1016/s0002-9610(73)80123-0)
 23. Menzies D, Ellis H (1990) Intestinal obstruction from adhesions—how big is the problem? *Ann R Coll Surg Engl* 72(1):60–63
 24. Ten Broek RP, Issa Y, van Santbrink EJ et al (2013) Burden of adhesions in abdominal and pelvic surgery: systematic review and met-analysis. *BMJ* 347:f5588. <https://doi.org/10.1136/bmj.f5588>
 25. Nkere UU (2000) Postoperative adhesion formation and the use of adhesion preventing techniques in cardiac and general surgery. *ASAIO J* 46(6):654–656. <https://doi.org/10.1097/00002480-200011000-00003>
 26. Bhandari TR, Khan SA, Jha JL (2021) Prediction of difficult laparoscopic cholecystectomy: an observational study. *Ann Med Surg (Lond)* 72:103060. <https://doi.org/10.1016/j.amsu.2021.103060>
 27. Kaushik R (2010) Bleeding complications in laparoscopic cholecystectomy: incidence, mechanisms, prevention and management. *J Minim Access Surg* 6(3):59–65. <https://doi.org/10.4103/0972-9941.68579>
 28. Melnyk R, Chen Y, Holler T et al (2022) Utilizing head-mounted eye trackers to analyze patterns and decision-making strategies of 3D virtual modelling platform (IRIS™) during preoperative planning for renal cancer surgeries. *World J Urol* 40(3):651–658. <https://doi.org/10.1007/s00345-021-03906-z>
 29. Lindeman B, Brasel K, Minter RM, Buyske J, Grambau M, Sarosi G (2021) A phased approach: the general surgery experience adopting entrustable professional activities in the United States. *Acad Med* 96(7s):S9–S13. <https://doi.org/10.1097/acm.00000000000004107>
 30. American Board of Surgery (2022) ABS announces transition to entrustable professional activities for general surgery resident evaluation. https://www.absurgery.org/default.jsp?news_epas0222. Accessed 7 March 2023
 31. Lam K, Chen J, Wang Z et al (2022) Machine learning for technical skill assessment in surgery: a systematic review. *NPJ Digit Med* 5(1):24. <https://doi.org/10.1038/s41746-022-00566-0>
 32. Vedula SS, Ghazi A, Collins JW et al (2022) Artificial intelligence methods and artificial intelligence-enabled metrics for surgical education: a multidisciplinary consensus. *J Am Coll Surg* 234(6):1181–1192. <https://doi.org/10.1097/xcs.0000000000000190>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.