Title:
Machine Learning for Postoperative Care

Short Title
ODIN-Pilot

Registration: NCT04877535

Version: 3.0; Dec 28 2022

Version 1.10 adds an education and implementation phase between the pre-intervention and post-intervention phases. This will be necessary to give all the participating nurses time to learn where the machine-learning information is located in the Epic environment.

Version 1.11 Clarifies the activities of the education and implementation phase, including the lack of additional consent for notification of the intervention.

Version 2.0 is a major change.

- Parent study (TECTONICS) is closing, remove status as a sub-study.
    a. Specify anesthesia-control-tower clinical protocols and equipment instead of TECTONICS protocol
    b. Remove TECTONICS SMB and DSMB
    c. Remove TECTONICS data retention
- Remove "short report card" comparator
- Change to inclusion criteria:
    a. remove “observation unit status” and replace it with “risk of unplanned ICU admission in 85%ile
    b. add Transplant surgery , Otolaryngology, hepatobiliary surgery to included surgical services
    c. Remove TECTONICS contact-arm inclusion criterion
    d. Remove PACU anesthesiologist
    e. Add previously implicit inclusion criteria: weekdays during business hours
- Modify design from pre- vs post-intervention to pre-intervention plus contemporaneous controls with intervention/no-intervention on alternating days
- Clarify many aspects of intervention screening and mechanics as these have updated
- Rename "run-in phase" to "education and adaptation phase"
- Remove investigator Wildes (leaving institution)
- Eliminate secondary outcome measure (handoff checklist)

These changes are based on (1) feedback from existing participants (2) technical issues and time-of-day making the existing recruitment strategy too slow (3) transition of parent RCT to usual clinical operations

Version 3

- **Adds additional time because of updates to the format of the report card in response to participant feedback.**
- **Adds anesthesia clinicians to those who are alerted to the report card**
- **Adds vascular surgery patients, because the underlying clinical alarms suggest that they are a very relevant group**

Roles and responsibilities:

Principal Investigator:
Christopher Ryan King
Assistant Professor
Washington University Department of Anesthesiology
660 South Euclid Ave Box 8054
St. Louis, MO 63110
P: 314-362-6978
E: christopherking@wustl.edu

Institutions:
Washington University School of Medicine
660 South Euclid Ave Box 8054
St. Louis, MO 63110
Michael S. Avidan, MBBCh
P: 314-273-4136
E: avidanm@wustl.edu

Other Key Personnel:
Philip Payne, PhD
Professor
Washington University School of Medicine
Department of Medicine and Institute for Informatics (I2)
4444 Forest Park Blvd 06309
St. Louis MO, 63110
314-273-1088
prpayne@wustl.edu

Joanna Abraham, PhD
Associate Professor
Washington University School of Medicine
Department of Anesthesiology and Institute for Informatics (I2)
660 S. Euclid Avenue
Campus Box 8054
St. Louis MO, 63110
314-286-1768
joannaa@wustl.edu

Yixin Chen, PhD
Professor of Computer Science and Engineering

Washington University Department of Computer Science and Engineering
660 S. Euclid Avenue
St. Louis MO, 63110
314-935-7528
Ychen25@wustl.edu

Michael S. Avidan, MBBCh
Professor and Chair,
Washington University Department of Anesthesiology
660 South Euclid Ave Box 8054
St. Louis, MO 63110
P: 314-273-4136
E: avidanm@wustl.edu

Funding:
National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number KL2 TR002346

Non-Technical Summary

Although surgery and anesthesia have become much safer on average, many patients still experience complications after surgery. Some of these complications are likely to be avoided or less severe with early detection and treatment. Barnes-Jewish Hospital has recently started using an Anesthesia Control Tower (ACT), which is a remote group lead by an anesthesiologist who reviews live data from BJH operating rooms and calls the anesthesia provider with concerns to improve reaction times and improve use of best-practices treatments. The ACT also uses machine learning (ML) to calculate patient risks during surgery as a way of measuring when the patient is doing better or worse.

