Novel Assessments of Technical and Non-Technical Cardiac Surgery Quality: Protocol for a Mixed Methods Study

Donald Likosky, Steven J. Yule, Michael R. Mathis, Roger D. Dias, Jason J. Corso, Min Zhang, Sarah L. Krein, Matthew D. Caldwell, Nathan Louis, Allison M. Janda, Nirav J. Shah, Francis D. Pagani, Korana Stakich-Alpirez, Milisa M. Manojlovich

Submitted to: JMIR Research Protocols on: July 22, 2020

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.
Table of Contents

Original Manuscript.................................................................................................................................................. 5
Supplementary Files.................................................................................................................................................. 24
Figures....................................................................................................................................................................... 25
Figure 1............................................................................................................................................................... 26
Figure 2............................................................................................................................................................... 27
Novel Assessments of Technical and Non-Technical Cardiac Surgery Quality: Protocol for a Mixed Methods Study

Donald Likosky1* PhD; Steven J. Yule2* PhD, FRCSE; Michael R. Mathis3* MD; Roger D. Dias4* MD, MBA, PhD; Jason J. Corso5* PhD; Min Zhang6* PhD; Sarah L. Krein7* PhD, RN; Matthew D. Caldwell1* MD; Nathan Louis5* BS; Allison M. Janda8* MD; Nirav J. Shah3* MD; Francis D. Pagani8* MD, PhD; Korana Stakich-Alpirez8* MS, CACS; Milisa M. Manojlovich9* PhD, RN

1Department of Cardiac Surgery, University of Michigan Ann Arbor US
2University of Edinburgh Edinburgh GB
3University of Michigan Department of Anesthesiology Ann Arbor US
4STRATUS Center for Medical Simulation | Brigham Health Instructor in Emergency Medicine Harvard Medical School Cambridge US
5University of Michigan Department of Electrical Engineering and Computer Science, School of Engineering Ann Arbor US
6University of Michigan Department of Biostatistics, School of Public Health Ann Arbor US
7University of Michigan Department of Internal Medicine Ann Arbor US
8University of Michigan Department of Cardiac Surgery Ann Arbor US
9University of Michigan School of Nursing Ann Arbor US
*these authors contributed equally

Corresponding Author:
Donald Likosky PhD
Department of Cardiac Surgery, University of Michigan
5346 CVC
Ann Arbor
US

Abstract

Background: Of the 150,000 patients annually undergoing coronary artery bypass grafting surgery, 35% develop complications that increase mortality 5-fold and expenditures by 50%. Differences in patient risk and operative approach explain only 2% of hospital variation in some complications. The intraoperative phase remains understudied as a source of variation, despite its complexity and amenability to improvement.

Objective: The objectives of this study are to: (i) investigate the relationship between peer assessments of intraoperative technical skills and non-technical practices with risk-adjusted complication rates and (ii) evaluate the feasibility of using computer-based metrics to automate the assessment of important intraoperative technical skills and non-technical practices.

Methods: This multi-center study will use video-recording, established peer assessment tools, electronic health record data, registry data and a high-dimensional computer vision approach to: (1) investigate the relationship between peer assessments of surgeon technical skills and variability in risk-adjusted patient adverse events; (2) investigate the relationship between peer assessments of intraoperative team-based non-technical practices and variability in risk-adjusted patient adverse events; (3) use quantitative and qualitative methods to explore the feasibility of using objective, data-driven computer-based assessments to automate the measurement of significant, intraoperative determinants of risk-adjusted patient adverse events.

Results: The project was funded by the National Heart, Lung and Blood Institute in 2019 and enrollment for the overall project is expected to begin in 2020.

Conclusions: We anticipate this project will substantially increase our ability to assess determinants of variation in complication rates by specifically studying a surgeon’s technical skills and operating room team member non-technical practices. These findings may provide effective targets for future trials or quality improvement initiatives to enhance the quality and safety of cardiac surgical patient care. Clinical Trial: -

(JMIR Preprints 22/07/2020:22536)
DOI: https://doi.org/10.2196/preprints.22536
Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ Please make my preprint PDF available to anyone at any time (recommended).
   Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
   Only make the preprint title and abstract visible.
   No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).
   Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.
   Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in the PubMed Now! service, my accepted manuscript PDF will automatically be made openly available.

https://preprints.jmir.org/preprint/22536
[unpublished, non-peer-reviewed preprint]
Novel Assessments of Technical and Non-Technical Cardiac Surgery Quality: Protocol for a Mixed Methods Study

Donald S. Likosky, Ph.D.*; Department of Cardiac Surgery, University of Michigan, Ann Arbor, MI, United States

Steven J. Yule, Ph.D.; Department of Clinical Surgery, University of Edinburgh, Edinburgh, Scotland; Department of Surgery, Brigham & Women’s Hospital/Harvard Medical School, Boston, MA, United States

Michael R. Mathis, M.D.; Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

Roger D. Dias, M.D., M.B.A., Ph.D.; Department of Emergency Medicine, Brigham & Women’s Hospital/ Harvard Medical School, Boston, MA, United States

Jason J. Corso, Ph.D.; Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, Ann Arbor, MI, United States

Min Zhang, Ph.D.; Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

Sarah L. Krein, Ph.D., R.N.; Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

Matthew D. Caldwell, M.D.; Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

Nathan Louis, B.S.; Department of Electrical Engineering and Computer Science, College of Engineering, University of Michigan, Ann Arbor, MI, United States

Allison M. Janda, M.D.; Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

Nirav J. Shah, M.D.; Department of Anesthesiology, University of Michigan, Ann Arbor, MI, United States

Francis D. Pagani, M.D., Ph.D.; Department of Cardiac Surgery, University of Michigan, Ann Arbor, MI, United States

Korana Stakich-Alpirez, MS, CCRP; Department of Cardiac Surgery, University of Michigan, Ann Arbor, MI, United States

https://preprints.jmir.org/preprint/22536 [unpublished, non-peer-reviewed preprint]
Milisa M. Manojlovich, Ph.D., R.N.; School of Nursing, University of Michigan, Ann Arbor, MI, United States

Address correspondence to:
Dr. Donald Likosky
Section of Health Services Research and Quality
Department of Cardiac Surgery
5346 CVC
Michigan Medicine
Ann Arbor, MI 48109-5864
phone: 734-232-4216, fax: 734-764-2255, email: likosky@umich.edu
Novel Assessments of Technical and Non-Technical Cardiac Surgery Quality: Protocol for a Mixed Methods Study

Abstract

Background: Of the 150,000 patients annually undergoing coronary artery bypass grafting surgery, 35% develop complications that increase mortality 5-fold and expenditures by 50%. Differences in patient risk and operative approach explain only 2% of hospital variation in some complications. The intraoperative phase remains understudied as a source of variation, despite its complexity and amenability to improvement.

