Development of a Reliable Surgical Quality Assurance System for 2-stage Esophagectomy in Randomized Controlled Trials

Alexander Harris, PhD, James Butterworth, MSc,* Piers R. Boshier, PhD,* Hugh MacKenzie, PhD,* Masanori Tokunaga, PhD,† Hideki Sunagawa, PhD,‡ Stella Mavroveli, PhD,* Melody Ni, PhD,* Sameh Mikhail, MD,§ Chi-Chuan Yeh, PhD,¶ Natalie S. Blencowe, PhD,||** Kerry N. L. Avery, PhD,** Richard Hardwick, MD,†† Arnulf Hoelscher, PhD,‡‡ Manuel Pera, PhD,§§ Giovanni Zaninotto, MD,* Simon Law, PhD,¶¶ Donald E. Low, MD,|||| Jan J. B. van Lanschot, PhD,*** Richard Berrisford, ChM,††† Christopher Paul Barham, MD,|| Jane M. Blazeby, MD,||** and George B. Hanna, PhD*✉*

Objective: The aim was to develop a reliable surgical quality assurance system for 2-stage esophagectomy. This development was conducted during the pilot phase of the multicenter ROMIO trial, collaborating with international experts.

Summary of Background Data: There is evidence that the quality of surgical performance in randomized controlled trials influences clinical outcomes, quality of lymphadenectomy and loco-regional recurrence.

Methods: Standardization of 2-stage esophagectomy was based on structured observations, semi-structured interviews, hierarchical task analysis, and a Delphi consensus process. This standardization provided the structure for the operation manual and video and photographic assessment tools. Reliability was examined using generalizability theory.

Results: Hierarchical task analysis for 2-stage esophagectomy comprised fifty-four steps. Consensus (75%) agreement was reached on thirty-nine steps, whereas fifteen steps had a majority decision. An operation manual and record were created. A thirty five-item video assessment tool was developed that assessed the process (safety and efficiency) and quality of the end product (anatomy exposed and lymphadenectomy performed) of the operation. The quality of the end product section was used as a twenty seven-item photographic assessment tool. Thirty-one videos and fifty-three photographic series were submitted from the ROMIO pilot phase for assessment. The overall G-coefficient for the video assessment tool was 0.744, and for the photographic assessment tool was 0.700.

Conclusions: A reliable surgical quality assurance system for 2-stage esophagectomy has been developed for surgical oncology randomized controlled trials.

Ethical approval: 11/NW/0895 and confirmed locally as appropriate, 12/SW/0161, 16/SW/0098.

Trial registration number: ISRCTN59036820, ISRCTN10386621.

Keywords: esophageal cancer, esophagectomy, surgical quality assurance

(*Ann Surg* 2022;275:121–130)

Surgery remains the cornerstone in the multimodality management of esophageal cancer. The outcome of esophagectomy is surgeon-dependent with wide variability in surgical technique and perioperative management. Systematic reviews of open and minimally invasive surgery have shown that formal lymphadenectomy was not performed in most studies and lymph node harvest fell below the minimum number recommended to achieve survival benefits.[1,2] The problem of surgical variability is heightened in randomized controlled trials (RCTs) where the quality of surgery could influence the final outcome and might compromise the generalizability of results.[3,4] A systematic review assessing surgery within RCTs for the treatment of esophagogastric cancer demonstrated significant heterogeneity in study design and quality assurance.[5] Surgical quality indicators identified in this systematic review were: pre-trial standardization of surgical technique, credentialing of surgeons to enter into the trial, and monitoring of surgical performance. Those measures influence in-hospital mortality, the quality of lymphadenectomy, and loco-regional recurrence.

Surgical quality assurance (SQA) was developed within the context of the pilot phase of the multicenter randomized oesophagectomy – minimally invasive or open (ROMIO) trial, which is comparing minimally invasive esophagectomy with open surgery,[6] where surgery is the trial intervention within a multimodality approach to the treatment of esophageal cancer. The specific

From the *Department of Surgery & Cancer, Imperial College London, United Kingdom; †Department of Gastrointestinal Surgery, Tokyo Medical and Dental University, Japan; ‡Department of Gastroenterological Surgery, New Tokyo Hospital, Japan; §Department of General Surgery, Faculty of Medicine University of Cairo, Egypt; ¶Department of Surgery, National Taiwan University Hospital, Taiwan; ||Division of Surgery, University Hospitals Bristol NHS Foundation Trust, United Kingdom; **National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol, United Kingdom; ††Upper gastrointestinal Unit, Cambridge University Hospitals NHS Foundation Trust, United Kingdom; ‡‡Center for Esophageal and Gastric Surgery, Agaplesion Markus Hospital, Germany; §§Department of Surgery, Hospital del Mar, Spain; ¶¶Department of Esophageal and Upper Gastrointestinal Surgery, The University of Hong Kong Queen Mary Hospital, Hong Kong; ||||Department of Thoracic Surgery, Virginia Mason Medical Center, Seattle, WA; ***Department of Surgery, Erasmus MC University Medical Center, The Netherlands; and †††Department of Surgery, University Hospitals Plymouth NHS Trust, United Kingdom.

✉g.hanna@imperial.ac.uk.

Re-prints will not be available from the corresponding author.