We think that many times the right interpretation of the data or the right treatment to give isn't clear until the surgery is nearly finished. The medical team in the recovery room (post-anesthesia care unit, PACU) and surgical wards is responsible for deciding the treatment strategy, but they don't have all the information available from the intraoperative monitors and events. The patient may still be too sedated or confused from the anesthesia to explain much about their preoperative history. Our goal is to use ML to create a report card for each patient that summarizes the preoperative assessment and intraoperative data in a way that is useful for postoperative providers. In addition to communicating specific facts and risks, the report card may cause the postoperative provider to ask different questions at handoff and spend more time with vulnerable patients.

The goal of this study is to integrate these ML reports into the clinical workflow and to determine if it does affect handoff behavior. The screened patients will be all adults having surgery at BJH with the division of Acute and Critical Care Surgery, Otolaryngology, Hepato-biliary Surgery, and Abdominal Transplant Surgery with elevated risk of unplanned ICU admission. Exclusion criteria are a planned ICU admission or outpatient surgery (no hospital admission).

For each included patient, the ACT clinician will review the report card information, and the postoperative providers will receive it via a print-out or an Epic Best Practices Advisory integrated into their workflow. Our study will include 3 phases: (1) pre-intervention observation of handoff practices, (2) education and workflow adaptation of the intervention, and (3) non-randomized evaluation. During the education and non-randomized evaluation phases, some days will be without-intervention and others with-intervention. On with-intervention days, all eligible patients on a given day will receive the intervention. On without-intervention days no patients will receive the intervention, but handoff outcome measures will be assessed. The main comparison will be between with-intervention and without-intervention days in the evaluation phase. The primary outcome measure we will study is handoff effectiveness in the recovery room or wards. We will also survey providers on any inaccurate items, or major omissions.

This study will be conducted with a waiver of informed consent. It offers non-significant risk to patients. The ML report card will not recommend specific treatments, and decisions will remain in the hands of the supervising physician in all phases of care. The postoperative provider will also be given information about the report card and its limitations.

1 Background and Rationale

The operating room (OR) to post-anesthesia care unit (PACU) handoff is a challenging transition for patients where communication errors frequently occur (Segall et al., 2012); subsequent adverse events are common and contribute substantially to mortality, costs, and other patient harms (Hines et al., 1992; Pearse et al., 2006; Story et al., 2010). Preventing loss of information during handoff is an important need to allow accurate early intervention in newly admitted patients. Handoff research has almost exclusively focused on checklists of expert-specified topics that should be discussed. These checklists by their nature do not adapt to the specific patient, and there is a tradeoff between checklist length-thoroughness and user fatigue-compliance. Machine learning risk prediction has been widely recognized for its potential to decrease medical errors and failure to rescue (Xiao et al., 2018). However, provider uptake has been slow, often preferring simple scoring systems (Sanchez-Pinto et al., 2018). Providers skepticism of black box predictions and the difficulty of interpreting ML outputs has been suggested to be a major hindrance (Maddox et al., 2018). As a result, no published research explores the ability of ML to augment information transfer during handoff. Additionally, few clinical ML projects have been rigorously evaluated for changes in user behavior or designed for end-user comprehension; development has focused on raw predictive accuracy, despite many attempts at creating “interpretable” ML. Innovations in ML reports that improve the adoption and provider-understanding of ML predictions are a scalable, low-cost intervention to address the need for improved information transfer and identification of at-risk patients in the postoperative setting.

At BJH, ML-assisted telemedicine for intraoperative providers is provided by the Anesthesia Control Tower (ACT) to allow early identification of errors, optimize early treatments, and prevent lapses in standard of care. The ACT includes ML-based predictions of multiple postoperative complications to identify at-risk and deteriorating patients; the intraoperative provider has indirect access to these predictions via a clinician in the ACT who filters the raw output and considers the ML prediction in their clinical context.

However, the core ACT clinical protocol does not contemplate providing the data summary and ML-prediction information to the receiving provider in the PACU or surgical wards. The proposed Outcome predictions, Data summaries, and Interpretive Notes Report (ODIN-Pilot) study will provide a "report card" containing automatically generated summaries of pre- and intraoperative data along with ML output. This information will be presented to multiple postoperative providers with different contexts and use cases. PACU physicians and registered nurses (RNs) receive handoff from the intraoperative team and assume care for the patient in the immediate postoperative phase (typically 45 minutes to several hours). They observe for appropriate recovery from anesthesia, surveil for developing complications, treat common side effects of anesthesia. When the patient is ready to leave the PACU and communicate their findings to the wards (surgical inpatient unit) RN. The PACU team also communicates to the intraoperative surgical team, who communicates to the wards "midlevel" (advanced practice nurse or resident MD) responsible for routine medical oversight, modifying medication strategies, and ordering additional diagnostic studies if necessary.