Objectives: The objectives of this study are to: (i) investigate the relationship between peer assessments of intraoperative technical skills and non-technical practices with risk-adjusted complication rates and (ii) evaluate the feasibility of using computer-based metrics to automate the assessment of important intraoperative technical skills and non-technical practices.

Methods: This multi-center study will use video-recording, established peer assessment tools, electronic health record data, registry data and a high-dimensional computer vision approach to: (1) investigate the relationship between peer assessments of surgeon technical skills and variability in risk-adjusted patient adverse events; (2) investigate the relationship between peer assessments of intraoperative team-based non-technical practices and variability in risk-adjusted patient adverse events; (3) explore the feasibility of using objective, data-driven computer-based assessments to automate the measurement of significant, intraoperative determinants of risk-adjusted patient adverse events.

Results: The project was funded by the National Heart, Lung and Blood Institute in 2019 and enrollment for the overall project is expected to begin in 2020.

Conclusions: We anticipate this project will substantially increase our ability to assess determinants of variation in complication rates by specifically studying a surgeon's technical skills and operating room team member non-technical practices. These findings may provide effective targets for future trials or quality improvement initiatives to enhance the quality and safety of cardiac surgical patient care.
Keywords:
Introduction
The Epidemiology of Cardiac Surgery
Nearly 150,000 coronary artery bypass grafting (CABG) procedures are performed annually in the U.S., a procedure associated with a high rate of major adverse events (35% of patients) that vary by hospital [1]. These adverse events increase a patient’s risk of mortality 4.7-fold, increase extended (greater than 14 days) index admission length of stay 7-fold, and are associated with more than $50K in additional healthcare expenditures per patient [1–4]. While understudied, intraoperative performance (including the surgeon’s technical skills and team-based nontechnical practices) is a significant potentially modifiable determinant of operative adverse events, Figure 1.

The Role of Technical Skills on Surgical Outcomes
Prior research has evaluated the association between technical skills as defined as “psychomotor action or related mental faculty acquired through practice and learning pertaining to a particular craft or profession [5]”) and operative outcomes [6]. While taxonomies exist to objectively and reliably assess a surgeon’s technical skills, they are often applied within simulated, structured scenarios that may not mimic real-world patient care (Table 1). In one exception, investigators applied the Objective Structured Assessment of Technical Skill (OSATS [7]) to real operative settings. In this instance, 10 clinician experts rated a single 25 to 50-minute video segment of a laparoscopic operation from 20 surgeons [8]. Assessments, linked to data from the last 2 years of each surgeon’s experience, were significantly, inversely associated with the surgeon’s adverse event and mortality outcomes. In another study, surgical skill was associated with outcomes from cancer surgery [9].
The Role of Non-Technical Practices on Surgical Outcomes

Non-technical practices (“the cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance [10]”) are both individual and team-based. While improvement in these practices has been associated with decreases in operative mortality [11], investigations thus far have focused on developing robust, validated taxonomies of behavior with corresponding assessment tools customized to the individual team members’ intraoperative care role. Dominant taxonomies [12–14] include: Non-Technical Skills for Surgeons (NOTSS), Anesthetists’ Non-Technical Skills (ANTS), Perfusionists Intraoperative Non-Technical Skills (PINTS) and Scrub Practitioners’ List of Intraoperative Non-Technical Skills (SPLINTS). These taxonomies enable assessments of four important categories of non-technical practices: situation awareness, decision making, communication & teamwork, and leadership/task management. Situation awareness [15] is the process of developing and maintaining a dynamic awareness of the operative situation based on gathering and interpreting data from the operative environment. This domain is essential for effective decision-making [16], skills for diagnosing a given situation to inform a judgment about appropriate actions. Successful surgery also depends on social skills allowing multiple individuals with task interdependencies and shared goals to communicate and work effectively as a team [17]. Dysfunctional team dynamics, ineffective communication, and ambiguous leadership [18] account for a substantial proportion of operative adverse events [19].

A surgeon’s non-technical practices, manifesting as diagnostic failure [20] or a breakdown in teamwork and information sharing [21] may contribute to higher risk of a major adverse event or death. The largest operative study of NOTSS conducted thus far involved 715 surgical procedures and 11,440 assessments [22]. Surgeons’ non-technical skills were rated as good (score = 4) in 18.8% of responses, acceptable (score = 3) in 49.1%, marginal (score = 2) in 21.9% and poor (score = 1) in 0.9%. In a video-based study including 82 cardiac surgeons, there was a 129% increased odds (after adjusting for technical skills) of higher patient safety scores with every 1-point increase in NOTSS score [23].