Study was supported by the NIHR-HTA Grant 10/50/65, Randomized Oesophagectomy – Minimally Invasive or Open (ROMIO) Trial. London Deanery Simulation and Technology enhanced Learning Initiative (STeLI). The Great Britain Sasakawa Foundation.

This work was undertaken with the support of the Medical Research Council ConDuCT-II (Collaboration and innovation for Difficult and Complex randomized controlled Trials In Invasive procedures) Hub for Trials Methodology Research (MR/K025643/1). (http://www.bristol.ac.uk/social-community-medicine/centres /conduct2/) and Royal College of Surgeons of England Bristol Surgical Trials Centre. KNLA and JMB were supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. JMB holds an NIHR Senior Investigator award. The NIHR Imperial Biomedical Research Centre also provided support and infrastructure. PB is a NIHR Lecturer. The views expressed are those of the authors and not necessarily those of the MRC, UK National Health Service, National Institute for Health Research or Department of Health and Social Care.

The authors declare no conflict of interests.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/20/27501-0121

DOI: 10.1097/SLA.0000000000003850

objectives were to standardize the performance of 2-stage esophagectomy and develop a competency assessment tool for trial. The intended deliverables were an operation manual and reliable competency assessment tools.

METHODS

Standardization of 2-stage Esophagectomy

Semi-structured Interviews and Structured Observations

Semi-structured interviews and structured observations[7] were performed with surgeons at specialist centers for esophagogastric cancer surgery in the United Kingdom, United States of America and Japan (Glossary of terms used provided in Supplemental Digital Content 1, http://links.lww.com/SLA/C33). These investigated similarities and variations in clinical practice of 2-stage esophagectomy performed worldwide. Interviews were digitally audio-recorded (with consent) before being transcribed, checked for accuracy, and qualitatively analyzed using Thematic Analysis.[8,9] For contingency, short-hand written records were also kept in case of digital data loss. Structured observations were written in a research diary kept by the primary researcher (AHa). A second researcher (PB) assisted with intraoperative data collection in Japan and the United Kingdom. A debrief held at the end of each observed operation permitted comparison of notes between researchers, with video recordings of selected operations used to support further in-depth analysis remotely.

Hierarchical Task Analysis (HTA)

Findings from the published literature and on-line digital media were combined with Thematic Analysis of the semi-structured interviews and structured observations to create a HTA for 2-stage esophagectomy.[10] Several iterations were written and revised in consultation with a panel of senior esophagogastric cancer surgeons. The accuracy of each HTA was verified against live and video recorded 2-stage esophagectomies performed by senior esophagogastric cancer surgeons until no additional changes were identified. The final HTA was then tested against a series of subsequent operations, for which the primary researcher was present as an observer.

Delphi Consensus Process for the Esophagectomy HTA

Ten peer-nominated expert esophagogastric cancer surgeons were electronically invited to participate in a Delphi consensus process.[11] This method was selected to ensure the underlying face and content validity of the final assessment tool, and surgeon acceptance of the procedural HTA. Esophagogastric cancer surgeons involved during the development of the HTA were excluded. In total, 9 surgeons consented to participate in the Delphi process.

In the first Delphi round, each surgeon was provided with the final HTA and a questionnaire where they were requested to rate the requirement for each of the defined steps as either mandatory, optional, or prohibited.[12] This method of rating was in accordance with the requirements of the ROMIO trial protocol.[6] Additional free space was available for comments to be made regarding each individual step within the HTA. Completed questionnaires were returned electronically and analyzed by the primary researcher. An arbitrary consensus agreement of 75% was sought for each step.[13]

This process was repeated in the second Delphi round, during which the 9 respondents from the first Delphi round were electronically sent the percentage agreement and anonymized comments for each step of the HTA. The original HTA was re-sent, along with a new Delphi questionnaire. If respondents' answers remained outside of the majority agreement, they were asked to provide reasons for this in the comments section. For steps where consensus agreement could not be reached by the end of the second Delphi round, it was deemed acceptable for a majority opinion to be upheld if it reflected the findings of the HTA and its original evidence base (ie, triangulation of the published literature, semi-structured interviews, and structured observations).

Development of an Operation Manual and Note

A complete operation manual was constructed for surgeons based on the Delphi consensus approved HTA. The key operative steps for open and minimally invasive 2-stage esophagectomy were identical. Each step of the operation, both mandatory and optional, was described in detail with photographs illustrating the required en-bloc lymphadenectomy.

Given the length of the full manual, a separate summary document describing 10 essential steps for the abdominal and thoracic phases of the operation was also produced. The operation manual and summary of essential steps were approved by the ROMIO pilot phase steering committee.

A standardized operation note was constructed for 2-stage esophagectomy in an iterative process. It was designed to reflect the clinical requirement of providing a formal record of the operation performed and as a requirement of SQA. The body of the operation note included a tick box version of the operation manual, permitting surgeons to rapidly provide a detailed outline of the procedure performed, with white space boxes available for additional information.

Development of Video and Photographic Assessment Tools

The details for image capture and data transfer are supplied in Supplemental Digital Content 2, http://links.lww.com/SLA/C34.