We hypothesize several mechanisms by which this ML report card may improve patient outcomes. First, the report card will function as an “alert” for patients at high risk and direct attention to them. This additional review may reduce postoperative complications. Second, the report card may frame handoff communications around important data and issues relevant to the patient. By focusing the postoperative provider on potential problems, it may improve anticipatory guidance between the intraoperative, supervisory, and frontline team members.

This pilot study will not focus on patient outcomes, but will attempt to measure these potential mechanisms. We will survey postoperative providers to measure perceived handoff quality and coverage of important topics. We will validate these surveys with direct observations, debriefing interviews, and structured handoff evaluation tools.

The second objective of this study is to improve the usability of the ML report card and determine the rate at which clinicians believe the information it contains to be inaccurate. Clinical data is often "messy," and it is possible that the mechanisms we use to filter data may make mistakes. We will survey providers on errors (including omissions) to determine the "safety" of our implementation of the postoperative ML report card.

1.1 Choice of Comparators
Based on the lack of existing objective evidence of benefit for ML predictions in the perioperative setting, and the widespread belief that clinicians do not trust ML predictions, we will compare the ML report card intervention to usual care.

2 Objectives
The objectives of the study are to determine the interpretability, workflow role, and effect on communications of showing report cards containing ML-based risk profiles based on pre- and intra-operative data to postoperative providers.

2.1 Primary Objective
The primary objective will be to determine the effect of postoperative ML report cards on handoff effectiveness.

2.2 Secondary Objectives
2.2.1 Key Secondary Objectives
The secondary objective is to determine the rate of clinician-noted inaccuracies or major omissions in postoperative ML report cards.

3 Trial Design
The study will be a non-randomized experimental design with 3 phases. First, data about current practices will be collected before interventions begin. Second, we will pilot the mechanics of delivering the report card while educating users about it. Third, we will measure handoff and usability outcomes on alternating days with and without the intervention. (2:1 intervention to non-intervention). We anticipate the number of intervention-deliver and education days to be 30, but may have to periodically update the report card delivery requiring an additional delivery-testing period. During the entire study, we will accept and record any safety concerns from participants.

Methods

4 Participants, Interventions, Outcomes
4.1 Study Setting
This will be a single center, non-randomized clinical trial at Barnes Jewish Hospital (BJH) in St. Louis, a large tertiary care academic medical center affiliated with Washington University in St Louis (WUSTL).

4.2 Participants and Eligibility
4.2.1 Inclusion criteria
- Adults undergoing OR procedures in BJH South campus, including all of "Pod 2", "Pod

3", "Pod 5" and excluding all procedure suites such as Interventional Radiology, Parkview Tower "Pod 1", Center for Advanced Medicine "Pod 4", Labor and Delivery suites.
- Surgical service is “Acute and Critical Care Surgery”, Otolaryngology, Hepato-biliary, Vascular, or Abdominal Transplant
- Planned non-ICU disposition ("floor" and "observation unit" collectively "ward" patients).
- Calculated risk of unplanned ICU admission in the first 7 postoperative days in the top 15% of historical records.

Risk of unplanned ICU admission will use the same calculator currently employed in the ACT. Risk factors considered include the surgery performed, demographics, laboratory values, comorbidities, vital signs, and medications.

Although patients are exposed to effects of the intervention, data is obtained from healthcare providers. Included providers are PACU nurses, PACU anesthesiology physicians, wards nurses, and wards midlevel providers (resident physicians, physician assistants, and advanced practice nurses) who are involved in a postoperative handoff for these patients.

4.2.2 Durations and sample sizes

The planned duration and size of each phase is

- Pre-intervention observation: 60 working days, 50 patients
- Education and intervention delivery testing: 30 working days, 50 patients
- Intervention evaluation: 90 working days, 350 patients divided into 30 without-intervention days and 60 with intervention days.

