BJS Open, 2026, zrag038
https://doi.org/10.1093/bjsopen/zrag038
Original Article

Video-based assessment tool for workflow analysis in minimally invasive colorectal surgery: expert consensus-based development and multicentre validation of ColoWorkflow

Pooja P. Jain[1,2,]†, Pietro Mascagni[1,3,4,*]†, Giuseppe Massimiani[1,3], Nabani Banik[2], Marta Goglia[5], Lorenzo Arboit[2], Britty Baby[2], Andrea Balla[6], Ludovica Baldari[7], Gianfranco Silecchia[5], Claudio Fiorillo[3]; The CompSurg Colorectal Experts Group, Sergio Alfieri[3,4], Salvador Morales-Conde[6], Deborah S. Keller[8,9], Luigi Boni[7] and Nicolas Padoy[1,2]

[1]Institute of Image-Guided Surgery, IHU Strasbourg, Strasbourg, France
[2]University of Strasbourg, CNRS, INSERM, ICube, UMR7357, Strasbourg, France
[3]Bioimage Analysis Center, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy
[4]General Surgery, Università Cattolica del Sacro Cuore, Rome, Italy
[5]Department of Medical–Surgical Sciences and Translational Medicine, Faculty of Medicine and Psychology, Sapienza University of Rome, Rome, Italy
[6]University Hospital Virgen Macarena, University of Sevilla, Seville, Spain
[7]Department of General and Minimally Invasive Surgery, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
[8]School of Biological and Health Systems Engineering, Arizona State University, Tempe, Arizona, USA
[9]Mayo Clinic, Phoenix, Arizona, USA

*Correspondence to: Pietro Mascagni, General Surgery Residency, Università Cattolica del Sacro Cuore, Largo Francesco Vito 1, Roma RM 00168, Italy (e-mails: pietro.mascagni01@icatt.it, pietro.mascagni@ihu-strasbourg.eu, pietro.mascagni@guest.policlinicogemelli.it)

Members of The CompSurg Colorectal Experts Group collaborators are co-authors of this study and are listed under the heading Collaborators.

†Joint first authors

Abstract

Background: Minimally invasive colorectal surgery is characterized by significant procedural variability, a difficult learning curve, and complications that affect patient outcomes. Video-based assessment offers an opportunity to generate data-driven insights to reduce variability, optimize training, and improve surgical performance. This mixed-methods study aimed to develop and validate a video-based assessment tool for workflow analysis across minimally invasive colorectal procedures.

Methods: An international steering committee panel of seven members coordinated a three-round modified Delphi process. Experts in colorectal surgery and video-based assessment were invited to reach consensus (≥ 70% agreement) on the phases and steps that describe workflows across colorectal surgery procedural videos. The resulting framework informed the development of ColoWorkflow. Four independent raters then applied ColoWorkflow to a multicentre video data set of laparoscopic and robotic colorectal procedures. Applicability and inter-rater reliability were evaluated.

Results: Of 66 invited experts, 41 (62%) from 11 countries completed round 1, 40 (98%) completed round 2, 20 attended the final online consensus meeting, and all (40 of 40) approved the final workflow descriptors. Consensus was achieved on nine procedure-agnostic phases (port placement and abdomen exploration; vascular dissection and ligation, mesocolon/mesorectum dissection, additional lymphadenectomy; colon and/or rectum mobilization; colorectal transection; anastomosis; completion of operation; preplanned additional procedures; unplanned procedures; extracorporeal procedures) and 34 procedure-specific steps that describe colorectal surgery workflows. ColoWorkflow was developed and applied to 54 colorectal surgery videos (left and right hemicolectomies, sigmoid and rectosigmoid resections, and total proctocolectomies) from five centres in four countries. The tool demonstrated broad applicability, as all phase and step labels, except one, were represented in at least one video. Inter-rater reliability was moderate, with a mean Cohen's κ value of 0.72 for phases and 0.67 for steps. Most discrepancies arose at phase transitions and step boundary definitions.

Conclusion: Based on this preliminary work, ColoWorkflow is a promising tool for workflow analysis across minimally invasive colorectal procedures. Its adoption may help standardize training, accelerate competency acquisition, and advance data-informed surgical quality improvement.

Keywords: lower gastrointestinal tract, surgical data science, artificial intelligence, quality improvement, robotic surgery, laparoscopic surgery
Received: December 30, 2025. **Revised:** February 25, 2026. **Accepted:** February 28, 2026

Introduction

Minimally invasive surgery has become the standard approach for benign and malignant elective colorectal resections owing to its reduced morbidity, shorter hospital stays, and faster recovery compared with open surgery[1,2]. Despite these advantages, outcomes after minimally invasive 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[4–6]. Even among experienced centres and surgeons, heterogeneity in intraoperative techniques and dissection sequences contributes to variation in operating time, blood loss, complication risks, and postoperative recovery[7]. Addressing this variability is critical to facilitating training, benchmarking performance, standardizing care, and in improving the quality of minimally invasive CRS.