Video Assessment Tool

In accordance with the Systems Engineering Initiative for Patient Safety model that describes a structure-process-outcome approach,[14] the results obtained from the semi-structured interviews and observations confirmed the importance of technical performance (process) and oncological quality (outcome) of the operation. After consideration of techniques that could permit independent remote blind evaluation of the technical performance and oncological quality of surgery, it was determined that a video assessment tool would best address all of these aspects.

An existing validated, consultant-level, surgical assessment tool[15,16] was deconstructed and its underlying principles adapted during the structural development of this video assessment tool. Elements relevant to the safety and efficiency of the operative process, and the oncological quality of the end product, were identified from Thematic Analysis of the semi-structured interviews and structured observations such that clear definitions for each of the terms used were composed.

Several different video assessment tools were written and piloted at the St Mary's Hospital, London, UK, over the course of 3 months. Each version placed a different emphasis on rating the element being assessed, with the intention to balance the operating surgeon's technical safety, efficiency, and oncological quality of their dissection in the final tool. (Please note that tasks have been labeled differently in the HTA and assessment tool, as the research evolved). The video assessment tool was approved for use by the ROMIO pilot phase steering committee and tested at the 2 centers involved in the pilot RCT.

Photographic Assessment Tool

The outcome section of the video assessment tool was purposefully written as a stand-alone photographic assessment tool, focusing on the completeness of the lymphadenectomy, and exposure of the relevant anatomical landmarks, should video submission have not proved feasible. The photographic tool was also approved for use by the ROMIO pilot phase steering committee and piloted as above.

Examining Reliability of the Video and Photographic Assessment Tools

Independent Assessment of Video and Photographic Records From the ROMIO Trial

Three esophagogastric cancer surgeons (1 based in the UK and 2 in Japan) were invited to assess and rate the intraoperative videos and photographs submitted by surgeons within the pilot ROMIO trial. Before commencing data analysis, the 3 assessors were trained by the senior author (GH) in 2 videoconference meetings on the pre-defined terms for using the assessment tools and to clarify any conceived variability. Before the second videoconference, each assessor was asked to independently rate 2 videos that had been chosen at random. These assessments formed a focal point for the discussion held during the second videoconference to minimize the discrepancy in assessments.

These 3 independent blinded assessors then applied the video and photographic esophagectomy assessment tools to rate each of the video and photographic records submitted to the pilot ROMIO trial. Assessments were completed on paper forms, which were subsequently submitted as a scanned PDF file and later transcribed into Excel (Microsoft Office, Redmond, WA).

Statistical Analysis

Generalizability (G) theory was used to assess the reliability of the video and photographic assessment tools because, in contrast to the classical test theory, G-theory includes several aspects of reliability (eg inter-rater, inter-test, and intra-test) in the same model. A decision (D) study was performed to determine the combination of components that yielded the maximum generalizability.[17] G-string software was used to conduct the generalizability theory, inter-rater reliability, and internal consistency analysis.[18] Cronbach alpha for internal consistency was performed using IBM SPSS statistics (Ver. 24, SPSS Inc., Chicago, IL) as part of cross-validation.

RESULTS

Standardization of 2-stage Esophagectomy

Semi-structured Interviews and Structured Observations

In total, 8 separate semi-structured interviews were performed with surgeons from the UK (n = 6) and the USA (n = 2). Themes arising from the qualitative analysis of these interviews are summarized in Supplemental Digital Content 3, http://links.lww.com/SLA/C35. Greater than fifty esophagectomies, performed by sixteen surgeons from the UK (n = 9), USA (n = 6), and Japan (n = 1), were observed in 7 different hospitals. Structured observation notes were combined with findings from the ‘‘operative procedure’’ theme identified from the semi-structured interviews and incorporated into the HTA.

HTA

Full details of the HTA are provided in Supplemental Digital Content 4, http://links.lww.com/SLA/C36. The abdominal component of 2-stage esophagectomy comprised 7 tasks, and the thoracic component comprised 6 tasks. Each task was then sub-divided into multiple steps. Overall, fifty-four steps were identified.

Delphi Consensus Process

Full details of the Delphi consensus process are provided in Supplementary Digital Content 5, http://links.lww.com/SLA/C37. In round 1 of the Delphi consensus process, 9 of the 10 invited surgeons responded representing the UK, Germany, Spain, Italy, the Netherlands, the USA, and Hong Kong. They reached consensus (75%) agreement on forty of the fifty-four steps comprising the HTA for esophagectomy according to whether each task was mandatory, optional, or prohibited.

In round 2 of the Delphi consensus process, the same 9 surgeons who had previously responded were provided with the results from the first round and asked to re-rate each task. This time, only 6 surgeons responded in the allotted timeframe. Thirty-nine tasks reached consensus agreement. Seven of the fifteen tasks without consensus agreement were the same in both rounds.

Given the diminishing number of responses, it was felt that a third round would not be beneficial. Furthermore, it was not logistically possible to gather all invited surgeons in person, or via conference call, to discuss their decisions. Accordingly, for tasks that did not reach consensus agreement in round 2, a majority decision was upheld if it reflected the findings of the evidence-based HTA. Thirteen of the fifteen tasks without agreement had a majority decision in the second round. Two of the fifteen tasks were equally split by respondents in the second round, but had had a majority decision in round 1. The earlier majority decision (based on a greater number of respondents) was therefore followed.