The exact number of providers surveyed is difficult to project, because some providers will care for multiple patients, nurses and midlevel providers change their assignments, and staff turnover may occur. We anticipate surveying fewer than 60 distinct PACU RNs, fewer than 15 PACU residents, fewer than 20 PACU physicians, fewer than 60 distinct floor RNs, and fewer than 25 distinct midlevel providers.

4.2.3 Case screening and identification

Patients eligibility is pre-screened automatically, and a researcher or ACT clinician verifies their eligibility. The first two enrollment criteria

- Adults undergoing OR procedures in BJH South campus, including all of "Pod 2", "Pod 3", "Pod 5" and excluding all procedure suites such as Interventional Radiology, Parkview Tower "Pod 1", Center for Advanced Medicine "Pod 4", Labor and Delivery suites.
- Surgical service is “Acute and Critical Care Surgery”, Otolaryngology, Hepato-biliary Surgery, or Abdominal Transplant Surgery

are taken from the surgical schedule. The third enrollment criteria

- Planned non-ICU disposition ("floor" and "observation unit" collectively "ward" patients).

is taken from (1) the procedure schedule request , (2) the anesthesiologist day-of-service anesthesia plan documentation, (3) a predictive model based on the scheduled surgery. Cases with > 75% likelihood of direct ICU admission or outpatient discharge are automatically excluded by the predictive model. The final criterion

- Calculated risk of unplanned ICU admission in the first 7 postoperative days in the top 10% of historical records.

is calculated using the same model described in 4.3.2

Once the server has identified patients as likely eligible, they are marked for review in the ACT

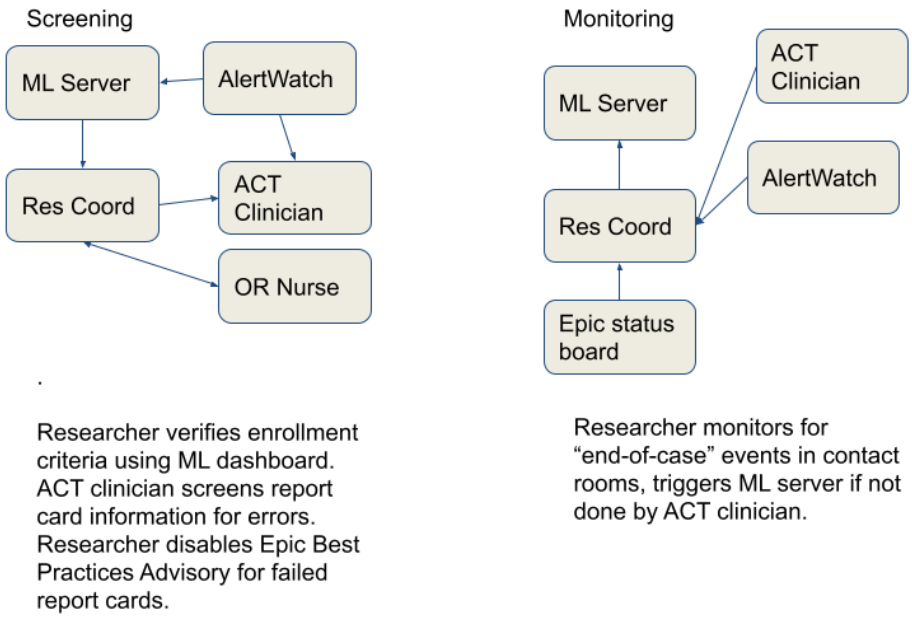
ML dashboard, where final eligibility is confirmed.

During the intervention evaluation (phase 3), each day will be with or without intervention for all eligible patients. These will rotate 2:1 over work days , eg week 1 = (with, with, without, with, with), week 2= (without, with, with, without, with), week 3 = (with, without, with, with, without)

4.3 Interventions

Figure 2 is a flow diagram of the activities during a patient's day and Figure 3 is a patient timeline.

