Expert Consensus-based Video-Based Assessment Tool for Workflow Analysis in Minimally Invasive Colorectal Surgery: Development and Validation of ColoWorkflow

Pooja P Jain[a,b], Pietro Mascagni[a,c,*], Giuseppe Massimiani[a,c], Nabani Banik[b], Marta Goglia[d], Lorenzo Arboit[b], Britty Baby[b], Andrea Balla[e], Ludovica Baldari[f], Gianfranco Silecchia[d], Claudio Fiorillo[c], CompSurg Colorectal Experts Group, Sergio Alfieri[c], Salvador Morales-Conde[e], Deborah S Keller[g,h], Luigi Boni[f], Nicolas Padoy[a,b]

[a] *IHU Strasbourg, Strasbourg, France*
[b] *University of Strasbourg, CNRS, INSERM, ICube, UMR7357, Strasbourg, France*
[c] *Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy*
[d] *Department of Medical-Surgical Sciences and Translational Medicine, Faculty of Medicine and Psychology, Sapienza University of Rome, Italy*
[e] *University Hospital Virgen Macarena, University of Sevilla, Seville, Spain*
[f] *Department of General and Minimally Invasive Surgery, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*
[g] *Arizona State University, Tempe, AZ, USA*
[h] *Mayo Clinic, Phoenix, AZ, USA*

Background: Minimally invasive colorectal surgery is characterized by significant procedural variability, a difficult learning curve, and complications that impact patients' outcomes. Video-based assessment (VBA) offers an opportunity to generate data-driven insights to reduce variability, optimize training, and improve surgical performance. This study aims to develop and validate a VBA tool for workflow analysis across minimally invasive colorectal procedures.

Methods: A three-round modified Delphi process was conducted among experts in colorectal surgery (CRS) and VBA to achieve consensus on generalizable workflow descriptors. The resulting framework informed the development of a new VBA tool, *ColoWorkflow*. Independent raters then applied *ColoWorkflow* to a multicentre video dataset of laparoscopic and robotic CRS. Applicability and inter-rater reliability were evaluated.

Results: Consensus was achieved for 10 procedure-agnostic phases and 34 procedure-specific steps describing CRS workflows. *ColoWorkflow* was developed and applied to 54 colorectal operative videos (left and right hemicolectomies, sigmoid and rectosigmoid resections, and total proctocolectomies) from five centres in four countries. The tool demonstrated broad applicability, with all but one label utilized. Inter-rater reliability was moderate, with mean Cohen's κ of 0.71 for phases and 0.66 for steps. Most discrepancies arose at phase transitions and step boundary definitions.

Conclusions: *ColoWorkflow* is the first consensus-based, validated VBA tool for comprehensive workflow analysis in minimally invasive colorectal surgery. It establishes a reproducible framework for video-based performance assessment, enabling benchmarking across institutions and supporting the development of artificial intelligence-driven workflow recognition. Its adoption may standardize training, accelerate competency acquisition, and advance data-informed surgical quality improvement.

1. Introduction

Minimally invasive surgery (MIS) has become the gold standard approach for benign and malignant elective colorectal resections due to its reduced morbidity, shorter hospital stays, and faster recovery compared with open surgery [1, 2]. Despite these advantages, outcomes after MIS colorectal surgery (CRS) remain highly variable across institutions and surgeons [3]. Procedural variability, difficult learning curves, and persistent complication rates underscore ongoing challenges in achieving consistent technical performance [4, 5, 6]. Even among experienced centres and surgeons, heterogeneity in intraoperative techniques and dissection sequences contributes to variation in operative time, blood loss, complication risks, and postoperative recovery [7]. Addressing this variability is critical to facilitating training, benchmarking performance, standardizing care, and in improving quality of MIS CRS.

Video-based assessment (VBA) has emerged as a robust method to evaluate and improve surgical procedures. Unlike direct observation, VBA enables asynchronous, detailed review of intraoperative workflows and critical events, supporting structured feedback and postoperative debriefing [8, 9]. Prior studies have demonstrated that technical performance on surgical video correlates with patient outcomes, providing an objective link between intraoperative skill and clinical results [10, 11, 12, 13]. However, the expansion of VBA programs has been limited by the scarcity of validated assessment tools, and the significant time and expertise required for manual video review. While advances in computer vision and artificial intelligence (AI) promise to automate aspects of surgical video analysis, these computational methods still require to be trained and validated on well-defined and consistent VBA tools [14].

A standardized VBA tool for workflow analysis could quan-

*Corresponding author: `pietro.mascagni@ihu-strasbourg.eu`

tify procedural variability in MIS CRS, identify best practices, and establish the foundation for automated workflow recognition.

This study aimed to develop and validate a VBA tool for comprehensive workflow characterization in minimally invasive colorectal surgery. We hypothesized that a consensus-derived, procedure-agnostic framework could reliably describe workflows across laparoscopic and robotic colorectal operations and enable reproducible VBA.

2. Materials and Methods

2.1. Study design

This study comprised two components: (1) a modified Delphi process to establish expert consensus on workflow descriptors for minimally invasive colorectal surgery (MIS CRS), and (2) the development and validation of a video-based assessment (VBA) tool, ColoWorkflow, derived from the consensus framework. The Delphi process was conducted and reported in accordance with the ACCORD (A Consensus-Based Checklist for Reporting of Delphi Studies) guidelines [15].

2.2. Delphi consensus on colorectal surgery workflow descriptors

A three-round modified Delphi study was undertaken to achieve consensus on the phases and steps describing workflows in minimally invasive colorectal surgery. A multidisciplinary steering committee of seven members with expertise in colorectal surgery, VBA, and artificial intelligence (AI) oversaw the process. Expert participants were peer-nominated based on demonstrated experience in colorectal surgery and/or video-based assessment.

The steering committee developed initial workflow definitions through a comprehensive review of key surgical textbooks and literature [16, 17, 18, 19, 20, 21, 22, 23], semi-structured interviews with colorectal surgeons, and a hierarchical task analysis (HTA) to deconstruct procedures into phases and steps. Definitions followed the SAGES consensus framework, in which *phases* represent the highest-level temporal components of an operation and *steps* denote procedure-specific segments achieving discrete clinical objectives [24]. Preliminary workflow definitions were piloted on a multicentre dataset comprising laparoscopic and robotic CRS videos from five institutions to ensure broad procedural representation. Rigid sequencing of phases and steps was intentionally avoided to capture real-world variability in surgical practice.

The Delphi survey was implemented using JotForm (JotForm Inc., San Francisco, CA). Information on participants demographics, clinical experience, and country of practice were collected. Each proposed phase or step was rated on two 5-point Likert scales: (1) inclusion relevance, and (2) clarity and completeness of description. Participants could also provide free-text comments, suggest edits, or propose new items. Consensus thresholds were based on prior Delphi studies [25]: items with ≥70% agreement (Likert ≥4/5) were accepted, those with 60–70% agreement or more than two revision suggestions were modified and carried forward, and items with <60% agreement were excluded. Following each round, aggregated and de-identified feedback was reviewed by the steering committee to refine workflow descriptors. A summary of results was distributed to participants after each round.

All Round 2 participants were invited to the third round which consisted of an online consensus meeting. Unresolved items were reviewed with presentation of agreement data, comment counts, and item progress summaries. A steering committee moderator facilitated discussion until full consensus was achieved. Revised definitions were circulated to all participants for final review and approval.