Development of an Operation Manual and Note

A summary describing 10 essential steps for both the abdominal and the thoracic phases of the operation can be found in Table 1. The final versions of the operation manual and note are provided in Supplemental Digital Contents 6, http://links.lww.com/SLA/C38 and 7, http://links.lww.com/SLA/C39.

Development of Video and Photographic Assessment Tools

The video and photographic assessment tools are shown in Figs. 1 and 2, respectively.

Examining Reliability of the Video and Photographic Assessment Tools

Thirty-one videos and 53 photographic series from patients undergoing 2-stage esophagectomy were submitted for assessment. The length of submitted video recordings varied widely, ranging from 1.0 minute to 447.0 minutes. In total, 4464.3 minutes of video footage were received, with a median duration of 119 minutes submitted per esophagectomy. Photographic submissions ranged from 2 images to 35 images per esophagectomy. In total, 451 images were received, with a median of 9 photographs per case.

Despite a large volume of data being submitted, the 3 assessors identified that there was also a significant amount of data missing. After an interim review, a videoconference between the 3 reviewers explored the possible reasons for missing data and potential strategies to mitigate its impact on data analysis. The original 3-point lymphadenectomy rating system of complete, incomplete, and not performed was deemed to be inadequate. Alternative solutions that were considered included making missing assessment values a mean value or coding them as not performed.[19] However, assessors were concerned that this would introduce bias and skew results. There was consensus that 2 additional categories could be used to re-code those parts of the assessment in which assessors were unable to provide a rating. The new categories acknowledged insufficient evidence for assessors to provide a rating (eg, videos with an obstructed field of view or blurred photographs) and absent data (ie, no video or photograph submitted). Overall, 32.3% of video and photographic data were absent. 6.8% of video data were insufficient for assessors to provide a rating, compared with 23.4% of photographic series.

TABLE 1. Essential Tasks for 2-stage Esophagectomy

Abdominal phase
Abdominal access
1. Confirm the absence of metastatic disease.
Diaphragmatic hiatus
2. Mobilize the gastroesophageal junction, resecting right and left paracardial lymph nodes (LN). (LN stations 1 and 2)
3. Resect a cuff of diaphragm and pleura to achieve a clear circumferential margin in advanced disease.
4. Dissect along the pre-aortic fascia.
Gastric mobilization
5. Mobilize the stomach based on the right gastroepiploic vessels
Celiac axis
6. Dissect LN tissue along the common hepatic artery, celiac artery, left gastric artery and proximal splenic artery. (LN stations 7, 8a, 9, 11p)
7. Ligate and divide the left gastric vein close to the portal vein and the left gastric artery at the celiac artery.
8. Dissect LN tissue from the left side of the celiac artery, to the left crus at the esophageal hiatus, and left side of Gerota's fascia.
9. Continue the dissection along the anterior surface of the proximal splenic artery towards the splenic hilum and ligate the posterior gastric vessels at their origin from the splenic artery.
Gastric tube
10. Create the gastric tube, removing tissue along the lesser curvature of the stomach. (LN stations 3a and 3b) This step may be done in the chest.
Thoracic phase
Thoracic access
1. Exclude metastatic disease in the chest.
Thoracic lymphadenectomy
2. Divide the inferior pulmonary ligament and ligate and divide the azygos arch.
3. Dissect along the pericardium until the left pulmonary vein is reached, including the left pleura in advanced disease.
4. Perform a sub-carinal lymphadenectomy (LN station 107) and clear both bronchi of LN tissue. (LN station 109)
5. Dissect the mediastinal pleura at the anterolateral border of the thoracic aorta and dissect along the pre-aortic fascia, from the proximal resection margin towards the diaphragm. (LN station 112)
6. Identify and ligate the thoracic duct at the proximal resection margin and above the diaphragm.
Specimen excision
7. Ensure that the thoracic part of the specimen is circumferentially free, from the previously completed diaphragmatic mobilization (performed during the abdominal phase) to at least the level of the aortic arch. (LN stations 108, 110, and 111)
8. Deliver the stomach into the right chest cavity, ensuring that the gastric tube can reach the site of anastomosis without tension or torsion.
9. Excise the specimen with suitable proximal and distal resection margins and send it to pathology as per the ROMIO trial protocol.
Anastomosis
10. Perform an esophago-gastrostomy using preferred technique.

LN indicates lymph nodes; ROMIO, randomized oesophagectomy – minimally invasive or open.

G-theory Results for the Video Assessment Tool

Generalizability analyses were performed to evaluate reliability of the 35-item video assessment tool with a fully crossed design using videos (V), items (I), and assessors (A), such that (V × I × A). In total, 93 assessment forms (31 videos rated by 3 assessors) of the 35-item video assessment tool were used in the analysis. Raw scores of the 35-item video assessment tool were generalized over the assessor and item. Overall reliability of video assessment was represented by a generalizability coefficient of G(AI) = 0.744. D-studies were performed to examine the effect of increasing numbers of assessors and video esophagectomies that they assessed (Fig. 3). The critical G coefficient of 0.8 was reached with 4 assessors each rating 26 video esophagectomies or 6 assessors each rating 16 video esophagectomies.

G-theory Results for the Photographic Assessment Tool

To evaluate the reliability of the 27-item photographic assessment tool, a fully crossed design using photographs (P), items (I), and assessors (A) such that (P × I × A) was used.