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

A standardized VBA tool for workflow analysis could therefore quantify procedural variability in minimally invasive CRS, help identify and benchmark best practices, and establish the foundation for automated workflow recognition.

The aim of this mixed-methods study was to develop and validate a VBA tool for comprehensive workflow characterization across laparoscopic and robotic colorectal procedures. It was hypothesized that a consensus-derived, procedure-agnostic framework could reliably describe workflows across laparoscopic and robotic colorectal operations.

Methods

Study design

This mixed-methods study comprised two components: a modified Delphi process to establish expert consensus on workflow descriptors for minimally invasive CRS; and the development and validation of a VBA tool, ColoWorkflow, derived from the consensus framework. The Delphi process was conducted and reported in accordance with ACCORD guidelines[15].

Ethical approval for the collection and analysis of deidentified surgical videos was obtained under the OPERATE protocol (ID 6456) from Fondazione Policlinico Universitario Agostino Gemelli IRCCS (Rome, Italy). All videos were pseudonymized before analysis, and no identifiable patient information was included.

Delphi consensus

A three-round modified Delphi study was undertaken to achieve consensus on the phases and steps describing workflows in minimally invasive CRS. A multidisciplinary steering committee of seven members from France, Italy, and Spain oversaw the process. The steering committee included expertise in CRS, VBA, and AI. Expert participants were nominated by the steering committee based on demonstrated experience in CRS and/or VBA. No predefined criteria were set to define expertise.

The steering committee developed initial workflow definitions by means of a comprehensive review of key surgical textbooks and literature[16–23] and semistructured interviews with colorectal surgeons (*Appendix S1*). Hierarchical task analysis was performed following the principles popularized by Annett[24] to deconstruct procedures into a sequence of phases and steps. Definitions followed the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) consensus framework[25], in which phases represent the highest-level temporal components of an operation and steps denote procedure-specific segments that achieve a discrete clinical objective. For example, anastomosis is a phase common across CRS procedures whereas rectal stump perforation and stapler firing is a specific step in left-sided anastomoses. Preliminary workflow definitions were piloted on a multicentre data set comprising laparoscopic and robotic CRS videos from five institutions to ensure broad procedural representation. Members of the steering committee are affiliated with four of the institutions that contributed CRS data specifically for this project, whereas the fifth institution publicly released CRS data[16]. These videos were not selected and no minimum number of videos per institutions was required. 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, San Francisco, CA, USA). Information on participants' demographics, clinical experience, and country of practice was collected. Each proposed phase or step was rated on two 5-point Likert scales: inclusion relevance; and clarity and completeness of description. Participants could also provide free-text comments, suggest edits, or propose new items. Consensus thresholds were based on previous Delphi studies[26]: items with ≥ 70% agreement (Likert ≥ 4 of 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, the steering committee deidentified, aggregated, and reviewed feedback 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 had been achieved. Revised definitions were circulated to all participants for final review and approval.

Development and validation of ColoWorkflow

ColoWorkflow was constructed using the consensus-derived workflow descriptors. 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 facilitate consistent annotation, no overlap between phases and between steps was allowed (that is each analysed video frame was labelled with a single phase and a single step); an example of ColoWorkflow temporal annotations is available in *Fig. S1*. An 'out-of-body' label was added to describe when the scope is outside of the abdomen, for cleaning for instance.

To assess the tool applicability, one surgeon from the steering committee used ColoWorkflow to annotate the above described data set of CRS procedural videos.

To assess tool reliability, the steering committee nominated four raters (a medical student, a physician, and 2 surgical residents) to apply ColoWorkflow independently to a representative subset of ten procedures, two per major procedure type (right hemicolectomy, left hemicolectomy, sigmoid resection, rectal resection, total proctocolectomy). All raters had previous VBA experience and used MOSaiC (IHU-Strasbourg, Strasbourg, France), a web-based platform for collaborative medical video analysis[27].

Finally, workflow elements with low inter-rater reliability (IRR) were reviewed collegially by raters and steering committee experts for root-cause analysis and refinements to ColoWorkflow. Discrepancies were classified as either due to start and end timestamps timing offsets, misinterpretation of phase/steps transitions, or disagreements about phase/steps classification. Based on this review, the visual cues used for annotation were refined, and guidance for the consistent application of ColoWorkflow was elaborated.

Outcomes of interest and statistical analysis

Outcomes of interest were the Delphi consensus agreement, and ColoWorkflow applicability and reliability.

Delphi participation metrics, including response and retention rates, were tracked by JotForm and summarized descriptively. Mean Likert scores for each round were compared to assess convergence towards consensus.

The applicability of annotations and the time spent annotating videos were tracked automatically by MOSaiC. Video and annotation duration were reported as mean(standard deviation, s.d.) values. To assess the mean analysis time per minute of video, the annotation time ratio was computed as the annotation time divided by the video duration. A ratio of 1.0 indicated real-time annotation speed.