2.3. Development and validation of ColoWorkflow

Workflow descriptors reaching final consensus were incorporated into the *ColoWorkflow* VBA tool. Start and end points for each phase and step were defined by the steering committee using stable, visually identifiable cues to ensure reproducible annotation.

To assess applicability, *ColoWorkflow* was used by one of the doctors in the steering committee to annotate a dataset of MIS CRS procedural videos from four European institutions and a public dataset [16]. To assess inter-rater reliability (IRR), four independent raters (a medical student, a physician, and two surgical residents) annotated a representative subset of 10 procedures, two per major procedure type. All raters had prior VBA experience and used the MOSaiC (IHU-Strasbourg, France), a web-based platform for collaborative medical video analysis [26].

2.4. Outcomes and statistical analysis

Participation metrics, including response and retention rates, were summarized descriptively. Mean Likert scores for each round were compared to assess convergence toward consensus. Inter-rater reliability was quantified using Cohen's κ and percentage agreement [27]. Cohen's κ represents the proportion of agreement beyond that expected by chance and is calculated as:

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

where $Pr(a)$ denotes observed agreement and $Pr(e)$ represents expected agreement by chance. Strength of agreement was interpreted as follows: almost perfect ($\kappa > 0.90$), strong (0.80–0.90), moderate (0.60–0.79), weak (0.40–0.59), minimal (0.21–0.39), or none (0–0.20)[27]. κ values were computed for all six rater pairs and averaged.

All analyses were performed in Python (version 3.12; Python Software Foundation, Wilmington, DE) using *pandas*, *numpy*, and *scikit-learn*. Visualizations were generated with *matplotlib* and *seaborn*.

2.5. Ethical statement

Ethical approval for the collection and analysis of de-identified surgical videos was obtained under the OPERATE protocol (ID 6456). All videos were pseudonymized prior to analysis, and no identifiable patient information was included.

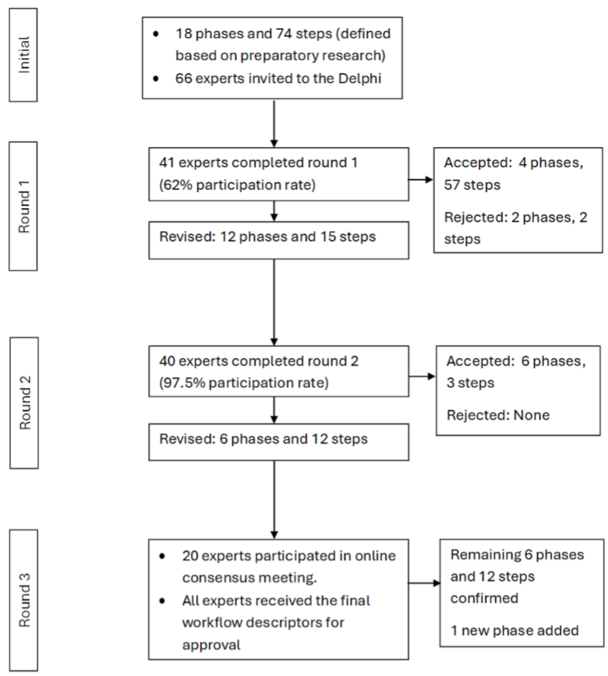


Fig. 1. Results of the modified Delphi process. The numbers for *phases* and *steps* represent both the inclusion of *phases/steps* and their corresponding descriptions as separate items for consensus.

3. Results

3.1. Modified Delphi study for expert consensus on colorectal workflow descriptors

The modified Delphi process was conducted over three rounds spanning five months (Figure 1). Panel selection and initial item generation were completed during the first two months. Round 1 (Dec 3, 2024) remained open for 14 days, followed by Round 2 (January 3, 2025) for 16 days, with one to two reminder emails per round. The final consensus meeting (Round 3) was held online on February 27th, 2025, followed by a 10-day feedback period for final approval.

Out of 66 invited experts from 13 countries, 41 (62%) from 11 countries completed Round 1. The panel was diverse in both geographic and procedural experience, with 40% having performed more than 1,000 colorectal surgeries (Supplementary Table 1). Forty participants (97.5%) continued to Round 2. The initial draft included 18 *phase* items and 74 *step* items describing minimally invasive (MIS) colorectal surgery (CRS) workflows. Items achieving ≥ 70% agreement (Likert score ≥ 4/5) with minimal qualitative feedback (≤ 2 revision suggestions) were accepted. Average Likert scores increased between rounds for *phase* items and remained stable for *step* items. Most *steps* (77%) items were accepted in Round 1, with 20% of the revised *steps* subsequently accepted in Round 2. Accepted *phases* items increased from 22% to 50% between Delphi Round 1 and 2, indicating progressive consensus among participants (Table 1). The distribution of qualitative feedback across *phases* and *steps* items is illustrated in

Table 1. Results of acceptance, rejection, and revision in Rounds 1 and 2 of the Delphi. Round 3 was a collegial discussion until unanimous agreement. The numbers represent both the inclusion of *phases/steps* and their corresponding descriptions.

Status	Round 1		Round 2	
	Phases (%)	Steps (%)	Phases (%)	Steps (%)
Accept	4 (22.22)	57 (77.02)	6 (50)	3 (20)
Revise	12 (66.67)	15 (20.27)	6 (50)	12 (80)
Reject	2 (11.1)	2 (2.70)	0	0

Supplementary Figure 1a–b, indicating that certain procedural domains required greater refinement.

Twenty experts attended the final consensus meeting. Remaining items were reviewed, discussed, and revised for clarity and precision. Key modifications included splitting the *phase* “Other Interventions” into 2 distinct *phases*: “Preplanned Additional Procedures” and “Unplanned Procedures”. The definition of “Leak Testing” was modified, and the description of “Mesorectal Dissection” was refined to specify circumferential dissection with or without vascular control. Following these revisions, expert consensus was achieved on 10 procedural *phases* and 34 workflow *steps* (Figure 2) and their description. Complete *phase* and *step* definitions are detailed in Supplementary Tables 3 and 4.

3.2. Development and validation of ColoWorkflow

The video-based assessment (VBA) tool, *ColoWorkflow*, was constructed using the consensus-derived workflow descriptors to ensure face and content validity according to Messick’s validity framework [28]. Each *phase* and *step* was operationally defined with clear visual start and end cues to facilitate reproducible annotation across videos.

ColoWorkflow was applied to 54 MIS CRS procedural videos encompassing: left hemicolectomies (10 cases), right hemicolectomies (12), rectal resections (10), sigmoid resections (12), and total proctocolectomies (10). Video durations ranged from 82 to 200 minutes with an average of 142.19 ± 59 minutes (see Supplementary Table 2). All *phase* and *step* labels except “Ileoanal pouch creation” were represented in at least one video, with no missing or redundant labels reported. On average, each video contained 7.52 ± 0.79 *phases* and 16.46 ± 2.41 *steps*. Mean *phase* duration was 18.91 ± 17.46 minutes, and mean *step* duration was 6.49 ± 7.01 minutes.