In total, 159 ratings (53 sets of operative photographs rated by 3 assessors) of a 27-item photographic assessment tool were used in the analysis. Raw scores of the 27-item photographic assessment tool were generalized over the assessor and item. Overall reliability of photographic assessment was represented by a generalizability coefficient of G(AI) = 0.700. D-studies were once again performed to examine the effect of increasing numbers of assessors and sets of photographs of esophagectomies that they assessed (Fig. 4). The critical G coefficient of 0.8 was reached with 6 assessors each rating 38 sets of esophagectomy photographs or 8 assessors each rating 33 sets of esophagectomy photographs.

Generalizability coefficients were also calculated separately and demonstrated consistently high reliability coefficients within the video and photographic assessment tools respectively. It was noted that video assessment had consistently higher reliability coefficients compared to photographic assessment across all tasks (Table 2).

Inter-rater Reliability and Internal Consistency

By treating 1 facet at a time as random, whilst fixing the other facets, it was possible to determine the equivalent of inter-rater reliability for the video assessment tool as: $Ep^2 = 0.492$. By setting the item as random and the assessor as fixed for the video assessment tool, it was possible to determine the equivalent of internal consistency as $Ep^2 = 0.991$, which was similar to the value calculated using SPSS with Cronbach alpha 0.986.

Through the same process, the inter-rater reliability and internal consistency for the photographic assessment tool were calculated as $Ep^2 = 0.438$ and $Ep^2 = 0.948$, respectively. Again, the internal consistency was similar to the value calculated using SPSS with Cronbach alpha 0.942.

Further analysis was performed to determine redundancy of items within each of the video and photographic assessment tools. Cronbach alpha remained constant at 0.986 on removal of any anatomical item from the video assessment tool, thereby demonstrating high inter-item reliability. In the photographic assessment tool, removal of the same anatomical items resulted in small improvements in Cronbach alpha. However, this variation was attributed to the high occurrence of absent data or insufficient evidence within this cohort.

ROMIO Centre:.......................... **ROMIO Case Reference:..................................** **Surgeon:……………………….**

Please tick the appropriate descriptions of the safety, efficiency, and quality of the end product for each task, according to the key below.

i) TECHNICAL SAFETY

Safe	No adverse events or near misses occurred.
Near miss	Potential harms were narrowly avoided.
Unsafe	Adverse event(s) that resulted in reversible harm occurred.
Dangerous	Adverse event(s) that resulted in permanent harm occurred.

ii) OPERATIVE EFFICIENCY

Optimal	Purposeful and progressive movements throughout.
Adequate	Some unnecessary movements, but generally progressive.
Inefficient	Repeated, unproductive, movements.
Poor	Wrong movements that compromised patient safety.

iii) QUALITY OF THE END PRODUCT

Complete	Anatomical structure is clearly demonstrated following complete dissection of all associated lymphatic (LN) tissue.
Incomplete	Incomplete LN clearance of the anatomical structure (quantify if possible please)

TASK 1: DIAPHRAGMATIC HIATUS

i) Safety	**Safe**	**Near miss**	**Unsafe**	**Dangerous**	**Comments**
	☐	☐	☐	☐	....

ii) Efficiency	**Optimal**	**Adequate**	**Inefficient**	**Poor**	**Comments**
	☐	☐	☐	☐	....

iii) Quality of end product	**Complete**	**Incomplete**	**Not performed**	**Comments**
Right crus	☐	☐	☐	....
Left crus	☐	☐	☐	....
Aorta	☐	☐	☐	....
Pericardium	☐	☐	☐	....
Right lung	☐	☐	☐	....
Left lung	☐	☐	☐	....

TASK 2: ABDOMINAL LYMPHADENECTOMY

i) Safety	**Safe**	**Near miss**	**Unsafe**	**Dangerous**	**Comments**
	☐	☐	☐	☐	....

ii) Efficiency	**Optimal**	**Adequate**	**Inefficient**	**Poor**	**Comments**
	☐	☐	☐	☐	....

iii) Quality of end product	**Complete**	**Incomplete**	**Not performed**	**Quantify if incomplete**
Portal vein	☐	☐	☐	....
Proper hepatic artery	☐	☐	☐	....
Common hepatic artery	☐	☐	☐	....
Celiac artery	☐	☐	☐	....
Left gastric artery (stump)	☐	☐	☐	....
Left gastric vein (stump)	☐	☐	☐	....
Proximal splenic artery	☐	☐	☐	....
Distal splenic artery	☐	☐	☐	....
Splenic vein	☐	☐	☐	....
Splenic hilum (if appropriate)	☐	☐	☐	....

Continued…

FIGURE 1. Video assessment tool.

TASK 3: THORACIC LYMPHADENECTOMY

i) Safety	Safe	Near miss	Unsafe	Dangerous	Comments
	☐	☐	☐	☐	....

ii) Efficiency	Optimal	Adequate	Inefficient	Poor	Comments
	☐	☐	☐	☐	....

iii) Quality of end product	Complete	Incomplete	Not performed	Quantify if incomplete
Carina	☐	☐	☐	....
Right main bronchus	☐	☐	☐	....
Left main bronchus	☐	☐	☐	....
Right pulmonary veins	☐	☐	☐	....
Left pulmonary veins	☐	☐	☐	....
Pericardium	☐	☐	☐	....
Aorta	☐	☐	☐	....