Figure 2: Flow of research activities



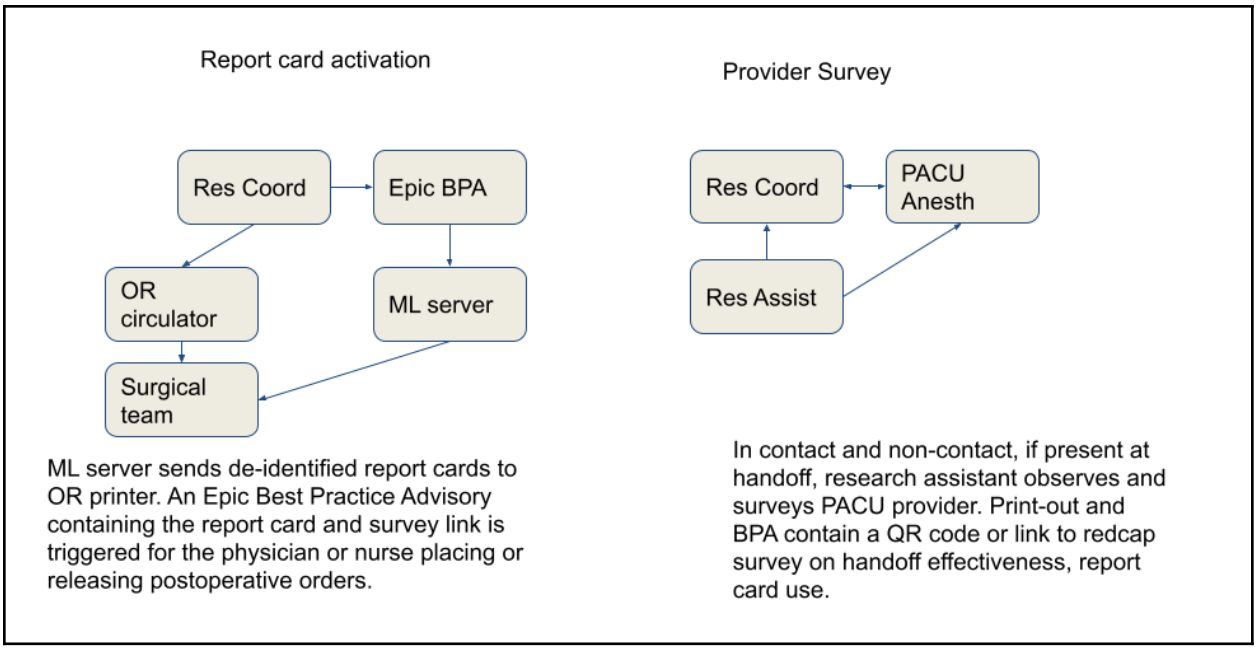


Figure 3: Patient activity timeline

Begin OR	Identify
T0: Surgery "close"	Validate enrollment Validate
T0 + 1	Survey
T0 + 3-6	Survey other

4.3.1 ML Report Card

The core intervention is the ML report card containing predictors of adverse events. Figure 4 is an example of the printed report card on a simulated patient. No personally identifiable information is included, other than a randomized study ID. The report card includes the operating room and date, to allow providers to know that it is for the current patient. Because the printouts are only created in operating rooms and PACU, this is sufficient information to place it in the correct chart. However, at other times the report card itself does not contain sufficient information for anyone without the studyID-patientID mapping file or pre-exising access to the EHR to determine the patient. Operating room number and time is included, but these can not be mapped to patients without the detailed surgical schedule in the EHR. The QR

code study links to an anonymous survey (no patient or provider information) for problems with the report card for the education and testing phase. The evaluation phase survey includes the additional questions in the appendix (triggered on date.

The ML risk profile generated for each patient will include risk of 30 day mortality, pulmonary complications and respiratory failure, acute kidney failure, transfusion, deep venous thrombosis, cardiac complications, and unplanned ICU admission. Additional predicted outcomes may be added based on feedback from healthcare team members. Risk profiles will display the average for all patients undergoing the same surgery, the risks calculated using preoperative assessments only (i.e. excluding intraoperative data), and the end-of-case values (i.e. including preoperative and intraoperative data). Literature-based estimates of probability of postoperative nausea and vomiting and severe postoperative pain will also be included where elevated. Expert-curated summaries of important variables will be included in the Epic report card, which include: estimated blood loss, fluids given, opioids given, duration of anesthesia, antibiotics given. Explanatory tools for the risk profile will also be generated based on SHAP values [Lundberg 2017] identifying the variables with greatest influence on predictions