On a subset of 10 videos annotated by four independent raters, the mean Cohen’s κ was 0.71 for *phases* and 0.66 for *steps*, indicating moderate agreement (Table 2). Pairwise κ values stratified by raters’ experience are reported in Table 2. Labels with low κ scores or high variability (highlighted in Supplementary Figure 2a–b) were examined during structured debriefings. Qualitative review revealed that disagreements stemmed primarily from inherent challenges in interpreting surgical videos, such as ambiguous visual cues or overlapping tasks, rather than inconsistent annotation practice. Raters’ feedback and insights from the qualitative review were used to refine the *ColoWorkflow* VBA framework. Common pitfalls,

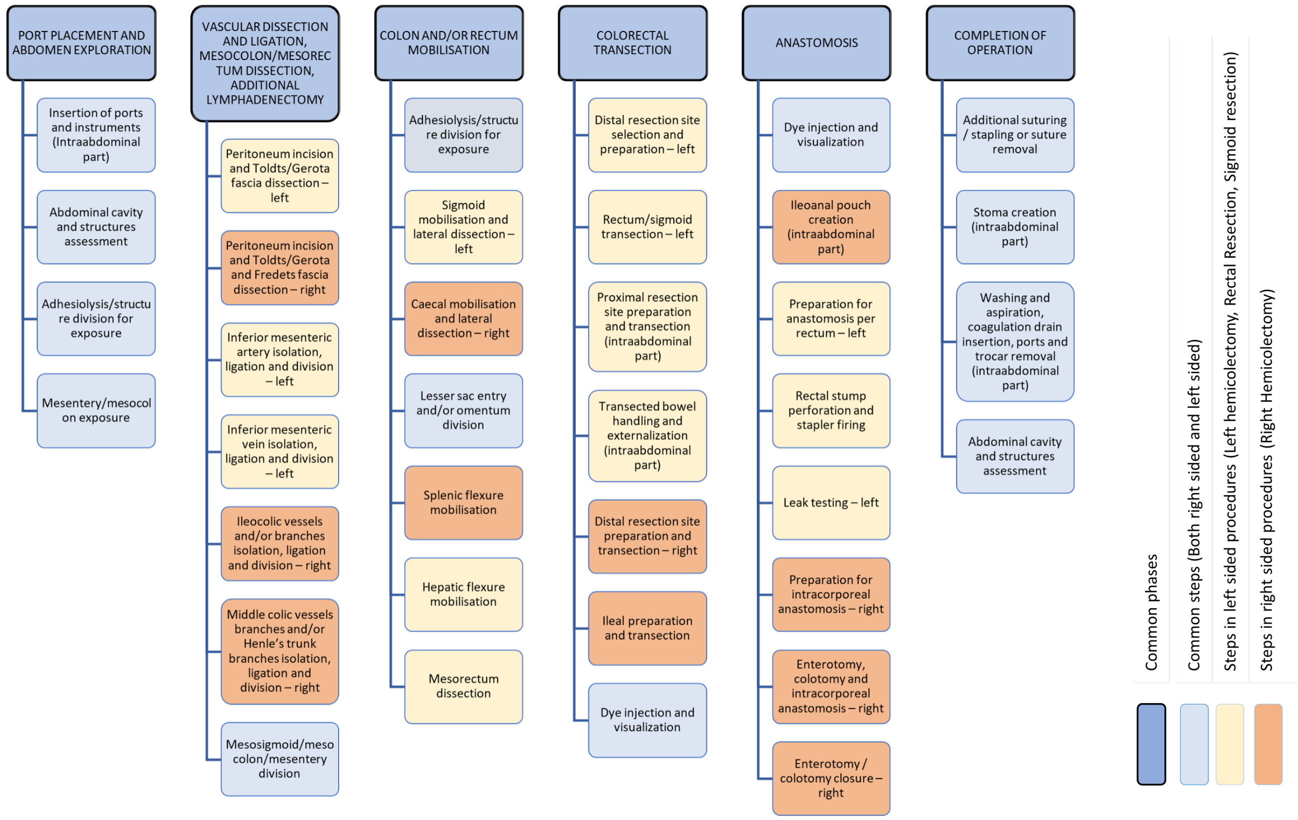


Fig. 2. Representation of *ColoWorkflow*. Common *phases* are shown in dark blue, steps shared across left- and right-sided procedures in light blue, steps specific to left-sided procedures (left hemicolectomy, rectal resection, sigmoid resection) in yellow, and steps specific to right-sided procedures (right hemicolectomy) in orange.

Table 2. Interrater reliability metrics, including Cohen's κ and agreement percentage.

Rater pairs	**Phases**		**Steps**	
	Cohen's κ	**Agreement (%)**	**Cohen's κ**	**Agreement (%)**
Resident 1–Resident 2	0.70	75.74	0.66	69.41
Resident 1–Med student	0.68	73.55	0.60	63.58
Resident 1–Doctor	0.74	78.56	0.73	75.28
Resident 2–Med student	0.69	73.96	0.64	66.73
Resident 2–Doctor	0.78	81.89	0.73	75.09
Med student–Doctor	0.72	76.48	0.66	68.47
Average	**0.72**	**76.70**	**0.67**	**69.76**

along with recommended strategies to improve annotations, are summarized in Table 3.

4. Discussion

This study developed and validated *ColoWorkflow*, a video-based assessment (VBA) tool designed to characterize procedural workflows across minimally invasive (MIS) colorectal surgeries (CRS). Consensus on 10 procedural phases and 34 workflow steps was established through a modified Delphi process involving an international panel of experts, ensuring *ColoWorkflow*'s face and content validity. Application of the tool to 54 surgical videos demonstrated its feasibility, consistency, and interrater reliability, providing preliminary evidence of response process and internal structure validity according to Messick's framework. Collectively, these findings support *ColoWorkflow* as a robust foundation for standardized workflow analysis in MIS CRS.

The present findings complement and extend prior work on VBA of CRS workflow. Previous Delphi-based frameworks have focused on individual procedures, such as laparoscopic right colectomy or sigmoid [20, 21, 29, 30], or on discrete segments like splenic flexure mobilization [31]. More recently, the Heidelberg Colorectal (HeiCo) dataset introduced a phase-based framework generalizable across left-sided CRS procedures [16]. *ColoWorkflow* builds on top of previous frameworks by integrating procedure-agnostic phases with procedure-specific steps within a unified hierarchical structure. It is flexible regarding the order of phases or steps to work across specific surgical approaches. This design promotes granularity where needed while enabling cross-procedure comparison, and captures observable surgical actions that are commonly encountered, regardless of institutional preferences or surgeon-specific styles.

The rigorous consensus methodology and diverse expert representation underpin the validity and generalizability of this work. The modified Delphi approach proved effective for harmonizing complex surgical descriptions. High expert engagement led to the direct acceptance of ~70% of step items in Round 1. Items with disagreement largely remained contested in Round 2 (80%), revealing that some steps are universally agreed upon, while others reflect inherent variability in sur-

Table 3. Identified challenges contributing to interrater variability and mitigation strategies to overcome them.

Problem	Mitigation strategy
Ambiguous start and end of phases and steps	Base transitions on clear visual cues and provide illustrative examples.
Hard to recognize anatomical landmarks	Apply the VBA tool by surgeons or trained annotators supported with image primers, expert video reviews, and videos of sufficient quality.
Flickering phases/steps	Define *a priori* whether to optimize for workflow stability (goal-oriented annotations, high threshold for phase/step change) or for granularity (action-oriented annotations, low threshold for phase/step change).
Learning and fatigue effects	Schedule interim calibration meetings and distribute analysis sessions over time.

gical technique. Disagreement often centered on terminology rather than content, emphasizing the importance of linguistic precision in workflow modeling [32]. The present findings also highlight the challenge of balancing annotation granularity for clinical interpretation with the coarser segmentation needed for efficient AI model training [33, 34].