Rationale for the Study

Our study, leveraging the infrastructure and track record of two established physician-led quality collaboratives integrated with a cutting-edge scalable video understanding platform, seeks to advance our understanding of how operative skills and non-technical practices impact outcomes. Evaluation of technical skills and non-technical practices has important implications for the 150,000 patients annually undergoing CABG. Patients undergoing CABG are at risk of harm due in part to the: (i) reconstruction of anatomical structures under high magnification, (ii) multiple high-risk phase transitions of care between team members (e.g., anesthesiologist, perfusionist), and (iii) need for communication and teamwork (e.g., instrument handoffs) across many team members. Our approach aimed at identifying key, modifiable intraoperative determinants of major adverse events may likely be applied to approximately 200,000 additional cardiac surgical procedures (e.g., transcatheter aortic valve replacement) or other high-risk non-cardiac surgical specialties (e.g., thoracic surgery, intra-abdominal surgery, neurosurgery, orthopedics, head and neck reconstructive surgery).
Methods

Study Population

Our population will include adult patients undergoing scheduled CABG operations using cardiopulmonary bypass performed by attending surgeons at 6 hospitals participating in the Multicenter Perioperative Outcomes Group (MPOG) Collaborative, a national physician-led collaborative of academic and community hospitals.

Digital Recording

We will digitally record 506 CABG operations at 6 hospitals.

Per protocol, the Study Coordinator will use a randomization protocol from the Data Coordinating Center (DCC) to select, by week, different cardiac surgical operating rooms for video recording. Prior to the recording, the Coordinator will synchronize the cameras with other operating room data sources (e.g., intraoperative record as submitted to MPOG). As part of existing MPOG workflows and data mapping, key transitions in phases of the patient’s care (e.g., patient room entry, surgical incision, onset of bypass) are routinely documented within the intraoperative electronic health record of participating MPOG hospitals; these data are validated and mapped to universal MPOG concepts, semantically interoperable across MPOG sites with varying source vendor software[24]. Digital recordings will be segmented based upon key transitions in phases of care; recordings of the operation will begin when the patient enters the operating room (and therefore prior to the pre-induction verification) and end when the patient exits the operating room.

Per protocol, the DCC will conduct audio and video quality checks across hospitals and recordings. Monthly video calls with the Study Coordinators and the DCC will minimize protocol deviations, anticipated to be 10% of cases. Initially, investigators at the DCC will review the entire recording to fine-tune the MPOG event timestamping to the exact second. Given the input operative data, a Hidden Markov Model [25,26] or Deep Learning based approach [27] will divide the procedure into temporal segments and associate them with the procedural step labels from the operative script. Standardized segments for assessment will only contain critical operative portions (Table 2 for illustrative Phases). Edited segments will be transferred to and assigned to assessors via a HIPAA-compliant web-based assessment platform. We will resubmit twenty percent of segments to the same (test-retest reliability) or other assessors (inter-rater reliability) using alpha = 0.70 for concordance [28].

Hospital Performance Feedback

The DCC will provide monthly reports to hospitals, including: number of digitally recorded operations, quality of transmitted digital recordings, and adherence to study operational protocols.

Peer Assessments

Assessors, blinded to the hospital and operative team, will provide technical (Aim 1) and non-
technical (Aim 2) assessments of cardiac surgical operations. Each operation will receive at least 12 assessments (three for each: surgeon, anesthesiologist, perfusionist, scrub nurse), with 20% of assessments re-reviewed. Surgeons will assess a surgeon’s technical skills (modified OSATS + cardiac surgery-specific skills) via a validated 5-point behaviorally anchored scale. Additionally, surgeons will evaluate non-technical practices using NOTSS [12], while anesthesiologists will use ANTS [13], scrub nurses use SPLINTS [14] and perfusionists use PINTS. All non-technical taxonomies leverage a four-point ordinal scale. Segments will capture each operation’s critical phases. Technical skills assessors will be given one operative field camera angle [Cam Surgical Field (SF)] for their assessment (Figure 2). Given the inter-dependence of intraoperative team members, non-technical assessors will be given alternative camera angles depicting the intraoperative team [Cam Operative Team (OT) #1 and #2] (Figure 2). Assessors will receive an Amazon coupon for completed reviews.

**e-Learning Training Module**
Assessor training will be conducted using a web-based assessor training package that will include: (i) foundational knowledge of the relevant tool and (ii) video examples of correct identification, categorization, and assessment. In the final assessment, assessors will have to reach 0.70 consistency with gold standard (investigative team) assessments to contribute to the study. A median 70% concordance with reference assessments (established by the DCC) will serve as a basis for eligibility of assessors to continue conducting real assessments [29].

**Linkage of Peer Assessments with Complications**
We will associate peer assessments from recorded operations with each surgeon’s adverse event rate for the previous 2 years using each hospital’s The Society of Thoracic Surgeons (STS) data. Next, we will evaluate the feasibility of using a video understanding platform to automate the identification and tracking of important skills and practices (identified in Aims 1 and 2).

**Aim 1: Investigate the relationship between peer-rater assessments of a surgeon’s technical skills and variability in risk-adjusted patient adverse events**

**Approach**
We will conduct peer reviewer assessments of digitally recorded CABG operations at 6 MPOG hospitals to identify technical skills significantly associated with major operative adverse events. We will digitally record surgeries from a total of 36 surgeons (6 surgeons at each of 6 centers). By recording multiple surgeries for each surgeon and conducting multiple reviews for each recorded surgery, we are able to better assess a surgeon’s skills as it considers the variability across surgeries within a surgeon, variability between surgeons within and across centers, and the variability due to assessors.

Each surgical operation will be divided (using our video segmentation protocol) into pre-specified phases containing the most critical operative portions. The DCC will distribute video segments for surgeon assessment via our HIPAA-compliant assessment platform that will provide assessors with a view of the operative field (from Cam SF, Figure 2). Reminder notices will be sent to the assessors until we receive at least 3 assessments per video. Surgeon assessors will provide domain-specific and an overall summary judgement (using a modified-OSATS taxonomy). Twenty percent of segments...
will be resubmitted for review to test assessor reliability. We will associate the average assessments with the adjusted risk of major morbidity and mortality over the prior two years for each surgeon.

**Measures**

Our primary exposure will be the average summary assessment of each surgeon’s technical skill. The primary outcome will be a surgeon’s STS’ composite major morbidity or mortality (i.e., permanent stroke, surgical re-exploration, deep sternal wound infection, renal failure, prolonged ventilation, or operative mortality) rate. We will use clinical data from each center to adjust for covariates incorporated within the STS’ risk prediction models [30,31].