Figure 4: Example ML report card

Please place this with the preop/pacu signout sheet

Generated at 06/20/2022, 12:11:07 PM for Room: BJH OR POD 3 ROOM 302

This is an experimental (IRB: 202103210) integration of AI to identify high risk postop patients. **This does not replace usual handoff, but if risks below are high, it may be worth adding the plan and assessment to your handoff.** Many factors are not included in the prediction models; always use your clinical judgement. With concerns or questions please contact **christopherking@wustl.edu**. Please use the QR code to take an anonymous survey (no PHI, phone is ok). For BJH computers with a scanners, open chrome, click the address bar, and scan; some scanners add a leading and trailing slash. We really want this feedback to make it work better! I can also be reached at (314) 323-1585.

Postoperative risk report

Outcome	Outcome % for patient	Risk percentile	Surgery rate	Average outcome rate
death_in_30	0.8	76	0.9	1.8
pulm	0.1	70	0.1	0.3
transfusion	6.8	69	20.8	11.5
DVT_PE	0.5	67	2.4	0.9
cardiac	1.3	66	5.5	2.6
ICU	4.8	63	24.3	15.2
post_dialysis	0.0	60	1.4	2.4
long_vent	1.0	55	4.4	6.0

first 7 days. death_in_30 = death in 30 days.

"Surgery rate" = prediction based on the words contained in the scheduled surgery, Average outcome = all inpatient postop at BJC. Outcome definitions: Cardiac = new diagnosis of CHF or MI, elevated troponin or BNP, new A Fib. Pulm = pneumonia, reintubation in 5 days. postop_dialysis = new dialysis. DVT_PE = diagnosis of DVT or PE. Transfusion = transfusion > than 1 unit of any product. ICU = ICU in

Table 1: Largest contributors to overall adverse events

Variable	Value	relative import
Deep Venuous ThrombosisMissing	0.0	1
Diabetes MellitusMissing	0.0	0.88
Partial Thromboplastin Time	29.0	0.63
Pulmonary Hypertension	0.0	0.63
Obstructive Sleep Apnea	0.0	0.54
Alkaline Phosphatase	416.0	0.49
Weight	70.8	0.29
PreOp Systolic	NA	0.25

Methods note

Risks generated using pre-anesthesia assessment, major comorbidities, laboratory values, and preoperative vital signs; errors in these values can significantly affect the calculations. Risk calculations use gradient boosted decision trees trained on BJH inpatient data 2019-2021. see https://pubmed.ncbi.nlm.nih.gov/31558311/

study ID 86915c307cbd5e01531052a04ff901d94371e8bf

The ACT detects that cases are near completion and that a report card should be triggered in two ways. First, the circulating nurse in each room electronically marks that a room is "close" in the EHR, usually during closure, to ensure that subsequent patients are called to the preoperative area and supplies for the next case are readied. This is performed with high fidelity as it is considered important to prevent case delays. The displayed OR schedule in Epic changes color, making this easily recognized. Second, an algorithm detects when changes in the anesthesia record suggest that the case is concluding. Both of these features cause the ML dashboard in the ACT (Figure 5) to highlight the patient as needing review to validate the report card. After manual validation, the report card will print directly in the operating room and a notification is sent to the circulating nurse to add it to the chart and provide it to the anesthesia team for handoff. The risk profile is then made active in a Best Practice Advisory in Epic.

Figure 5: Anesthesia Control Tower ML Dashboard with patient information redacted. Yellow highlight implies "ready for review." Additional columns not shown for space.

Room Name	Description	akigr1	cardiac	death_in_30	DVT_PE	ICU	post_dialysis
BJH OR POD 2 ROOM 209	###LAPAROSCOPIC CHOLECYSTECTOMY	2.90%	1.60%	0.10%	8.80%	5.20%	0.10%
BJH OR POD 3 ROOM 301	EXPLORATION LIVER TRANSPLANT WITH WASHOUT LIVER BIOPSY AND BILE DUCT REVISION	67.10%	6.90%	7.60%	1.20%	93.40%	0.90%
BJH OR POD 2 ROOM 211	TRACHEOSTOMY PERCUTANEOUS	40.30%	2.70%	3.50%	7.40%	94.50%	0.20%