IRR was quantified using Cohen's κ and percentage agreement[28]. Cohen's κ represents the proportion of agreement beyond that expected by chance and was 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. Cohen's κ was computed at 2.5 frames per second. 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)[28]. κ values were computed for all six rater pairs, and the mean was taken.

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

Results

Modified Delphi study and development of ColoWorkflow

The modified Delphi process was conducted over three rounds over 5 months (Fig. 1). Expert panel selection and initial item generation were completed during the first 2 months. Round 1 (3 December 2024) remained open for 14 days, followed by round 2 (3 January 2025) for 16 days, with one or two reminder e-mails per round. The final consensus meeting (round 3) was held online on 27 February 2025, followed by a 10-day feedback period for final approval.

Of 66 invited experts from 13 countries, 41 (62%) from 11 countries completed round 1. The panel was diverse in both geographical and procedural experience, with 40% having performed > 1000 colorectal procedures (Table S1). Forty participants (98%) continued to round 2. The initial draft included 18 phase items and 74 step items describing minimally invasive CRS workflows. Items achieving ≥ 70% agreement (Likert score ≥ 4 of 5) with minimal qualitative feedback (≤ 2 revision suggestions) were accepted. Mean Likert scores increased between rounds for phase items and remained stable for step items. Most steps (77%) were accepted in round 1, with 20% of the revised steps subsequently accepted in round 2. Accepted phase items increased from 22 to 50% between Delphi rounds 1 and 2, indicating progressive consensus among participants (Table 1). The distribution of qualitative feedback across phases and steps is illustrated in Fig. S2, which indicates 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 two distinct phases: 'preplanned additional procedures' (for example additional resections) and 'unplanned procedures' (for example rectification of an intraoperative adverse event). 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 nine phases (port placement and abdomen exploration; vascular dissection and ligation, mesocolon/mesorectum dissection, additional lymphadenectomy; colon and/or rectum mobilization; colorectal transection; anastomosis; completion of operation; preplanned additional procedures; unplanned procedures; extracorporeal procedures) and 34 steps and their description.

These phases and steps were used to build ColoWorkflow. Start and end visual cues were defined based on a pilot annotation. An out-of-body phase was added to describe when the scope is outside of the abdomen, for cleaning, for instance. Figure 2 represents ColoWorkflow, plotting common phases and how steps can repeat across phases. Complete phase and step definitions are detailed in Tables S2 and S3.

Validation of ColoWorkflow

ColoWorkflow was applied to 54 minimally invasive CRS procedural videos encompassing: left hemicolectomies (10 patients), right hemicolectomies (12), rectal resections (10), sigmoid resections (12), and total proctocolectomies (10). Video duration ranged from 82 to 200 minutes (min) with a mean(s.d.) of 142.19(59) min (Table S4). All phase and step labels except those for ileoanal pouch creation were represented in at least one video, with no missing or redundant labels reported. Each video contained a mean of 7.52(0.79) phases and 16.46(2.41) steps. Mean phase duration was 18.91(17.46) min, and mean step duration was 6.49(7.01) min. It took a mean of 78(32) min to apply ColoWorkflow to each video, with a mean annotation time ratio of 0.59(0.23). Right hemicolectomy was the more time-consuming procedure to annotate, with a mean annotation time ratio of 0.77(0.32). A complete description of the data set and annotation effort is available in Table S4.