TASK 4: RECONSTRUCTION

i) Safety	Safe	Near miss	Unsafe	Dangerous	Comments
	☐	☐	☐	☐	....

ii) Efficiency	Optimal	Adequate	Inefficient	Poor	Comments
	☐	☐	☐	☐	....

iii) Quality of end product	Yes	No	Borderline	Comments
Viable color of gastric tube	☐	☐	☐	....
Lesser curve cleared of LN tissue	☐	☐	☐	....
Tension free anastomosis	☐	☐	☐	....
Appropriate approximation of sutures	☐	☐	☐	....

ANY OTHER COMMENTS

ROMIO ASSESSOR (Print):... **SIGN:**... **DATE:**..................

FIGURE 1. (Continued).

ROMIO Centre:......................... ROMIO Case Reference:................................. Surgeon:…………………

Instructions: Please tick the appropriate description for the quality of the end product for each task.

Key:
Complete Anatomical structure is clearly demonstrated following complete dissection of all associated lymphatic (LN) tissue.
Incomplete Incomplete LN clearance of the anatomical structure (Quantify if incomplete).

Task 1 Diaphragmatic hiatus	**Complete**	**Incomplete**	**Not performed**	**Quantify if incomplete**
Right crus	☐	☐	☐	………………………………………………..
Left crus	☐	☐	☐	………………………………………………..
Aorta	☐	☐	☐	………………………………………………..
Pericardium	☐	☐	☐	………………………………………………..
Right lung	☐	☐	☐	………………………………………………..
Left lung	☐	☐	☐	………………………………………………..

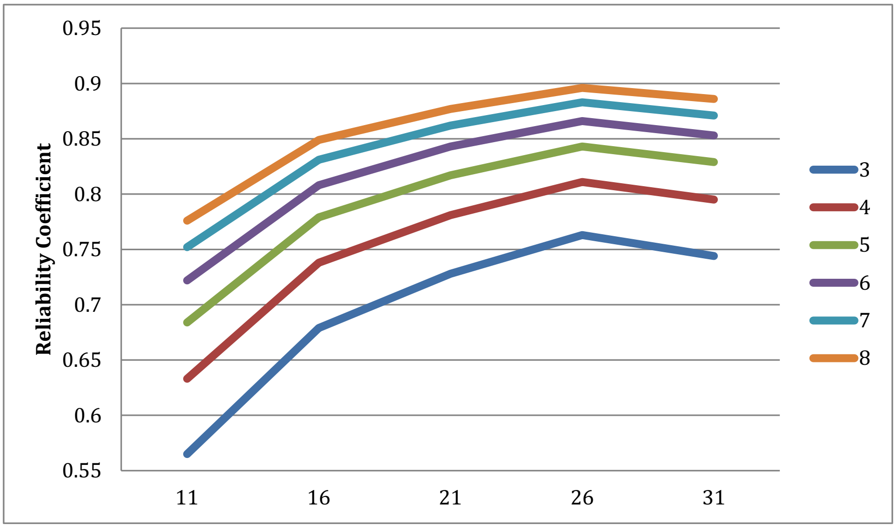
Task 2 Abdominal lymphadenectomy	**Complete**	**Incomplete**	**Not performed**	**Quantify if incomplete**
Portal vein	☐	☐	☐	………………………………………………..
Proper hepatic artery	☐	☐	☐	………………………………………………..
Common hepatic artery	☐	☐	☐	………………………………………………..
Celiac artery	☐	☐	☐	………………………………………………..
Left gastric artery (stump)	☐	☐	☐	………………………………………………..
Left gastric vein (stump)	☐	☐	☐	………………………………………………..
Proximal splenic artery	☐	☐	☐	………………………………………………..
Distal splenic artery	☐	☐	☐	………………………………………………..
Splenic vein	☐	☐	☐	………………………………………………..
Splenic hilum (if appropriate)	☐	☐	☐	………………………………………………..

Task 3 Thoracic lymphadenectomy	**Complete**	**Incomplete**	**Not performed**	**Quantify if incomplete**
Carina	☐	☐	☐	………………………………………………..
Right main bronchus	☐	☐	☐	………………………………………………..
Left main bronchus	☐	☐	☐	………………………………………………..
Right pulmonary veins	☐	☐	☐	………………………………………………..
Left pulmonary veins	☐	☐	☐	………………………………………………..
Pericardium	☐	☐	☐	………………………………………………..
Aorta	☐	☐	☐	………………………………………………..

Task 4 Reconstruction	**Yes**	**No**	**Borderline**	**Quantify if incomplete**
Viable color of gastric tube	☐	☐	☐	………………………………………………..
Lesser curve cleared of LN tissue	☐	☐	☐	………………………………………………..
Tension free anastomosis	☐	☐	☐	………………………………………………..
Appropriate approximation of sutures	☐	☐	☐	………………………………………………..