**Analytical Plan**

For analyses of technical skills, we will use linear mixed effect models to model assessments of surgical procedures where assessors and surgeons are included as random effects. Based on the linear mixed effect models’ fit, we will quantify variation in peer-assessor assessments of a surgeon’s technical skills and use the intraclass correlation coefficient to measure inter-assessor reliability. We will use predictions of each surgeon’s technical skill (overall and each individual domain and element) from the linear mixed effect models as summary measures of a surgeon’s technical skills in subsequent analyses. Generalized linear mixed effect models with a logit link will then be used to associate a surgeon’s technical skills with our composite outcome. We model surgeons and hospitals as random effects, accounting for the nesting structure of the data (i.e., patients nested within surgeons and hospitals). We will adjust for patient (e.g., demographic and clinical factors), and surgeon (e.g., number of years since surgical fellowship training), factors by including them as fixed effects in the models. The factors of interest are summary measures of a surgeon’s technical skills, which are included as surgeon-level explanatory variables. We will consider the overall assessments of a surgeon’s technical skill, averaged across three assessors and each domain individually.

**Power Analysis**

We used simulations to evaluate statistical power for a two-sided test, alpha = 0.05. Our analysis will be based on: (i) outcomes for approximately 7200 operations over 2 years from 36 surgeons (~100 operations per surgeon) at 6 hospitals. We estimate having approximately 98% power in detecting an odds ratio of 0.85 per one unit (standardized) increase in a surgeon’s technical skill on the rate of adverse events.

**Aim 2**: Investigate the relationship between peer rater assessments of intraoperative team-based non-technical practices and variability in risk-adjusted patient adverse events.

**Approach**

We will leverage each hospital’s intraoperative electronic health record system for video segmentation. For example, electronic documentation workflow will capture the timing surrounding the preparation and weaning from cardiopulmonary bypass using pre-computed, validated, publicly available MPOG phenotypes “cardiopulmonary bypass start” and “cardiopulmonary bypass end”[24,32]. Segments will be reviewed by at least 12 assessors (3 surgeons will rate surgeons using NOTSS, 3 anesthesiologists will rate anesthesiologists using ANTS, 3 perfusionists will rate perfusionists using PINTS, and three scrub nurses using SPLINTS. We will assess the association between peer assessments of non-technical practices and surgeon measures of postoperative major morbidity and mortality, adjusted for patient risk factors and surgeon technical skills.
Measures
Our primary exposure will be the average summary peer assessment of each provider’s non-technical practices. Similar to Aim 1, the primary outcome will be the rate of major morbidity or mortality, adjusting for covariates contained within STS’ published risk models [30,31].

Analytical Plan
We will use generalized linear mixed effect models with a logit link to associate peer-assessor assessments of non-technical practices of the surgeon with the surgeon’s STS composite score for major morbidity and mortality. Models will be similar to those described in Aim 1, although will include average summary measures of surgeon’s non-technical skill as surgeon-level explanatory variables and hospital-level average summary measures of anesthesiologists, perfusionists, and scrub nurses. Both overall summary measures and individual scale domains will be considered. We will focus primarily on assessing the effects of non-technical practices on morbidity and mortality rates, while adjusting for patient level risk factors and a surgeon’s technical skills. We will explore the influence of non-technical practices on the relationship between a surgeon’s technical skills and our composite endpoint by including non-technical practices as an interaction term in models with technical skills.

Power Analysis
The power analysis is based on approximately 7,200 cases across 36 surgeons from 6 hospitals. As surgeon’s non-technical practices are considered a surgeon-level variable, there will be sufficient power in detecting the same effect sizes as reported in the Aim 1. For example, there will be ~98% power in detecting an odds ratio of 0.85 for every one unit (standardized by standard deviation) increase in a surgeon’s non-technical practices. There will be an approximately 85% power in detecting an odds ratio of 0.88 on major adverse events per one unit increase in non-technical practices that are summarized as hospital-level variables.

Aim 3: Explore the feasibility of using objective, data-driven computer-based assessments to automate the identification and tracking of significant, intraoperative determinants of risk-adjusted patient adverse events

High-dimensional computer-based assessments of digital recordings are used to recognize and track human activity (computer vision). Computer vision focuses on training computers to derive meaning from visual imagery. Video understanding, a specialty within computer vision, focuses on identifying and tracking objects over time from video and developing mathematical models to train computers to extract the meaning within these moving images. This field may offer unparalleled capabilities for conducting objective peer assessments by automatically identifying and tracking human activity, in non-simulated environments, comparable to that of expert human assessors.

Surgical Technical Skills
Video understanding may address some of the limitations in traditional mentored or simulation-based approaches for assessing a surgeon’s technical skill, including human assessor bias and limited scalability. Prior investigations have documented the reliability of video-based surgical motion analyses for assessing laparoscopic performance in the operating room as compared to the traditional time intensive, human assessor approach [33,34]. Wazari and colleagues compared expert surgeon’s rating assessments to computer-based assessments of technical skills (e.g., suturing, knot tying).
including fluidity of motion, tissue handling, and motion economy [35]. Computer-based assessments had less variance relative to expert assessors. Additionally, Corso and colleagues evaluated the feasibility of computer-based methods for technical skill assessment (e.g., suturing, knot tying) by 10 surgeons of varying experience with robotic-assisted surgery [36]. This evaluation included acquiring 99 unique videos with 22,467 total frames and the development of a state-of-the-art deep learning-based surgical tool tracking system. The quantitative assessment against gold standard (human annotated) tool-tracks found a 90.7% mean average precision over all test videos across all surgeon skill levels.