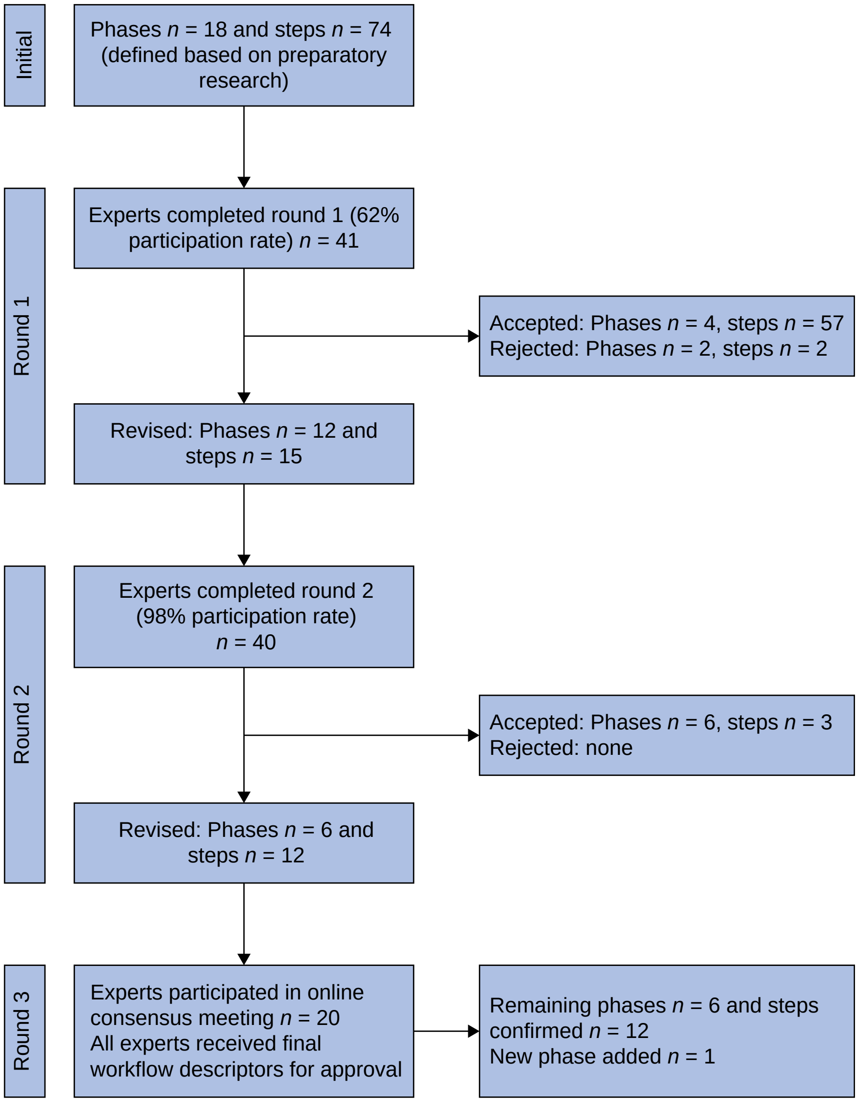

Fig. 1 Flow chart showing modified Delphi procedure

Experts were asked to find consensus on naming of phases and steps as well as their description.

On a subset of ten videos annotated by four independent raters, the mean IRR measured with Cohen's κ was 0.72 for phases and 0.67 for steps, indicating moderate agreement. Raters with more surgical experience tended to show higher IRR (*Table S4*). Labels with low κ scores or high IRR were examined during structured debriefings; *Fig. 3* shows Cohen's κ values across phases, and *Fig. S3* Cohen's κ values across steps. A qualitative review revealed that disagreements stemmed primarily from inherent challenges in interpreting surgical videos, such as ambiguous visual cues and overlapping tasks, rather than from inconsistent annotation practices. Raters' feedback and insights from the qualitative review were used to refine the ColoWorkflow VBA framework. Common pitfalls and recommended strategies to improve annotations are summarized in *Table 2*, and further guidance is provided in *Appendix S2*.

Discussion

This mixed-methods study developed and provided early validity evidence for ColoWorkflow, a VBA tool designed to analyse procedural workflows across laparoscopic and robotic colorectal procedures. Consensus on 9 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 ColoWorkflow to 54 surgical videos demonstrated its feasibility, consistency, and IRR, providing preliminary evidence of response process and internal structure validity according to Messick's framework[29]. Collectively, these findings support ColoWorkflow as a robust foundation for standardized workflow analysis in minimally invasive CRS.

Table 1 Results of modified Delphi procedure

Status	Round 1		Round 2	
	Phases	Steps	Phases	Steps
Accepted	4 (22%)	57 (77%)	6 (50%)	3 (20%)
Revised	12 (67%)	15 (20%)	6 (50%)	12 (80%)
Rejected	2 (11%)	2 (3%)	0 (0%)	0 (0%)

Values are *n* (%). The number of accepted, revised, and rejected items is reported for rounds 1 and 2. Round 3 comprised a collegial discussion until unanimous agreement had been reached.