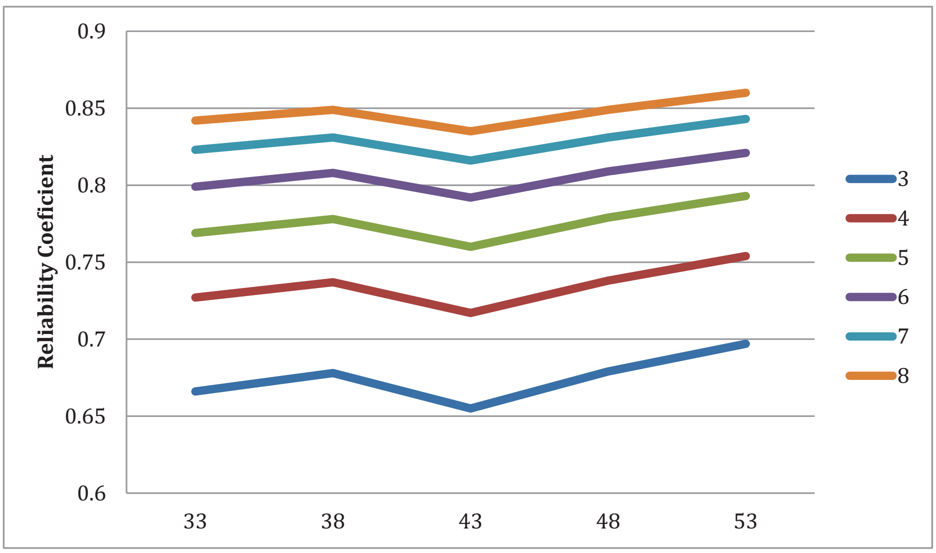
ROMIO ASSESSOR (Print):…………………………………………… SIGN:... DATE:………………

FIGURE 2. Photographic assessment tool.



The reliability coefficient (y-axis) is affected by the number of videos rated (x-axis). Colored lines represent a variable number of assessors.

FIGURE 3. D-study for the video assessment tool.



The reliability coefficient (y-axis) is affected by the number of sets of photographs rated (x-axis). Colored lines represent a variable number of assessors.

FIGURE 4. D-study for the photographic assessment tool.

TABLE 2. Generalizability Coefficients by Task

G Coefficient	Task 1. Diaphragm	Task 2. Abdomen	Task 3. Thorax	Task 4. Reconstruction
Photographs	0.758	0.770	0.695	0.759
Videos	0.983	0.919	0.975	0.961

DISCUSSION

This SQA system has been developed to assess the anatomy and tissues that remain after oncological resection to complement histological examination of the removed specimen, including resection margins and lymph node yield, which are the traditional markers of surgical quality. The study describes the development of key components of a SQA system that defines the operative standard for 2-stage esophagectomy and assesses operative competency. The deliverables were an operation manual and reliable video and photographic assessment tools for use within surgical oncology RCTs. Those deliverables address the required SQA measures identified in 2 systematic reviews to standardize surgical techniques, credential surgeons before entry into RCTs and monitor performance during the trial.[5,20]

The output of the standardization process provided the structure for the operation manual and basis for the competency assessment tools. A set of mandatory and optional steps was specified as a guide for surgical performance and a framework to examine variability of task execution between surgeons during the trial. Categorizing steps into mandatory and optional tasks allows for flexibility in surgical performance whilst maintaining a minimum quality standard.[12] Seventeen international esophago-gastric surgeons with trial experience participated in the interviews, observations, and Delphi consensus process to set a proposed standard for the trial. This international composition affords a degree of generalizability for the developed tools to be applied in esophageal RCTs worldwide. Nevertheless, this operative template may need to be modified by leading surgeons in specific trials to provide a balance between standardization and practicalities in exploratory and pragmatic trials. Although not ideal, the differences in surgical rigor between countries, cancer centers, and trial designs are a reality in surgery. The real performance of surgeons in the trial will be the practical level against which trial outcomes will be judged and explained.

The developed tools have a high level of content and face validity as they are based on a HTA and a Delphi consensus process of surgical performance. Both video and photographic assessment tools have a high reliability score using generalizability theory. However, capturing video or photographic data during open esophagectomy presents a challenge. This research shows that a high proportion of video and photographic data was absent or insufficient to rate performance. The operative time for esophagectomy and the potential intrusiveness of audio-visual recordings, given the restricted surgical access and limited operative field in open surgery, explains the challenge in capturing the image data. Clear instructions for data capturing and adequate resources and strong engagement from participating surgeons are required. An anecdotal observation from fieldwork performed in Japan, was how motivated surgeons were to capture high-quality video and photographic evidence of the procedure.

An alternative approach that could be explored is a short video recording, performed at key stages of the operation, to demonstrate the extent of the dissection and characteristics of the reconstruction. The benefits of this approach would be the avoidance of long video recordings and the frequent inadequacies of photographic images. The short recording would allow a dynamic snapshot of the operative field, permitting visualization of anatomical structures from multiple angles, and better assessment of conduit health and tension at the anastomosis. A limitation of such an approach would be the inability to assess the process, including safety and efficiency of operative tasks, as it would only show the quality of the end product. In addition, a feedback system will be developed for surgeons participating in the trial.[21] Nevertheless, the adoption of surgical assessment tools would be enhanced by overcoming potential practical challenges in routine practice that create the perception of being impractical and time consuming. Provision of measures that facilitate the ease and convenience of video/photo capture, sharing and assessment is critical for the uptake of SQA in clinical trials. Recording systems and instructions for imaging the operative field should be provided and tested at the outset of the trial. Future iterations of the tool could be hosted on a web-based platform to support the exchange and assessment through digital media. The time required for assessment could be shortened by limiting the assessment to short videos and/or photographs of the operative field at the end of the procedure. The ROMIO trial does; however, provide an opportunity to examine the tools' practicality and to consider any changes required to make such an approach more feasible.