Non-Technical Practices
Non-technical practice assessments have predominantly occurred within simulated environments and relied on trained human observers, thereby limiting fidelity and ease of real-world deployment [37,38]. While potentially feasible, investigators have not evaluated whether video understanding could provide an objective alternative for high fidelity assessments of non-technical practices in real-world operative environments, generalizable across hospitals with varied operating room layouts and camera configurations. Video understanding may be used to assess features aligned with non-technical practices without relying on verbal communication (e.g., recognizing and tracking team members during critical phases of the operation as a proxy for communication and teamwork; transferring of operative instruments between team members as a proxy for decision making) [39]. Video understanding requires time-limited human observer involvement to provide labels for training algorithms after which the automated system may be deployed at scale.

Approach
We will explore the feasibility of using a video understanding platform to identify important features associated with assessor ratings of technical skills and non-technical practices in recorded operations. To support developing the video understanding platform, we will conduct interviews and site visits at a subset of low and high performing hospitals to enhance understanding of a hospital’s contextual characteristics (e.g., culture) and important ‘usual practices.’

Video Understanding
The video understanding approach will focus on two specific techniques (i.e., visual detection, visual tracking), which will be applied to identify and measure surgeons’ technical skills (Aim 1) and team based non-technical practices (Aim 2). We will apply ambiguity reduction across the three time-synchronized video recordings to harmonize (rather than duplicate) elements within and across video angles. We will use proven methods for video understanding (e.g., boosting [40] and deep learning [41]). We will use boosting for cases of limited data and deep learning for cases with ample data. We will learn detection models to ascertain kinematic features potentially associated with surgical technical skills (e.g., path length of the surgeon’s suturing and non-suturing hands) and non-technical practices (e.g., identifying and tracking gaze direction of team members at critical times of the surgical procedure) based on Aims 1 and 2. We will learn these features using mutually exclusive datasets containing video segments: (i) a training dataset (used for training the video understanding algorithms); (ii) a computer vision validation dataset (used to mitigate risk of over-fitting, for example, of the video understanding algorithms); (iii) a computer vision testing dataset (used for computing the error statistics of the computer vision system to meet human feature annotation); and (iv) a study set (video segments for peer assessments). Investigators will observe the raw video from the training dataset to provide bounding-box annotations for each feature, within contextual feedback provided by members of the investigative team who work in the operating room. A certain detection model is initialized with a random set of parameters and then the training algorithm iteratively
refines them based on the model’s empirical performance (ability to automatically detect the phenomena bounding boxes) based on the annotations in the training data. The validation set is used during this training process to protect against over-training and bias. Some technical assessments will require detection in a video frame and tracking of the detected object throughout the video frames (“visual tracking.”) For example, to measure the surgeon’s economy of motion, we will detect the surgeon’s hands at frame t, track the surgeon’s hands at all future frames t+k, and then compute a trajectory of the centroid of the detected bounding boxes. We will use both classical physics-based tracking models (e.g., Lucas-Kanade tracking [42]) and modern deep-learning based methods [43]. We will compute a range of validated kinematic features [35] and quantifiers of economy of motion (e.g., path length of the surgeon’s suturing and non-suturing hands, variance of local change in the trajectory against a linear or smoothed trajectory).

Qualitative Interviews

Concurrent with developing and testing the video understanding platform, we will randomly select up to 4 of the 6 hospitals (equal representation of low and high outlier hospitals) participating in Aim 2 for more detailed investigation. We will conduct semi-structured interviews with interdisciplinary cardiac surgery operating room team members. To enhance our understanding of technical and non-technical operating room practices, we will collect data (through interviews with intraoperative team members) concurrent with conducting analyses. Using our conceptual model and insights from Aims 1 and 2, we will develop a semi-structured interview guide to encourage new and/or unexpected ideas or concepts to surface. For each interview, the interviewer will play back video segments from an operation involving the interviewee and ask the interviewee to describe his/her role within that operative phase. The interviewer will ask questions seeking to better understand team member roles and influences on technical skills and non-technical practices. We expect the guide, which will be modified as new findings or themes emerge, will consist of 7 to 9 open-ended questions, with probes. Interviewers will participate in a 3-day training program at the University of Michigan Health Communications Laboratory. Interviews will continue until reaching informational redundancy “saturation” (often occurring at 10 to 12 interviews or when sources have been exhausted) at each hospital. We will: (i) conduct 40 to 60-minute interviews in private rooms, (ii) digitally record and transcribe transcripts verbatim, (iii) compare 10% of transcripts (and correct as needed) against the recordings, and (iv) provide interviewees with a gift certificate. Our thematic analysis will include both within and across case comparisons. We expect that: (i) in reviewing the videos, providers will complement peer assessments regarding how and why contextual factors influence performance (technical and non-technical) and (ii) interviewees will validate the video content to maximize our video understanding algorithm’s fidelity. Thus, our interview findings will improve our interpretation of the video content to iteratively inform and enhance our video understanding platform training.

Measures

Our primary outcome will be the features derived from a video understanding platform, which will be compared to a gold standard human identifying the same features. Features, as economy of motion, are derived from the raw output of the video understanding platform, which naturally performs visual detection and visual tracking. The gold standard uses the analogous “raw output” from humans and the same method for the computation of the derived feature.

Analytical Plan

We will assess our video understanding’s ability to correctly identify and track features within our testing dataset. Using the raw video in the testing set, we will provide the necessary bounding-box
annotations for each feature, which will be compared to the automatically generated features from the video understanding system using standard metrics (e.g., Intersection over Union [44] and DICE coefficient [45]). For example, when we compute the economy of motion of the surgeon’s hand, we will provide bounding-box annotations of the surgeon’s hand. The video understanding system will use these annotations to learn a mathematical visual detection model capable of producing the detections of the hand on novel video. Then, the economy of motion feature will be derived on the output bounding boxes. We plan a 2-phased analysis. First, we will measure agreement and associate each feature with each component of the technical and non-technical assessments (specific to each operative phase) using Pearson/Spearman correlation coefficients or Kendall’s Tau, depending on data distribution. Second, we will identify the best combination of video understanding features that are most closely associated with technical and non-technical score domains (specific to each operative phase). We will use regression (e.g., linear, ridge, deep) to model each domain and technical and non-technical summary scores as dependent variables, including features from the video understanding platform as independent variables. We will: (i) select features using variable selection and (ii) quantify the magnitude of information in peer assessment that can be identified by the computer using (generalized) R squares. We expect our video understanding platform will perform similarly to human assessor in terms of identifying important surgical technical skills and team-based non-technical practices.