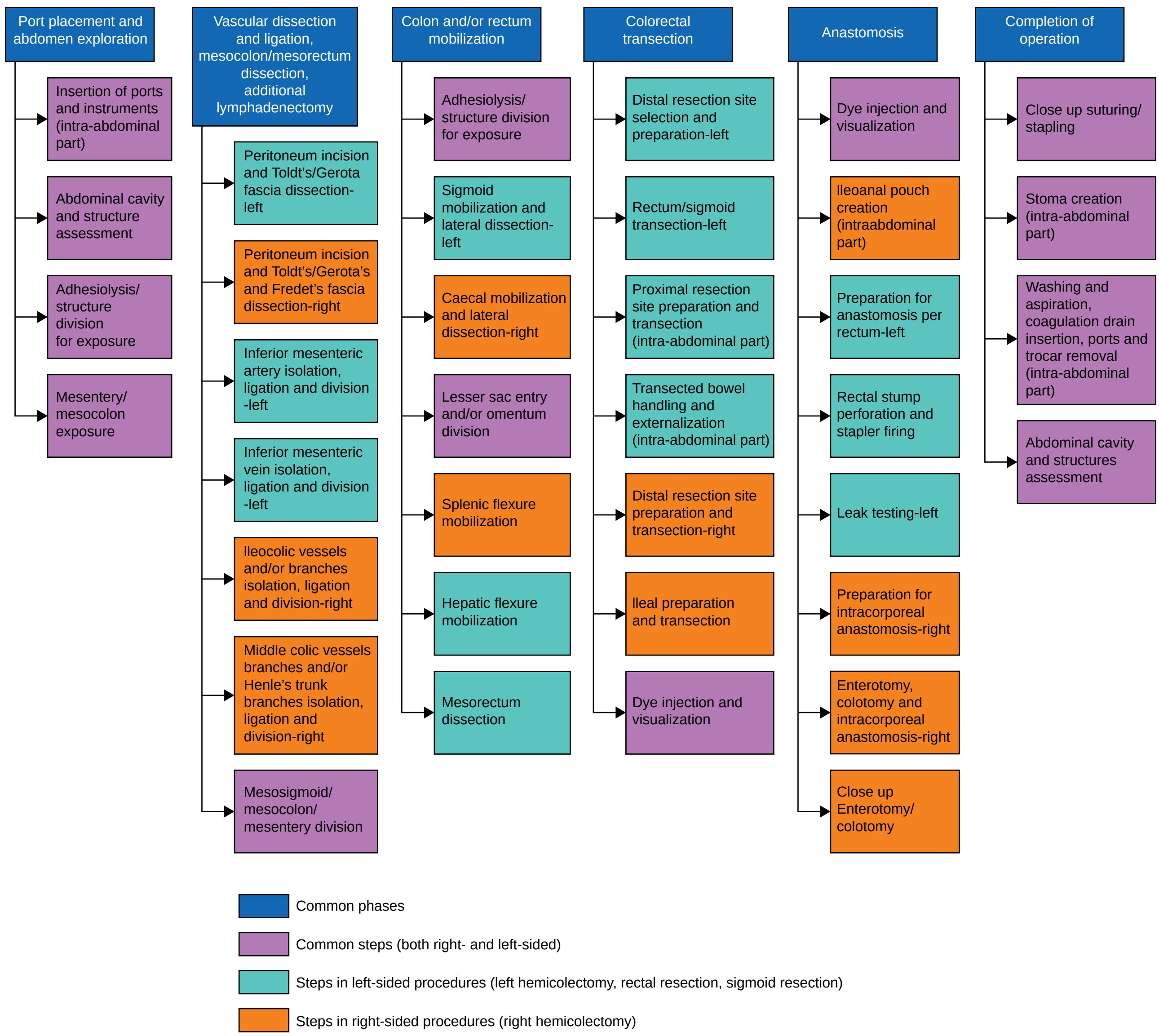

Fig. 2 Representation of ColoWorkflow

To improve readability, the following phases were not included in the figure: preplanned additional procedures, unplanned procedures, and extracorporeal procedures.

The present findings complement and extend previous work on the VBA of CRS workflow. Previous Delphi-based frameworks have focused on individual procedures, such as laparoscopic right colectomy or sigmoidectomy[20,21,30,31], or on discrete segments, such as splenic flexure mobilization[32]. More recently, the Heidelberg Colorectal data-set introduced a phase-based framework generalizable across left-sided CRS procedures[16]. ColoWorkflow builds on top of previous frameworks, integrating procedure-agnostic phases with procedure-specific steps within a unified hierarchical structure that can be applied across different laparoscopic and robotic CRS procedures. It is flexible in 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 approximately 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, whereas others reflect inherent variability in surgical technique. Disagreement often centred on terminology rather than content, emphasizing the importance of linguistic precision in workflow modelling[33]. The present findings also highlight the challenge of balancing annotation granularity for clinical interpretation with the coarser segmentation needed for efficient AI model training[34,35].

The moderate IRR (κ 0.6–0.8) suggests that ColoWorkflow is reliable despite the complexity of comprehensively assessing granular workflow elements across multicentre colorectal procedures. An analysis of disagreements informed the refinement of ColoWorkflow, integrating expert-based consensus recommendations with practical annotation insights.