The study has several limitations. The development of SQA was not based on clinical outcomes, but on a HTA of surgical procedures and a consensus view that is constrained by the definition and selection of surgical expertise. However, in the absence of established SQA methods, it is reasonable to start the process with observational data and expert consensus. The Delphi methodology has advantages not observed in other traditional qualitative methods.[22] Whilst anonymity was preserved across panel members and only the primary researcher could identify the responses, the expert nomination process could have biased the results.

In conclusion, a reliable SQA system for 2-stage esophagectomy has been developed for surgical oncology RCTs. Key components of SQA include standardization of 2-stage esophagectomy and assessment of competent performance. The predictive clinical validity of these assessment tools is still to be examined.

ACKNOWLEDGMENTS

The authors would like to acknowledge the input of Dr Chris Metcalfe, Professor of Medical Statistics & Co-director Bristol Randomized Trials Collaboration, with this research. We are grateful to Professor Mitsuru Sasako (Japan) and Professor David Farley (USA) for facilitating the interviews and task analysis conducted at the Hyogo College of Medicine in Japan and the Mayo Clinic in Rochester, USA. We also thank Jane M Blazeby, C Paul Barham, Dan Titcomb, Christopher Streets, Andrew Hollowood, Richard Krysztopik, Richard Berrisford, Tim Wheatley, and Grant Sanders who submitted operative data during the pilot phase of the study.

REFERENCES

1. Hanna G, Arya S, Markar S. Variation in the standard of minimally invasive esophagectomy for cancer - systematic review. *Semin Thorac Cardiovasc Surg*. 2012;24:176–187.
2. Boshier P, Anderson O, Hanna G. Transthoracic versus transhiatal esophagectomy for the treatment of esophagogastric cancer: a meta-analysis. *Ann Surg*. 2011;254:894–906.
3. Macdonald J, Smalley SR, Benedetti J, et al. Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N Engl J Med*. 2001;345:725–730.
4. Macdonald J, Smalley S, Benedetti J, et al. Postoperative combined radiation and chemotherapy improves disease-free survival and overall survival in resected adenocarcinoma of the stomach and GE junction. Results of intergroup study INT-0116 (SWOG 9008). *Eur J Cancer*. 2001;37:S10.
5. Markar S, Wiggins T, Ni M, et al. Assessment of the quality of surgery within randomised controlled trials for the treatment of gastro-oesophageal cancer: a systematic review. *Lancet Oncol*. 2015;16:e23–e31.
6. Avery K, Metcalfe C, Berrisford R, et al. The feasibility of a randomized controlled trial of esophagectomy for esophageal cancer - the ROMIO (randomized oesophagectomy: minimally invasive or open) study: protocol for a randomized controlled trial. *Trials*. 2014;15:200.
7. Marshall C, Rossman G. In: *Designing Qualitative Research*. 2nd ed., London: SAGE Publications; 1995.
8. Flick U. In: *An Introduction to Qualitative Research*. 4th ed., London: SAGE Publications; 2009.

9. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3:77–101.
10. Sarker S, Chang A, Albrani T, et al. Constructing hierarchical task analysis in surgery. *Surg Endosc*. 2008;22:107–111.
11. Palter V, MacRae H, Grantcharov T. Development of an objective evaluation tool to assess technical skills in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg*. 2011;201:251–259.
12. Blencowe N, Mills N, Cook JA, et al. Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. *Br J Surg*. 2016;103:1377–1384.
13. Vernon W. The Delphi technique: a review. *Int J Ther Rehabil*. 2009;16:69–76.
14. Carayon P, Hundt SA, Karsh BT, et al. Work system design for patient safety: the SEIPS model. *Qual Saf Health Care*. 2006;15(suppl I):i50–i58.
15. Miskovic D, Ni M, Wyles SM, et al. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surg Endosc*. 2012;26:796–803.
16. Miskovic D, Ni M, Wyles SM, et al. Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England. *Ann Surg*. 2013;257:476–482.
17. Bloch R, Norman G. Generalisability theory for the perplexed: a practical introduction and guide: AMEE guide no 68. *Med Teacher*. 2012;34:960–992.
18. Generalisability Theory. [cited 1st December 2017; Available at: http://fhsperd.mcmaster.ca/g_string/index.html
19. Ware J, Harrington D, Hunter DJ, et al. Missing data. *N Engl J Med*. 2012;367:1353–1354.
20. Foster J, Mackenzie H, Nelson H, et al. Methods of quality assurance in multicenter trials in laparoscopic colorectal surgery: a systematic review. *Ann Surg*. 2014;260:220–229.
21. Blencowe N, Skilton A, Gaunt D, et al. Protocol for developing quality assurance measures to use in surgical trials: an example from the ROMIO study. *BMJ Open*. 2019;9:e026209.
22. Avella J. Delphi panels: research design, procedures, advantages, and challenges. *Int J Doctoral Stud*. 2016;11:305–321.