The moderate ($0.6 \leq \kappa \leq 0.8$) interrater reliability suggests the reliability of *ColoWorkflow*, providing evidence of its internal structure validity. An analysis of disagreements informed the refinement of *ColoWorkflow*, integrating expert-based consensus recommendations with practical annotation insights. However, some disagreements seemed to derive from variability in mental models, VBA tool familiarity, and labeling strategies, reflecting inherent and often unavoidable differences in conceptualizing workflow definitions [35]. A structured orientation phase before any large-scale video-based analysis—including a co-review of sample videos and clearly defined usage guidelines for the VBA tool—could help overcome these challenges.

Altogether, several aspects distinguish *ColoWorkflow* from other frameworks for workflow analysis. Its thorough development through a highly participated modified Delphi exercise and its application to a multicenter dataset spanning CRS procedures provide early but convincing evidence of the tool's validity. *ColoWorkflow*'s hierarchical and flexible structure allows analysis across procedure types, surgical platforms (laparoscopic and robotic), and institutional protocols; such generalizability should favor applicability, cross-center benchmarking, and continuous professional development. Focusing on observable actions rather than prescriptive sequencing, it reflects real-world diversity while maintaining analytic rigor, facilitating VBA research. Importantly, the tool's explicit visual definitions and modular structure facilitate the development of AI for CRS workflow analysis, providing a bridge between expert consensus and machine learning scalability.

This work has limitations. First, the collegial discussion during the final Delphi round may have reduced independent expression due to group dynamics. Nonetheless, such discussion followed two independent consensus rounds and was of critical importance to guarantee convergence on workflow descriptors, probably contributing to *ColoWorkflow*'s face and content validity. Despite broad international participation, the expert panel may not have fully captured regional or technique-specific variations in surgical practice. Additionally, while the multicenter dataset included multiple CRS procedures, some were not represented (e.g., pouch creation). Future studies will easily integrate more steps or procedural variations into *ColoWorkflow* thanks to its hierarchical design and flexibility regarding phase and step sequencing. Finally, raters did not include expert CRS surgeons. While including more experienced raters might have improved interrater reliability, it was preferred to enroll medical students and residents trained on VBA, as workflow analysis does not entail assessments of high surgical semantics (e.g., analyzing anastomosis quality). Considering how overburdened expert surgeons are, a tool trainees can use should facilitate implementation.

Despite these limitations, this work provides a critical foundation for future research and clinical translation. Establishing a validated, generalizable colorectal workflow schema enables more consistent measurement of intraoperative performance and facilitates cross-institutional comparisons. When combined with advances in computer vision and AI, *ColoWorkflow* can accelerate the development of automated systems for workflow recognition, skill assessment, and intraoperative guidance. Moreover, standardized workflow mapping can enhance training curricula and quality improvement initiatives by linking intraoperative processes to patient outcomes.

In conclusion, *ColoWorkflow* represents the first expert-validated, procedure-generalizable framework for analyzing minimally invasive colorectal surgical workflows. Through a rigorous consensus and validation process, it establishes a shared language for procedural analysis that is both granular and adaptable. Future work will focus on expanding validity evidence, integrating additional procedure types, and leveraging *ColoWorkflow* as a foundation for AI-based workflow analysis and surgical education.

5. Disclosures

Pietro Mascagni and Nicolas Padoy are co-founders and shareholders of Scialytics. Jim Khan and Eloy Espin-Basany serve as proctors for Intuitive Surgical. Isacco Montroni has received compensation for participation in courses organized by Olympus and Fresenius Kabi. Monica Ortenzi is a paid collaborator with Theator. Willem A. Bemelman is a shareholder in Semiflex. All other authors declare no conflicts of interest.

6. Acknowledgement

This study was not preregistered in an independent institutional registry. De-identified consensus data and analytical methods data may be made available from the corresponding author on reasonable request.

6.1. Data Availability Statement

De-identified video data may be made available from the corresponding author on reasonable request, pending institutional approvals and data-use agreements.

6.2. Funding Statement

This work has received funding from the European Union (ERC, CompSURG, 101088553). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. This work was also developed within the Interdisciplinary Thematic Institute HealthTech (ITI 2021-2028 program of the University of Strasbourg, CNRS and Inserm), supported by IdEx Unistra (ANR-10-IDEX-0002) and SFRI (STRATUS project, ANR-20-SFRI-0012) under the framework of the French Investments for the Future Program. This work was partially supported by French state funds managed by the ANR under Grant ANR-10-IAHU-02 (IHU Strasbourg).

CompSurg Colorectal Experts Group Collaborators

Alberto Arezzo[a], Alfonso Lapergola[b], Amjad Parvaiz[c], Deena Harji[d], Domenico D'Ugo[e], Elisa Cassinotti[f], Eloy Espin-Basany[g], Emanuele Rausa[h], Francesco Brucchi[i], Gabriele Anania[j], Gaya Spolverato[k], Gianluca Rizzo[l], Giovanni Guglielmo Laracca[m], Giuseppe Quero[n], Graziano Pernazza[o], Isacco Montroni[p], Jared Torkington[q], Jim Khan[r], Manish Chand[s], Maria Vannucci[t], Michel Adamina[u], Monica Ortenzi[v], Nariaki Okamoto[w], Paola De Nardi[x], Paolo Pietro Bianchi[y], Patricia Sylla[z], Ronan Cahill[aa], Stefano Rausei[ab], Umberto Bracale[ac], Vincenzo Tondolo[ad], Willem A. Bemelman[ae]

a *Department of Surgical Sciences, University of Turin, Turin, Italy*

b *Visceral and Digestive Surgery Department, Colorectal and Robotic Unit, Nouvel Hôpital Civil, University Hospitals of Strasbourg, Strasbourg, France; IRCAD (Research Institute against Digestive Cancer), Strasbourg, France; IHU (Institut Hospitalo-Universitaire), Strasbourg, France*

c *Champalimaud Foundation, Lisbon, Portugal*

d *Department of Colorectal Surgery, Manchester University NHS Foundation Trust, Manchester, UK; Clinical Trials Research Unit, University of Leeds, Leeds, UK*

e *Fondazione Policlinico Universitario Gemelli – IRCCS, Rome, Italy*

f *Department of General and Minimally Invasive Surgery, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy*

g *General Surgery Department, Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Spain*

h *Colorectal Surgery Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy*

i *University of Milan, Milan, Italy*

j *Department of Medical Sciences, University of Ferrara, Ferrara, Italy*

k *General Surgery 3, Azienda Ospedale Università di Padova; Department of Surgical Oncological and Gastrointestinal Sciences, University of Padova, Padova, Italy*

l *UOC Chirurgia Digestiva e del Colon-Retto, Ospedale Isola Tiberina Gemelli Isola, Rome, Italy*

m *AOU Sant'Andrea, Surgery Department, Rome, Italy*

n *Fondazione Policlinico Universitario Agostino Gemelli IRCCS di Roma; Università Cattolica del Sacro Cuore di Roma, Italy*

o *San Giovanni Addolorata Hospital, Department of Surgery, Rome, Italy*

p *IRCCS Fondazione Istituto Nazionale dei Tumori, Division of Colon and Rectal Cancer Surgery, Milan, Italy*