Results
Results of Aims 1 and 2 will likely yield assessments that identify a wide range of variation in both surgeons’ technical skills and non-technical practices as has already been documented in the literature. Where our study will make a significant contribution is in associating these assessments with adverse event rates. The novel contribution of Aim 3 will be to associate computer-based assessments with adverse event rates, as a more objective and reliable replacement for human peer assessors, moving us closer to our overall goal of improving outcomes for cardiac surgery patients. We will use study results to develop data-driven technical skill and non-technical practice coaching interventions across a subset of hospitals.

Discussion
Strengths
There is increasing demand from the public and payers to improve healthcare value (quality divided by expenditures). Despite wide variability in cardiac surgical quality and robust clinical data from STS for risk adjustment and outcomes ascertainment, only 2% of hospital variability in some outcomes are explainable by currently recorded data elements [46]. Analysis of operative videos may reveal unique opportunities for advancing operative quality improvement beyond that provided through traditional data sources [47].

Our proposed study, leveraging the infrastructure and track record of 2 established physician-led quality collaboratives, integrated with a cutting-edge scalable video understanding platform, will advance our understanding of how surgical skills and non-technical practices impact outcomes. Our approach aimed at identifying key, modifiable intraoperative determinants of major adverse events may likely be applied to approximately 200,000 additional cardiac surgical procedures involving valve repair or replacement, aortic procedures and percutaneous cardiac procedures (e.g., transcatheter aortic valve replacement) or other high-risk non-cardiac surgical specialties (e.g., neurosurgery, orthopedics, head and neck reconstructive surgery).
Limitations

Although unlikely, there are a few potential challenges with this study.

Aims 1 and 2

There is a remote possibility that we do not find that investigated technical skills are associated with adverse events. If needed, we will expand our review of surgical operations to include: (i) hospitals with lower operative volume, (ii) longer segments for peer rating, (iii) an expanded list of operative phases that might distinguish between high and low performing surgeons, (iv) high risk or technically challenging operations (e.g., procedures involving circulatory arrest). We will consider expanding to other hospitals if: (i) hospital variability in adjusted adverse events is less than anticipated or (ii) we are unable to amass sufficient digital recordings from our initial 6 hospitals.

There is a remote possibility that we will not be able to recruit sufficient numbers of assessors, that there is attrition at sites before we complete data collection or that assessor reliability will not be satisfactory. If needed, we will (i) expand our sampling pool to include providers who have expressed desire to partner on this project but were not selected initially, (ii) provide monthly feedback and engagement support to participating assessors, and (iii) provide expanded assessor training and calibration.

Aim 3

Our video understanding may not be completely automated. Alternatively, we will consider a semi-automated platform that relies, for instance, on a human periodically manually annotating the relevant features in the video at a certain segment and then allowing the video understanding platform to interpolate those annotations. This strategy would reduce overall human annotation time but maintain the reproducibility and accuracy of the automated system.
Acknowledgements

The authors thank Holly Neilson for her editorial review of this manuscript.

Donald S. Likosky and Francis D. Pagani receive extramural support from the Agency for Healthcare Research and Quality (AHRQ: R01HS026003) and the National Heart, Lung, and Blood Institute (NHLBI: R01HL146619). Francis D. Pagani is a member of the scientific advisory board of FineHeart, Inc., member of the Data Safety Monitoring Board for Carmat, Inc., member of the Data Safety Monitoring Board for the NHLBI PumpKIN clinical trial, and Chair of The Society of Thoracic Surgeons, Intermacs Task Force. Sarah L. Krein is supported by a Department of Veterans Affairs HSR&D research career scientist award. Michael R. Mathis receives extramural support from the NHLBI (K01HL14170103). Allison M. Janda receives extramural support from the NIH through a T32 Research Fellowship (5T32GM103730-07). Opinions expressed in this manuscript do not represent those of The NIH or The AHRQ or the US Department of Health and Human Services or the US Department of Veterans Affairs.

Conflicts of Interest

None

Abbreviations

ANTS: Anesthetists’ Non-Technical Skills
CABG: Coronary Artery Bypass Grafting
DCC: Data Coordinating Center
MPOG: Multicenter Perioperative Outcomes Group
MSTCVS-QC: Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative
NOTSS: Non-Technical Skills for Surgeons
OSATS: Objective Structured Assessment of Technical Skill
OT: Operative Team
PI: Principal Investigator
PINTS: Perfusionists Intraoperative Non-Technical Skills
SF: Surgical Field
SPLINTS: Scrub Practitioners’ List of Intraoperative Non-Technical Skills
STS: Society of Thoracic Surgeons

Figures

Figure 1. Conceptual Model
Figure 2. Proposed Intraoperative Recording Configuration

Tables

Table 1. Application of Objective Structured Assessment of Technical Skills (OSATS) to Cardiac Surgery
Table 2. Illustrative Operative Phases for Video Assessments
References

1. Data Analyses of The Society of Thoracic Surgeons National Adult Cardiac Surgery Database. Society of Thoracic Surgeons; 2018.