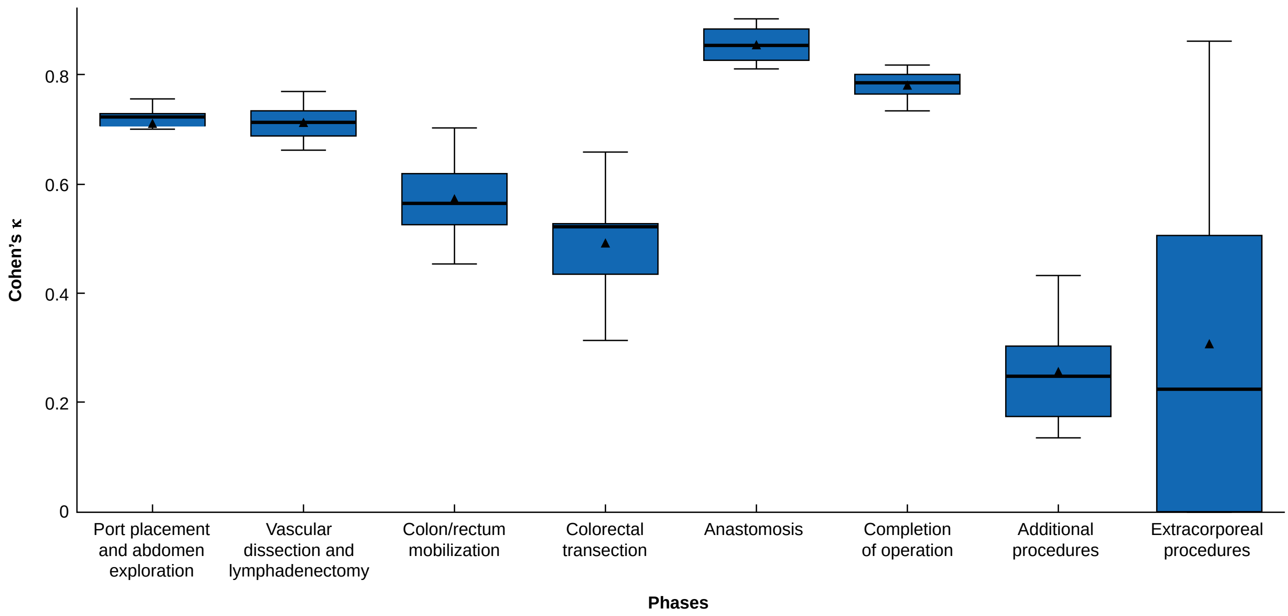

Fig. 3 Box plots showing Cohen's κ values across phases

Triangles represent mean values, and bold lines median values. Boxes span the interquartile range, and whiskers extend to show the range. The open circle represents an outlier. The phases preplanned additional procedures and unplanned procedures were analysed as one phase, labelled additional procedures, to allow reporting of reliable κ statistics despite low incidence.

Table 2 Challenges contributing to inter-rater variability and suggested mitigation strategies

Challenge	Suggested mitigation
Ambiguous start and end of phases and steps	Base transition on clear visual clues, provide examples
Hard-to-recognize anatomical landmarks	The VBA tool should be applied by surgeons or lay people with specific training and support (image primers, surgeons reviews, etc.) on videos of sufficient quality
Flickering phases/steps	Establish *a priori* whether to optimize for workflow stability with goal-oriented annotations (i.e. labelling a certain phase/step until it has been completed, high threshold to change phase/step) or for granularity with action-oriented annotations (i.e. labelling a certain phase/step based on the action being performed, low threshold to change phase/step)
Learning and fatigue effects	Schedule interim calibration meetings. Distribute analysis over time

VBA, video-based assessment.

However, some disagreements seemed to stem from variability in mental models, familiarity with the VBA tool, and labelling strategies, reflecting inherent and often unavoidable differences in conceptualizing workflow definitions[36]. A structured orientation phase before any large-scale VBA, including a co-review of sample videos, could help but it will likely not get to a strong (κ 0.80–0.90) or almost perfect (κ > 0.90) IRR. Higher IRR could facilitate applicability and the development of AI to automated workflow assessments. To support such use cases, AI models for surgical workflow recognition should be developed to be robust and generalizable across the high variability of real surgical practices, rather than simplifying the annotation task at the expense of clinical value. In addition, ColoWorkflow modular and hierarchical structures allow AI researchers to tackle the problem step-by-step, for instance, focusing on coarser workflow elements (for example phases) first.

Altogether, several aspects distinguish ColoWorkflow from other workflow analysis frameworks. Its thorough development through a modified Delphi exercise involving over 40 international experts and its application to a multicentre data set spanning CRS procedures provide early but convincing evidence of the tool's validity. The hierarchical and flexible structure of ColoWorkflow enables analysis across procedure types, surgical platforms (laparoscopic and robotic), and institutional protocols; this generalizability should enhance applicability, cross-centre benchmarking, and continuous professional development. Focusing on observable actions rather than prescriptive sequencing, it reflects real-world diversity while maintaining analytical rigour, 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. Regarding the Delphi process, the collegial discussion during the final round may have reduced independent expression owing to group dynamics. Nonetheless, such discussion followed two independent consensus rounds and was critical to ensuring convergence on workflow descriptors, likely contributing to ColoWorkflow's face and content validity. In addition, despite broad international participation, the expert panel may not have fully captured regional or technique-specific variations in surgical practice. With regard to the validation, although the multicentre data set included multiple colorectal operations, some procedures were not represented (for example pouch creation), limiting the current generalizability and applicability of the evidence. Nevertheless, more steps and procedural variations should be easy to integrate into ColoWorkflow thanks to its hierarchical design and flexibility in phase and step sequencing. In addition, among the four independent raters, none was an expert colorectal surgeon. Although including more experienced raters might have improved IRR, it was preferred to enrol medical students and residents trained on VBA as workflow analysis does not entail assessments of high surgical semantic (for example analysing anastomosis quality) and, considering how overburdened expert surgeons are, a tool trainees can use should facilitate implementation. Altogether, the validity evidence provided here should be considered preliminary and limited to face, content, response process, and internal structure, according to Messick's validity framework. Future validation studies, including a larger, more representative patient pool, more independent raters (including experts), assessment of IRR, and correlation with clinical outcomes, are warranted.