q *University Hospital of Wales, Cardiff, UK*

r *Portsmouth Hospitals University NHS Trust, Portsmouth, UK; University of Portsmouth, UK*

s *University College London, London, UK*

t *University of Torino, Department of General Surgery, Torino, Italy; Institute of Image-Guided Surgery, IHU-Strasbourg, Strasbourg, France*

u *Department of Visceral Surgery and Medicine, Inselspital, Bern University Hospital, Bern, Switzerland*

v *Università Politecnica delle Marche, Ancona, Italy*

w *IRCAD (Research Institute against Digestive Cancer), Strasbourg, France; Department of Colorectal Surgery, National Cancer Center Hospital East, Kashiwa, Japan*

x *Colorectal Surgery, IRCCS San Raffaele Scientific Institute, Milan, Italy*

y *ASST Santi Paolo e Carlo, Università di Milano, Milan, Italy*

z *Mount Sinai Hospital, New York, NY, USA*

aa *UCD Centre for Precision Surgery, University College Dublin, Dublin, Ireland; Department of Surgery, Mater Misericordiae University Hospital, Dublin, Ireland*

ab *Department of Surgery, ASST Sette Laghi, Cittiglio–Angera (Varese), Italy*

ac *Department of Medicine, Surgery and Odontoiatry, University of Salerno, Salerno, Italy*

ad *UOC Chirurgia Digestiva e Colon-Retto, Ospedale Isola Tiberina–Gemelli Isola, Rome, Italy; Università Cattolica del Sacro Cuore, Rome, Italy*

ae *Department of Surgery, Amsterdam University Medical Centers, Amsterdam, The Netherlands*

References

[1] J. C. Carmichael, D. S. Keller, G. Baldini, L. Bordeianou, E. Weiss, L. Lee, M. Boutros, J. McClane, L. S. Feldman, S. R. Steele, Clinical Practice Guidelines for Enhanced Recovery After Colon and Rectal Surgery From the American Society of Colon and Rectal Surgeons and Society of American Gastrointestinal and Endoscopic Surgeons, Diseases of the Colon & Rectum 60 (8) (2017) 761. doi:10.1097/DCR.0000000000000883.
URL https://journals.lww.com/dcrjournal/fulltext/2017/08000/clinical_practice_guidelines_for_enhanced_recovery.3.aspx

[2] M. Greco, G. Capretti, L. Beretta, M. Gemma, N. Pecorelli, M. Braga, Enhanced Recovery Program in Colorectal Surgery: A Meta-analysis of Randomized Controlled Trials, World Journal of Surgery 38 (6) (2014) 1, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1007/s00268-013-2416-8. doi:10.1007/s00268-013-2416-8.
URL https://onlinelibrary.wiley.com/doi/abs/10.1007/s00268-013-2416-8

[3] G. B. Hanna, H. Mackenzie, D. Miskovic, M. Ni, S. Wyles, P. Aylin, A. Parvaiz, T. Cecil, A. Gudgeon, J. Griffith, J. M. Robinson, C. Selvasekar, T. Rockall, A. Acheson, C. Maxwell-Armstrong, J. T. Jenkins, A. Horgan, C. Cunningham, I. Lindsey, T. Arulampalam, R. W. Motson, N. K. Francis, R. H. Kennedy, M. G. Coleman, Laparoscopic Colorectal Surgery Outcomes Improved After National Training Program (LAPCO) for Specialists in England, Annals of Surgery 275 (6) (2022) 1149. doi:10.1097/SLA.0000000000004584.
URL https://journals.lww.com/annalsofsurgery/abstract/2022/06000/laparoscopic_colorectal_surgery_outcomes_improved.19.aspx

[4] C. S. McArdle, D. Hole, Impact of variability among surgeons on postoperative morbidity and mortality and ultimate survival., British Medical Journal 302 (6791) (1991) 1501–1505, publisher: British Medical Journal Publishing Group Section: Research Article. doi:10.1136/bmj.302.6791.1501.
URL https://www.bmj.com/content/302/6791/1501

[5] C. L. Bennett, S. J. Stryker, M. R. Ferreira, J. Adams, R. W. Beart, Jr, The Learning Curve for Laparoscopic Colorectal Surgery: Preliminary Results From a Prospective Analysis of 1194 Laparoscopic-Assisted Colectomies, Archives of Surgery 132 (1) (1997) 41–44. doi:10.1001/archsurg.1997.01430250043009.
URL https://doi.org/10.1001/archsurg.1997.01430250043009

[6] D. S. Keller, J. J. Stulberg, J. K. Lawrence, H. Samia, C. P. Delaney, Initiating statistical process control to improve quality outcomes in colorectal surgery, Surgical Endoscopy 29 (12) (2015) 3559–3564. doi:10.1007/s00464-015-4108-y.
URL https://doi.org/10.1007/s00464-015-4108-y

[7] N. J. Curtis, J. D. Foster, D. Miskovic, C. S. B. Brown, P. J. Hewett, S. Abbott, G. B. Hanna, A. R. L. Stevenson, N. K. Francis, Association of Surgical Skill Assessment With Clinical Outcomes in Cancer Surgery, JAMA surgery 155 (7) (2020) 590–598. doi:10.1001/jamasurg.2020.1004.

[8] A. C. W. Reitz, N. N. Massarweh, Video-Based Assessment of Surgical Quality—Will Video Kill the Radio Star?, JAMA Network Open 7 (4) (2024) e246477. doi:10.1001/jamanetworkopen.2024.6477.
URL https://doi.org/10.1001/jamanetworkopen.2024.6477

[9] C. M. Pugh, D. A. Hashimoto, J. R. Korndorffer, The what? How? And Who? Of video based assessment, The American Journal of Surgery 221 (1) (2021) 13–18. doi:10.1016/j.amjsurg.2020.06.027.
URL https://www.sciencedirect.com/science/article/pii/S0002961020303809

[10] S. Balvardi, A. Kammili, M. Hanson, C. Mueller, M. Vassiliou, L. Lee, K. Schwartzman, J. F. Fiore, L. S. Feldman, The association between video-based assessment of intraoperative technical performance and patient outcomes: a systematic review, Surgical Endoscopy 36 (11) (2022) 7938–7948. doi:10.1007/s00464-022-09296-6.
URL https://doi.org/10.1007/s00464-022-09296-6

[11] B. C. Brajcich, J. J. Stulberg, B. E. Palis, J. W. Chung, R. Huang, H. Nelson, K. Y. Bilimoria, Association Between Surgical Technical Skill and Long-term Survival for Colon Cancer, JAMA Oncology 7 (1) (2021) 127–129. doi:10.1001/jamaoncol.2020.5462.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7600051/

[12] J. J. Stulberg, R. Huang, L. Kreutzer, K. Ban, B. J. Champagne, S. R. Steele, J. K. Johnson, J. L. Holl, C. C. Greenberg, K. Y. Bilimoria, Association Between Surgeon Technical Skills and Patient Outcomes, JAMA Surgery 155 (10) (2020) 960–968. doi:10.1001/jamasurg.2020.3007.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7439214/