2. Likosky DS, Wallace AS, Prager RL, Jacobs JP, Zhang M, Harrington SD, Saha-Chaudhuri P, Theurer PF, Fishstrom A, Dokholyan RS, Shahian DM, Rankin JS, Michigan Society of Thoracic and Cardiovascular Surgeons Quality Collaborative. Sources of Variation in Hospital-Level Infection Rates After Coronary Artery Bypass Grafting: An Analysis of The Society of Thoracic Surgeons Adult Heart Surgery Database. Ann Thorac Surg 2015 Nov;100(5):1570–5; discussion 1575–6. PMID:26321440

3. LaPar DJ, Crosby IK, Rich JB, Fonner E Jr, Kron IL, Ailawadi G, Speir AM, Investigators for Virginia Cardiac Surgery Quality, Initiative. A contemporary cost analysis of postoperative morbidity after coronary artery bypass grafting with and without concomitant aortic valve replacement to improve patient quality and cost-effective care. Ann Thorac Surg 2013 Nov;96(5):1621–1627. PMID:23972932

4. Fowler VG Jr, O’Brien SM, Muhlbaier LH, Corey GR, Ferguson TB, Peterson ED. Clinical predictors of major infections after cardiac surgery. Circulation 2005;112(9 Suppl):I358–65. PMID:16159846

5. Shorter Oxford English Dictionary on Historical Principles: N-Z. Oxford University Press; 2002. ISBN:9780198604570

6. Hance J, Aggarwal R, Stanbridge R, Blauth C, Munz Y, Darzi A, Pepper J. Objective assessment of technical skills in cardiac surgery. Eur J Cardiothorac Surg 2005 Jul;28(1):157–162. PMID:15982599

7. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997 Feb;84(2):273–278. PMID:9052454

8. Birkmeyer JD, Finks JF, O’Reilly A, Oerline M, Carlin AM, Nunn AR, Dimick J, Banerjee M, Birkmeyer NJO, Michigan Bariatric Surgery Collaborative. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013 Oct 10;369(15):1434–1442. PMID:24106936

9. Curtis NJ, Foster JD, Miskovic D, Brown CSB, Hewett PJ, Abbott S, Hanna GB, Stevenson ARL, Francis NK. Association of Surgical Skill Assessment With Clinical Outcomes in Cancer Surgery. JAMA Surg [Internet] 2020 May 6; PMID:32374371

10. Flin R, O’Connor P. Safety at the sharp end: a guide to non-technical skills. CRC Press; 2017.

11. Neily J, Mills PD, Young-Xu Y, Carney BT, West P, Berger DH, Mazzia LM, Paull DE, Bagian JP. Association between implementation of a medical team training program and surgical mortality. JAMA 2010 Oct 20;304(15):1693–1700. PMID:20959579

12. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons’ non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. World J Surg 2008 Apr;32(4):548–556. PMID:18259809

13. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists’ Non-Technical Skills (ANTS): evaluation of a behavioural marker system. Br J Anaesth 2003 May;90(5):580–588. PMID:12697584

14. Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G. Evaluation of the Scrub Practitioners’ List of Intraoperative Non-Technical Skills system. Int J Nurs Stud 2012 Feb;49(2):201–211. PMID:21974792

15. Schulz CM, Endsley MR, Kochs EF, Gelb AW, Wagner KJ.
16. Flin R, Youngson G, Yule S. How do surgeons make intraoperative decisions? Qual Saf Health Care 2007 Jun;16(3):235–239. PMID:17545353
17. Baker DP, Day R, Salas E. Teamwork as an essential component of high-reliability organizations. Health Serv Res 2006 Aug;41(4 Pt 2):1576–1598. PMID:16898980
18. Hu Y-Y, Parker SH, Lipsitz SR, Arriaga AF, Peyre SE, Corso KA, Roth EM, Yule SJ, Greenberg CC. Surgeons’ Leadership Styles and Team Behavior in the Operating Room. J Am Coll Surg 2016 Jan;222(1):41–51. PMID:26481409
19. Lagoo J, Berry WR, Miller K, Neal BJ, Sato L, Lillemoe KD, Doherty GM, Kasser JR, Chaikof EL, Gawande AA, Haynes AB. Multisource Evaluation of Surgeon Behavior is Associated with Malpractice Claims. Ann Surg [Internet] 2018 Mar 23; PMID:29578910
20. Way LW, Stewart L, Gantert W, Liu K, Lee CM, Whang K, Hunter JG. Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. Ann Surg 2003 Apr;237(4):460–469. PMID:12677139
21. Mazzocco K, Petitti DB, Fong KT, Bonacum D, Brookey J, Graham S, Lasky RE, Sexton JB, Thomas EJ. Surgical team behaviors and patient outcomes. Am J Surg 2009 May;197(5):678–685. PMID:18789425
22. Crossley J, Marriott J, Purdie H, Beard JD. Prospective observational study to evaluate NOTSS (Non-Technical Skills for Surgeons) for assessing trainees’ non-technical performance in the operating theatre. Br J Surg 2011 Jul;98(7):1010–1020. PMID:21480195
23. Yule S, Gupta A, Gazarian D, Geraghty A, Smink DS, Beard J, Sundt T, Youngson G, McIlhenny C, Paterson-Brown S. Construct and criterion validity testing of the Non-Technical Skills for Surgeons (NOTSS) behaviour assessment tool using videos of simulated operations. Br J Surg 2018 May;105(6):719–727. PMID:29601087
24. Colquhoun DA, Shanks AM, Kapeles SR, Shah N, Saager L, Vaughn MT, Buehler K, Burns ML, Tremper KK, Freundlich RE, Aziz M, Kheterpal S, Mathis MR. Considerations for Integration of Perioperative Electronic Health Records Across Institutions for Research and Quality Improvement: The Approach Taken by the Multicenter Perioperative Outcomes Group. Anesth Analg 2020 May;130(5):1133–1146. PMID:32287121
25. Kuehne H, Richard A, Gall J. Weakly supervised learning of actions from transcripts. Comput Vis Image Underst 2017 Oct 1;163:78–89.
26. Fellow R, IEEE, R. L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Readings in Speech Recognition San Francisco: Morgan Kaufmann; 1990. p. 267–296.
27. Zhou L, Xu C, Corso JJ. Towards automatic learning of procedures from web instructional videos. AAAI Conference on Artificial Intelligence [Internet] 2018. Available from: http://youcook2.eecs.umich.edu/static/YouCookII/youcookii_readme.pdf
28. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol 2012;8(1):23–34. PMID:22833776
29. Hull L, Arora S, Symons NRA, Jalil R, Darzi A, Vincent C, Sevdalis N, Delphi Expert Consensus Panel. Training faculty in nontechnical skill assessment: national guidelines on program requirements. Ann Surg 2013 Aug;258(2):370–375. PMID:23222032
30. Shahian DM, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC Jr, Lobdell KW, Vassileva C, Wyler von Ballmoos MC, Thourani VH, Rankin JS, Edgerton JR, D’Agostino RS, Desai ND, Feng L, He X, O’Brien SM. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development.