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

In conclusion, ColoWorkflow represents a promising framework for analysing workflows across laparoscopic and robotic CRS procedures. Through a rigorous consensus and early 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.

Collaborators

The CompSurg Colorectal Experts Group: M. Adamina (Inselspital, Bern University Hospital, Bern, Switzerland); G. Anania (University of Ferrara, Ferrara, Italy); A. Arezzo, M. Vannucci (University of Turin, Turin, Italy); W. A. Bemelman (Amsterdam University Medical Centers, Amsterdam, The Netherlands); P. P. Bianchi (ASST Santi Paolo e Carlo, Milan, Italy); U. Bracale (University of Salerno, Salerno, Italy); F. Brucchi (University of Milan, Milan, Italy); R. Cahill (University College Dublin, Dublin, Ireland); E. Cassinotti (Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy); M. Chand (University College London, London, UK); D. D'Ugo, G. Quero (Fondazione Policlinico Universitario Gemelli IRCCS, Rome, Italy); P. De Nardi (IRCCS San Raffaele Scientific Institute, Milan, Italy); E. Espin-Basany (Hospital Vall d'Hebron, Barcelona, Spain); D. Harji (Manchester University NHS Foundation Trust, Manchester, UK); J. Khan (Portsmouth Hospitals University NHS Trust, Portsmouth, UK); A. Lapergola (University Hospitals of Strasbourg, Strasbourg, France); G. G. Laracca (AOU Sant'Andrea, Rome, Italy); I. Montroni, E. Rausa (Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy); N. Okamoto (National Cancer Center Hospital East, Chiba, Japan); M. Ortenzi (Università Politecnica delle Marche, Ancona, Italy); A. Parvaiz (Champalimaud Foundation, Lisbon, Portugal); G. Pernazza (San Giovanni Addolorata Hospital, Rome, Italy); S. Rausei (ASST Sette Laghi, Varese, Italy); G. Rizzo, V. Tondolo (Ospedale Isola Tiberina Gemelli Isola, Rome, Italy); G. Spolverato (University of Padova, Padova, Italy); P. Sylla (Mount Sinai Hospital, New York, NY, USA); J. Torkington (University Hospital of Wales, Cardiff, UK).

Funding

This work 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 programme 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 Programme. This work was partially supported by French state funds managed by the Agence Nationale de la Recherche (ANR) under grant ANR-10-IAHU-02 (IHU Strasbourg).

Acknowledgements

This study was not preregistered in an independent institutional registry.

Disclosure

P.M. and N.P. are co-founders and shareholders of Scialytics. J. Khan and E. Espin-Basany serve as proctors for Intuitive Surgical. I. Montroni has received compensation for participation in courses organized by Olympus and Fresenius Kabi. M. Ortenzi is a paid collaborator with Theator. W. A. Bemelman is a shareholder in Semiflex. The authors declare no other conflict of interest.

Supplementary material

Supplementary material is available at *BJS Open* online.

Data availability

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

References

1. Carmichael JC, Keller DS, Baldini G, Bordeianou L, Weiss E, Lee L *et al.* 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. *Dis Colon Rectum* 2017;**60**:761–784
2. Greco M, Capretti G, Beretta L, Gemma M, Pecorelli N, Braga M. Enhanced recovery program in colorectal surgery: a