[13] O. A. Varban, J. R. Thumma, J. F. Finks, A. M. Carlin, A. A. Ghaferi, J. B. Dimick, Evaluating the Effect of Surgical Skill on Outcomes for Laparoscopic Sleeve Gastrectomy: A Video-based Study, Annals of Surgery 273 (4) (2021) 766. doi:10.1097/SLA.0000000000003385.
URL https://journals.lww.com/annalsofsurgery/abstract/2021/04000/evaluating_the_effect_of_surgical_skill_on.21.aspx

[14] P. Mascagni, D. Alapatt, L. Sestini, M. S. Altieri, A. Madani, Y. Watanabe, A. Alseidi, J. A. Redan, S. Alfieri, G. Costamagna, I. Boškoski, N. Padoy, D. A. Hashimoto, Computer vision in surgery: from potential to clinical value, npj Digital Medicine 5 (1) (2022) 163, publisher: Nature Publishing Group. doi:10.1038/s41746-022-00707-5.
URL https://www.nature.com/articles/s41746-022-00707-5

[15] W. T. Gattrell, P. Logullo, E. J. v. Zuuren, A. Price, E. L. Hughes, P. Blazey, C. C. Winchester, D. Tovey, K. Goldman, A. P. Hungin, N. Harrison, ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi, PLOS Medicine 21 (1) (2024) e1004326, publisher: Public Library of Science. doi:10.1371/journal.pmed.1004326.
URL https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1004326

[16] L. Maier-Hein, M. Wagner, T. Ross, A. Reinke, S. Bodenstedt, P. M. Full, H. Hempe, D. Mindroc-Filimon, P. Scholz, T. N. Tran, P. Bruno, A. Kisilenko, B. Müller, T. Davitashvili, M. Capek, M. D. Tizabi, M. Eisenmann, T. J. Adler, J. Gröhl, M. Schellenberg, S. Seidlitz, T. Y. E. Lai, B. Pekdemir, V. Roethlingshoefer, F. Both, S. Bittel, M. Mengler, L. Mündermann, M. Apitz, A. Kopp-Schneider, S. Speidel, F. Nickel, P. Probst, H. G. Kenngott, B. P. Müller-Stich, Heidelberg colorectal data set for surgical data science in the sensor operating room, Scientific Data 8 (1) (2021) 101, publisher: Nature Publishing Group. doi:10.1038/s41597-021-00882-2.
URL https://www.nature.com/articles/s41597-021-00882-2

[17] M. Farquharson, J. Hollingshead, B. Moran, Farquharson's Textbook of Operative General Surgery, CRC Press, 2014, google-Books-ID: JtEgCAAAQBAJ.

[18] S. D. Wexner, J. W. Fleshman, Colon and Rectal Surgery: Anorectal Operations, Lippincott Williams & Wilkins, 2012.

[19] M. L. Corman, Colon and Rectal Surgery, Lippincott Williams & Wilkins, 2005, google-Books-ID: imEv3PqkJSEC.

[20] F. A. Dijkstra, R. J. I. Bosker, N. J. G. M. Veeger, M. J. van Det, J. P. E. N. Pierie, Procedural key steps in laparoscopic colorectal surgery, consensus through Delphi methodology, Surgical Endoscopy 29 (9) (2015) 2620–2627. doi:10.1007/s00464-014-3979-7.
URL https://doi.org/10.1007/s00464-014-3979-7

[21] V. N. Palter, H. M. MacRae, T. P. Grantcharov, Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology, The American Journal of Surgery 201 (2) (2011) 251–259. doi:10.1016/j.amjsurg.2010.01.031.
URL https://www.sciencedirect.com/science/article/pii/S0002961010002606

[22] D. Kitaguchi, N. Takeshita, H. Matsuzaki, H. Takano, Y. Owada, T. Enomoto, T. Oda, H. Miura, T. Yamanashi, M. Watanabe, D. Sato, Y. Sugomori, S. Hara, M. Ito, Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach, Surgical Endoscopy 34 (11) (2020) 4924–4931. doi:10.1007/s00464-019-07281-0.
URL https://doi.org/10.1007/s00464-019-07281-0

[23] C. W. Strey, C. Wullstein, M. Adamina, A. Agha, H. Aselmann, T. Becker, R. Grützmann, W. Kneist, M. Maak, B. Mann, K. T. Moesta, N. Runkel, C. Schafmayer, A. Türler, T. Wedel, S. Benz, Laparoscopic right hemicolectomy with CME: standardization using the "crit-

ical view" concept, Surgical Endoscopy 32 (12) (2018) 5021–5030. doi:10.1007/s00464-018-6267-0.
URL https://doi.org/10.1007/s00464-018-6267-0

[24] O. R. Meireles, G. Rosman, M. S. Altieri, L. Carin, G. Hager, A. Madani, N. Padoy, C. M. Pugh, P. Sylla, T. M. Ward, D. A. Hashimoto, the SAGES Video Annotation for AI Working Groups, SAGES consensus recommendations on an annotation framework for surgical video, Surgical Endoscopy 35 (9) (2021) 4918–4929. doi:10.1007/s00464-021-08578-9.
URL https://doi.org/10.1007/s00464-021-08578-9

[25] M. Kommer, C. Hawthorne, L. Moss, I. Piper, R. O'Kane, M. Czosnyka, P. Enblad, J. C. Hemphill, A. Spiegelberg, J. S. Riddell, M. Shaw, International e-Delphi survey to define best practice in the reporting of intracranial pressure monitoring recording data, Brain and Spine 4 (2024) 102860. doi:10.1016/j.bas.2024.102860.
URL https://www.sciencedirect.com/science/article/pii/S2772529424001164

[26] J.-P. Mazellier, A. Boujon, M. Bour-Lang, M. Erharhd, J. Waechter, E. Wernert, P. Mascagni, N. Padoy, MOSaiC: a Web-based Platform for Collaborative Medical Video Assessment and Annotation, arXiv:2312.08593 [cs]. doi:10.48550/arXiv.2312.08593.

[27] M. L. McHugh, Interrater reliability: the kappa statistic, Biochemia Medica 22 (3) (2012) 276–282, publisher: Hrvatsko društvo za medicinsku biokemiju i laboratorijsku medicinu.
URL https://hrcak.srce.hr/89395

[28] S. Messick, Standards of Validity and the Validity of Standards in Performance Asessment, Educational Measurement: Issues and Practice 14 (4) (1995) 5–8, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3992.1995.tb00881.x. doi:10.1111/j.1745-3992.1995.tb00881.x.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.1995.tb00881.x

[29] G. S. Sica, G. Anania, C. Fiorani, L. Siragusa, D. Vinci, M. Caricato, P. Delrio, A. Agrusa, G. Baldazzi, R. Reddavid, G. Pellino, RRoC-STAR Collaborative Group, Standardization of the surgical technique and reporting for radical right colectomy with central vascular ligation and complete mesocolic excision (RRoC-STAR): Delphi consensus, BJS Open 9 (3) (2025) zraf066. doi:10.1093/bjsopen/zraf066.
URL https://doi.org/10.1093/bjsopen/zraf066