https://preprints.jmir.org/preprint/22536 [unpublished, non-peer-reviewed preprint]
Ann Thorac Surg 2018 May;105(5):1411–1418. PMID:29577925

31. O’Brien SM, Feng L, He X, Xian Y, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC Jr, Lodibell KW, Vassileva C, Wyler von Ballmoos MC, Thourani VH, Rankin JS, Edgerton JR, D’Agostino RS, Desai ND, Edwards FH, Shahian DM. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. Ann Thorac Surg 2018 May;105(5):1419–1428. PMID:29577924

32. MPOG Phenotype Browser [Internet]. Available from: http://phenotypes.mpog.org/

33. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, Bello F, Darzi A. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 2007 Jun;245(6):992–999. PMID:17522527

34. Dosis A, Aggarwal R, Bello F, Moorthy K, Munz Y, Gillies D, Darzi A. Synchronized video and motion analysis for the assessment of procedures in the operating theater. Arch Surg 2005 Mar;140(3):293–299. PMID:15781796

35. Azari DP, Frasier LL, Quamme SRP, Greenberg CC, Pugh CM, Greenberg JA, Radwin RG. Modeling Surgical Technical Skill Using Expert Assessment for Automated Computer Rating. Ann Surg [Internet] 2017; Available from: http://europepmc.org/abstract/med/28885509

36. Sarikaya D, Corso JJ, Guru KA. Detection and Localization of Robotic Tools in Robot-Assisted Surgery Videos Using Deep Neural Networks for Region Proposal and Detection. IEEE Trans Med Imaging 2017 Jul;36(7):1542–1549. PMID:28186883

37. Bierer J, Menu E, Leeper R, Fortin D, Fréchette E, Inculet R, Malthaner R. The Development of an in-situ Thoracic Surgery Crisis Simulation focused on Non-Technical Skill Training. Ann Thorac Surg [Internet] Elsevier; 2018 Feb 27; [doi: 10.1016/j.athoracsur.2018.01.058]

38. Dedy NJ, Bonrath EM, Ahmed N, Grantcharov TP. Structured Training to Improve Nontechnical Performance of Junior Surgical Residents in the Operating Room: A Randomized Controlled Trial. Ann Surg 2016 Jan;263(1):43.

39. Atkins MS, Tien G, Khan RSA, Meneghetti A, Zheng B. What do surgeons see: capturing and synchronizing eye gaze for surgery applications. Surg Innov 2013 Jun;20(3):241–248. PMID:22696024

40. Schapire RE, Freund Y, Bartlett P, Lee WS. Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. Ann Stat Institute of Mathematical Statistics; 1998;26(5):1651–1686.

41. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. SSD: Single Shot MultiBox Detector. Computer Vision – ECCV 2016 Springer International Publishing; 2016. p. 21–37.

42. Baker S, Matthews I. Lucas-Kanade 20 Years On: A Unifying Framework. Int J Comput Vis 2004 Feb 1;56(3):221–255.

43. Held D, Thrun S, Savarese S. Learning to Track at 100 FPS with Deep Regression Networks. Computer Vision – ECCV 2016 Springer International Publishing; 2016. p. 749–765.

44. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis 2010 Jun 1;88(2):303–338.

45. Duda RO, Hart PE, Stork DG, Others. Pattern classification. 2nd. Edition New York 2001;55.

46. Brescia AA, Rankin JS, Cyr DD, Jacobs JP, Prager RL, Zhang M, Matsouaka RA, Harrington SD, Dokholyan RS, Bolling SF, Fishstrom A, Pasquali SK, Shahian DM, Likosky DS, Michigan Society of, Thoracic, Cardiovascular Surgeons Quality C. Determinants of Variation in Pneumonia Rates After Coronary Artery Bypass Grafting. Ann Thorac Surg 2018 Feb;105(2):513–520. PMID:29174785
47. Dimick JB, Scott JW. A Video Is Worth a Thousand Operative Notes. JAMA Surg. 2019. p. 389–390. PMID:30673061
Supplementary Files
Figures
Conceptual Model.

Figure 1. Conceptual Model

- **Surgeon Technical Skills**: Surgical Approach, Experience, Speed, Agility, Coordination, Smoothness of Movements/Maneuvers
- **Intraoperative Performance**: Teamwork, Communication, Situational Awareness, Leadership, Decision Making, Task Management
- **Patient Factors**: Demographics, Comorbidities, Laboratory Values, Previous Interventions
- **Organizational Factors**: Surgical Case Volume, Complication Rates

**Complications**
- Permanent Stroke
- Surgical Re-exploration
- Sternal Wound Infection
- Renal Failure
- Prolonged Ventilation
- Mortality

**Window of Opportunity for Quality Measurement and Improvement**

- Increased Mortality
- Prolonged Hospital Length of Stay
- Increased Healthcare Expenditures
Proposed Intraoperative Recording Configuration.

**Figure 2. Proposed Intraoperative Recording Configuration**

- **Digital Video Cameras**
  - **Surgical Field (SF) Views**: Focuses on interactions among surgical team & surgeon technical skills
  - **Operating Room Team (OT) Views**: Focuses on all team members (including anesthesiologist, perfusionist and other foot traffic in and out of OR)

- **Microphone**: Lavalier lapel microphone for surgeon; Handheld dynamic microphone at head of bed