meta-analysis of randomized controlled trials. *World J Surg* 2014; **38**:1531–1541
3. Hanna GB, Mackenzie H, Miskovic D, Ni M, Wyles S, Aylin P *et al.* Laparoscopic colorectal surgery outcomes improved after national training program (LAPCO) for specialists in England. *Ann Surg* 2022;**275**:1149–1155
4. McArdle CS, Hole D. Impact of variability among surgeons on postoperative morbidity and mortality and ultimate survival. *BMJ* 1991;**302**:1501–1505
5. Bennett CL, Stryker SJ, Ferreira MR, Adams J, Beart RW Jr. The learning curve for laparoscopic colorectal surgery: preliminary results from a prospective analysis of 1194 laparoscopic-assisted colectomies. *Arch Surg* 1997;**132**:41–44
6. Keller DS, Stulberg JJ, Lawrence JK, Samia H, Delaney CP. Initiating statistical process control to improve quality outcomes in colorectal surgery. *Surg Endosc* 2015;**29**:3559–3564
7. Curtis NJ, Foster JD, Miskovic D, Brown CSB, Hewett PJ, Abbott S *et al.* Association of surgical skill assessment with clinical outcomes in cancer surgery. *JAMA Surg* 2020;**155**:590–598
8. Reitz ACW, Massarweh NN. Video-based assessment of surgical quality—will video kill the radio star? *JAMA Netw Open* 2024;**7**: e246477
9. Pugh CM, Hashimoto DA, Korndorffer JR. The what? How? And who? Of video based assessment. *Am J Surg* 2021;**221**:13–18
10. Balvardi S, Kammili A, Hanson M, Mueller C, Vassiliou M, Lee L *et al.* The association between video-based assessment of intraoperative technical performance and patient outcomes: a systematic review. *Surg Endosc* 2022;**36**:7938–7948
11. Brajcich BC, Stulberg JJ, Palis BE, Chung JW, Huang R, Nelson H *et al.* Association between surgical technical skill and long-term survival for colon cancer. *JAMA Oncol* 2021;**7**:127–129
12. Stulberg JJ, Huang R, Kreutzer L, Ban K, Champagne BJ, Steele SR *et al.* Association between surgeon technical skills and patient outcomes. *JAMA Surg* 2020;**155**:960–968
13. Varban OA, Thumma JR, Finks JF, Carlin AM, Ghaferi AA, Dimick JB. Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy: a video-based study. *Ann Surg* 2021;**273**:766–771
14. Mascagni P, Alapatt D, Sestini L, Altieri MS, Madani A, Watanabe Y *et al.* Computer vision in surgery: from potential to clinical value. *Npj Digit Med* 2022;**5**:163
15. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P *et al.* ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med* 2024;**21**:e1004326
16. Maier-Hein L, Wagner M, Ross T, Reinke A, Bodenstedt S, Full PM *et al.* Heidelberg colorectal data set for surgical data science in the sensor operating room. *Sci Data* 2021;**8**:101
17. Farquharson M, Hollingshead J, Moran B. *Farquharson's Textbook of Operative General Surgery*. CRC Press, 2014
18. Wexner SD, Fleshman JW. *Colon and Rectal Surgery: Anorectal Operations*. Lippincott Williams & Wilkins, 2012
19. Corman ML. *Colon and Rectal Surgery*. Lippincott Williams & Wilkins, 2005
20. Dijkstra FA, Bosker RJI, Veeger NJGM, van Det MJ, Pierie JPEN. Procedural key steps in laparoscopic colorectal surgery, consensus through Delphi methodology. *Surg Endosc* 2015;**29**: 2620–2627
21. Palter VN, MacRae HM, Grantcharov TP. Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg* 2011;**201**:251–259
22. Kitaguchi D, Takeshita N, Matsuzaki H, Takano H, Owada Y, Enomoto T *et al.* Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc* 2020;**34**:4924–4931
23. Strey CW, Wullstein C, Adamina M, Agha A, Aselmann H, Becker T *et al.* Laparoscopic right hemicolectomy with CME: standardization using the 'critical view' concept. *Surg Endosc* 2018;**32**:5021–5030
24. Annett J. Hierarchical task analysis. In *Handbook of Cognitive Task Design*. CRC Press, 2003
25. Meireles OR, Rosman G, Altieri MS, Carin L, Hager G, Madani A *et al.* SAGES consensus recommendations on an annotation framework for surgical video. *Surg Endosc* 2021; **35**:4918–4929
26. Kommer M, Hawthorne C, Moss L, Piper I, O'Kane R, Czosnyka M *et al.* International e-Delphi survey to define best practice in the reporting of intracranial pressure monitoring recording data. *Brain Spine* 2024;**4**:102860
27. Mazellier JP, Boujon A, Bour-Lang M, Erharhd M, Waechter J, Wernert E *et al.* MOSaiC: a Web-based Platform for Collaborative Medical Video Assessment and Annotation. *arXiv* 2023
28. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica* 2012;**22**:276–282
29. Messick S. Standards of validity and the validity of standards in performance asessment. *Educ Meas Issues Pract* 1995;**14**:5–8
30. Sica GS, Anania G, Fiorani C, Siragusa L, Vinci D, Caricato M *et al.* 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* 2025;**9**:zraf066
31. Lorenzon L, Biondi A, Carus T, Dziki A, Espin E, Figueiredo N *et al.* Achieving high quality standards in laparoscopic colon resection for cancer: a Delphi consensus-based position paper. *Eur J Surg Oncol* 2018;**44**:469–483
32. Lee YS. Three surgical approaches of laparoscopic splenic flexure mobilization. *J Minim Invasive Surg* 2019;**22**: 85–86
33. Liu Z, Chen K, Wang S, Xiao Y, Zhang G. Deep learning in surgical process modeling: a systematic review of workflow recognition. *J Biomed Inform* 2025;**162**:104779
34. Padoy N, Horn M, Feussner H, Marie-Odile B, Navab N. Recovery of surgical workflow: a model-based approach. In: *21st International Congress and Exhibition on Computer Assisted Radiology and Surgery*, 2007
35. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW *et al.* Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg* 2019;**270**:414–421
36. Ward TM, Fer DM, Ban Y, Rosman G, Meireles OR, Hashimoto DA. Challenges in surgical video annotation. *Comput Assist Surg (Abingdon)* 2021;**26**:58–68