[30] L. Lorenzon, A. Biondi, T. Carus, A. Dziki, E. Espin, N. Figueiredo, M. G. Ruiz, T. Mersich, I. Montroni, P. J. Tanis, S. R. Benz, P. P. Bianchi, M. Biebl, I. Broeders, R. De Luca, P. Delrio, M. D'Hondt, A. Fürst, J. Grosek, J. F. Guimaraes Videira, F. Herbst, D. Jayne, G. Lázár, D. Miskovic, A. Muratore, O. Helmer Sjo, T. Scheinin, A. Tomazic, A. Türler, C. Van De Velde, S. D. Wexner, C. Wullstein, W. Zegarski, D. D'Ugo, Achieving high quality standards in laparoscopic colon resection for cancer: A Delphi consensus-based position paper, European Journal of Surgical Oncology 44 (4) (2018) 469–483. doi:10.1016/j.ejso.2018.01.091.
URL https://linkinghub.elsevier.com/retrieve/pii/S0748798318301288

[31] Y. S. Lee, Three Surgical Approaches of Laparoscopic Splenic Flexure Mobilization, The Journal of Minimally Invasive Surgery 22 (2) (2019) 85–86. doi:10.7602/jmis.2019.22.2.85.
URL http://www.e-jmis.org/journal/view.html?doi=10.7602/jmis.2019.22.2.85

[32] Z. Liu, K. Chen, S. Wang, Y. Xiao, G. Zhang, Deep learning in surgical process modeling: A systematic review of workflow recognition, Journal of Biomedical Informatics 162 (2025) 104779. doi:10.1016/j.jbi.2025.104779.
URL https://www.sciencedirect.com/science/article/pii/S1532046425000085

[33] N. Padoy, M. Horn, H. Feussner, B. marie odile, N. Navab, Recovery of Surgical Workflow: a Model-based Approach, in: 21st International Congress and Exhibition on Computer Assisted Radiology and Surgery, 2007.

[34] D. A. Hashimoto, G. Rosman, E. R. Witkowski, C. Stafford, A. J. Navarette-Welton, D. W. Rattner, K. D. Lillemoe, D. L. Rus, O. R. Meireles, Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy, Annals of Surgery 270 (3) (2019) 414. doi:10.1097/SLA.0000000000003460.
URL https://journals.lww.com/annalsofsurgery/abstract/2019/09000/computer_vision_analysis_of_intraoperative_video_.3.aspx

[35] T. M. Ward, F. , Danyal M., B. , Yutong, R. , Guy, M. , Ozanan R., , D. A. Hashimoto, Challenges in surgical video annotation, Computer Assisted Surgery 26 (1) (2021) 58–68, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/24699322.2021.1937320. doi:10.1080/24699322.2021.1937320.
URL https://doi.org/10.1080/24699322.2021.1937320

Supplementary Materials

eTable 1. Demographics of participants of the Delphi study.

Variable	N (%)
Age (years)	
26–30	4 (9.8)
31–35	2 (4.9)
36–40	9 (22.0)
41–45	5 (12.2)
46–50	6 (14.6)
51–55	4 (9.8)
>55	11 (26.8)
Highest professional qualification	
Head of Department	10 (24.4)
Professor	16 (39.0)
Consultant	9 (22.0)
Fellow	2 (4.9)
Resident year 4	2 (4.9)
Resident year 5	2 (4.9)
Total number of colorectal surgeries performed	
1–200	8 (19.5)
201–400	9 (22.0)
401–600	2 (4.9)
601–800	4 (9.8)
801–1000	2 (4.9)
>1000	16 (39.0)
Number of colorectal surgeries performed per year	
1–50	13 (31.7)
51–100	12 (29.3)
101–200	3 (7.3)
201–300	12 (29.3)
>300	1 (2.4)

eTable 2. Description of the dataset used for validation of *ColoWorkflow*.

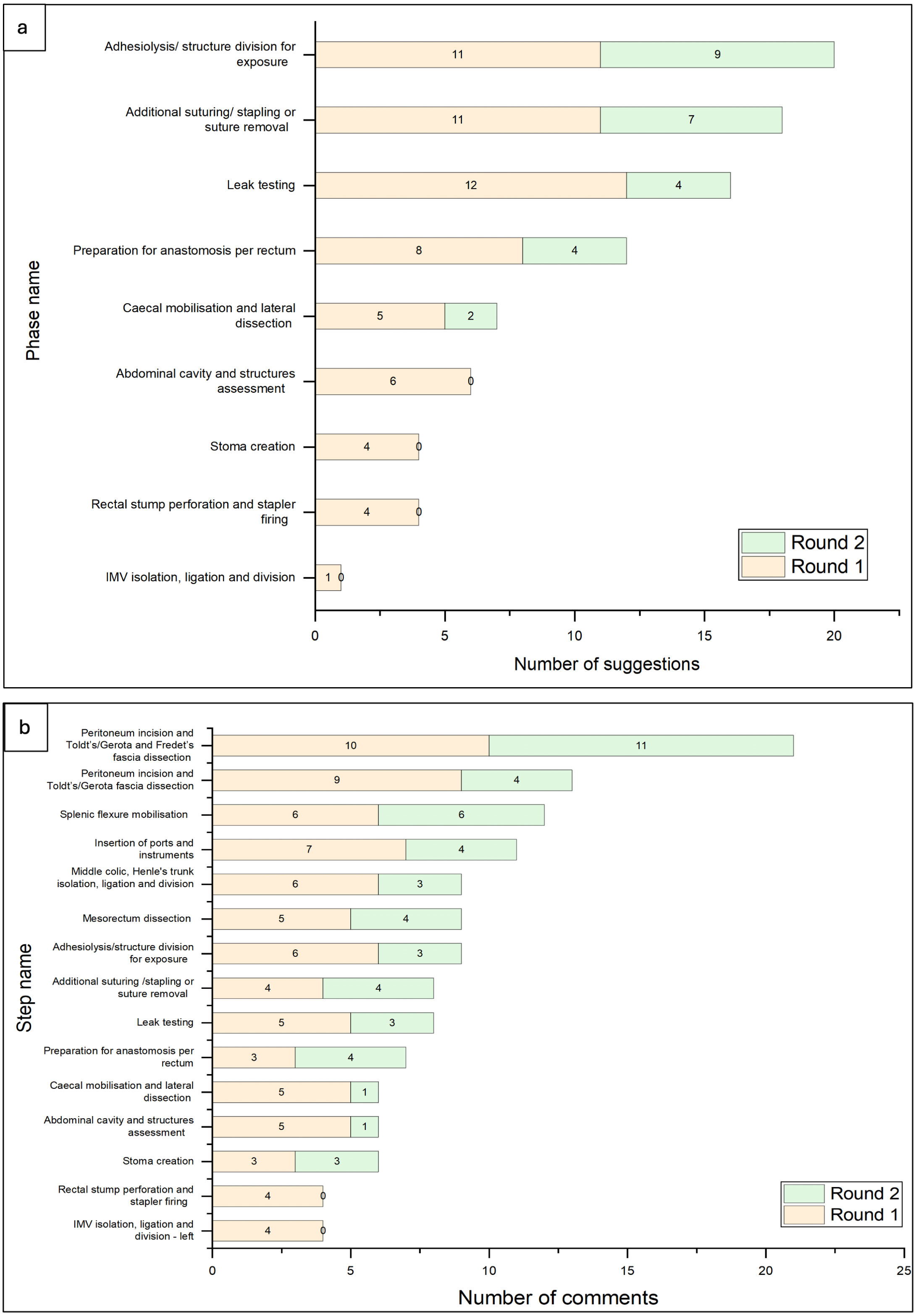
Procedure type	Source of data	Number of videos	Average duration (min)
Left Hemicolectomy	Center 1, Center 2, Center 3	10	100.82
Right Hemicolectomy	Center 1, Center 2, Center 3	12	93.00
Rectal Resection	Center 2, Center 3, HeiCo dataset	10	184.85
Sigmoidectomy	Center 2, Center 3, Center 4, HeiCo dataset	12	134.31
Total Proctocolectomy	HeiCo dataset	10	209.32

eTable 3. List of phases and their descriptions reaching consensus after all rounds of Delphi. Numbering is for reference only and does not necessarily reflect the required order of phases and steps.

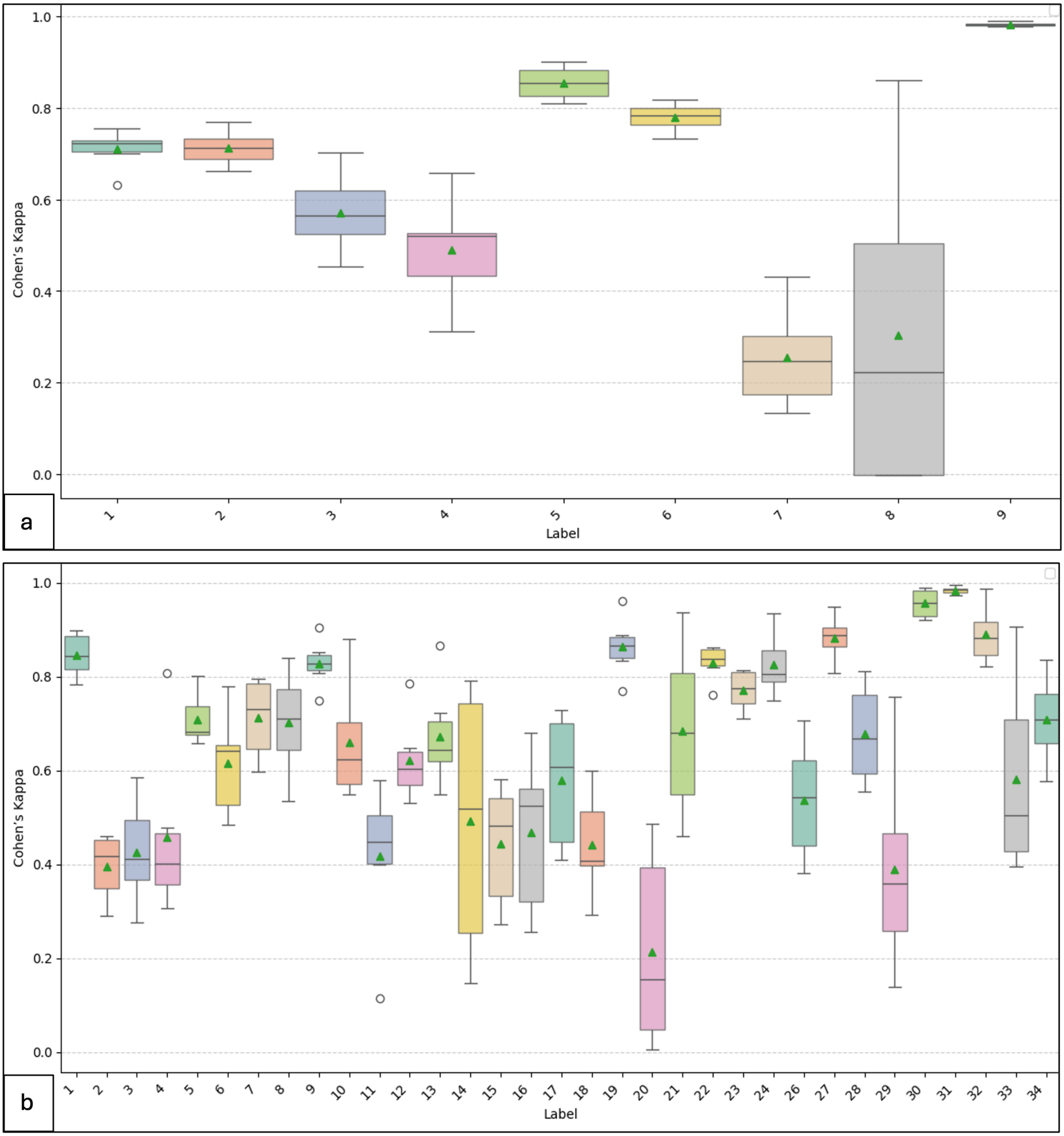
No.	Phase name	Description
1	Port placement and abdomen exploration	Laparoscopic camera insertion, additional port placement, cavity exploration, adhesiolysis, and manipulation of tissues/organs to expose the surgical field before starting executive phases.
2	Vascular dissection and ligation, mesocolon/mesorectum dissection, additional lymphadenectomy	Identification, dissection, and ligation of vascular pedicles to achieve procedure-specific lymphadenectomy, and dissection of the mesentery (mesocolon or mesorectum).
3	Colon and/or rectum mobilisation	Dissection of lateral attachments of the colon (ascending, descending, sigmoid), with/without continued splenic or hepatic flexure mobilisation.
4	Colorectal transection	Dissection and transection of proximal and distal bowel margins, with or without perfusion assessment.
5	Anastomosis	Reestablishment of bowel continuity.
6	Completion of operation	Final inspection of the abdominal cavity (hemostasis, viability, orientation, foreign body check), with/without mesenteric defect closure, drain placement, stoma creation (if planned), specimen retrieval, and port removal.
7	Preplanned additional procedures	Interventions not routinely part of colorectal surgery but preplanned (e.g., concurrent organ work, additional resections).
8	Unplanned procedures	Intraoperative deviations from planned workflow due to complications or judgment-based decisions.
9	Extracorporeal procedures	Dye visualization of the extracted colon for vascular assessment prior to proximal resection; may include ileoanal pouch construction.

eTable 4. Video-based assessment tool: list of steps and their descriptions reaching consensus after all rounds of Delphi. Numbering is for reference only and does not indicate sequential order.

No.	Step name	Description
1	Insertion of ports and instruments (intraabdominal part)	Trocar entry into the abdominal cavity and eventual insertion of instruments.
2	Abdominal cavity and structures assessment	Visual/instrumental assessment of bowel, liver, peritoneum, and other structures.
3	Adhesiolysis/structure division for exposure	Dividing adhesions and/or falciform ligament to delineate anatomy and optimize exposure of the target area.
4	Mesentery/mesocolon exposure	Lifting and repositioning omentum, small bowel, or colon to expose the mesentery/mesocolon root.
5	Peritoneum incision and Toldt's/Gerota fascia dissection – left	Incision of the peritoneum with subsequent dissection to isolate key vessels and develop the retroperitoneal plane between the Toldt and Gerota fasciae.



eFigure 1. Phases and steps in the video-based assessment (VBA) tool ranked by number of expert suggestions given for each phase/step naming and description. All phases shown in (a) (above); Top 15 steps shown in (b) (below). Comments mostly revolved around alternate approaches, inclusion of anatomical landmarks, and clarity in wording.



eFigure 2. Inter-annotator differences between the annotators and distribution across the labels. The plots show the Cohen's κ metric and its variability for each label, indicating that some labels have high agreement with low variation (e.g., step 1), while others may have low agreement and high variability (e.g., step 20). Phases shown in (a) (above); Steps in (b) (below). Phase and step names correspond to the numbering given in eTables 3 and 4.