Communicating from the ACT to the PACU physicians presents several advantages. First, the ACT clinician is familiar with the general model of ML output generated by AlertWatch and will filter out obvious errors from reaching the care team and potentially harming the patient, minimizing risk to participants. Second, because on a given day a single attending physician and resident oversees the PACU at BJH South Campus, friction related to communication between the ACT and these providers will be minimized as these contacts become routine. Both parties will have the opportunity to rapidly adjust to the presence of the report card as they will process multiple such reports per day. Third, because the ACT-postop communication happens *before* formal handoff, it will allow these providers to look into higher risk cases before they arrive and ask more detailed questions during handoff. We anticipate that most interactions will consist of simply forwarding the report card. The ACT will not communicate directly to PACU nurses or ward providers.

ML output will be generated on a server controlled by the investigators and administered by Washington University Information Technology. This is the same server and data access mechanism used by the ACT team. A feed of preoperative and intraoperative data extracted from the Epic EHR designed for use in the AlertWatch system is currently saved in a database maintained by BJH and exposed for authorized users to access, which we will access using encrypted transport mechanisms. The server will also have a TLS-encrypted and password protected web interface (dashboard) for debugging, message generation, and manual end-of-case triggering. The web interface will allow study technical staff and ACT clinicians to look up current and previous risk profiles by study identifier, patient identifier and date, or date, time, and operating room number. It will also allow them to manually enter that a case is concluding, and to mark a day as "with-intervention". The study identifiers, date, time, end of case marker, intervention indicator, and ML output will be stored on the WUSTL Research Infrastructure Services Data Storage Platform. The RIS Data Storage Platform is a scalable distributed storage solution with an integrated long-term archive. It is institutionally approved for protected health information, and requires WUSTL-key authentication and 2-factor authentication outside WUSTL facilities. Data access is granted only on a per-user basis and logged. The ML server is behind the Washington University firewall and inaccessible from off-campus computers.

To avoid misleading postoperative providers, we will create a brief educational sheet explaining the purpose of the study and the items included in the risk profile (attached). The educational sheet will include the calibration accuracy and other appropriate metrics regarding the predictions. For example, at the screening threshold we will use the sensitivity for AKI is about 50%, meaning that the provider should not be falsely reassured by not receiving an alert. We will stress that these risk predictions were formed in the context of usual care at BJH, and that a

lower risk prediction does not justify withholding a therapy that they would otherwise give. These sheets will be linked in the report card. A draft of the sheet is included with this application, but it may be modified based on feedback from providers. We also plan to periodically recalibrate the ML algorithms based on new data, and will update the accuracy metrics accordingly.

4.3.2 Forecasting algorithms.

The core of the intervention is a ML-based summary of predicted postoperative risks, the change in those risks during the case, the most relevant input variables for ML predicted risks, expert-curated important variables, and other "explanatory tools" derived from the ML risk prediction method. Transcribed below from the TECTONICS (IRB# 201903026) protocol is the method for generating ML risks from preoperative and intraoperative data (internal citations omitted):

We are integrating the machine-learning forecasting algorithms for adverse outcomes with our other alerting tools, and will be testing and improving them over the course of the study. The feasibility of this integration is supported by a previous successful trial, where members of our investigative team, using live data from the EHR, implemented machine learning algorithms to guide a rapid response team in medical wards. Furthermore, the results of the prior study provide strong support for using the multi-path convolutional neural network (MPCNN) model to build and regularly improve predictive machine-learning algorithms that will be used to forecast adverse outcomes (e.g., renal failure, respiratory failure, delirium, death) in the TECTONICS trial. We found that our algorithms, which are based on extensive patient information (i.e., demographic characteristics [including sex as a biological variable], surgical risk, co-morbidities) as well as time series physiological data (e.g., blood pressure, temperature, heart rate) predict outcomes such as death with high area under receiver operator characteristics curve (~90%), acceptable sensitivity (~50%) and excellent specificity (~95%). These algorithms will augment dynamic risk assessment in the ACT. Our MPCNN algorithms are unusual in that they are designed to handle heterogeneous data types, including numerical features, categorical features, and multivariate time series. By integrating feature preprocessing and model learning phase together, our algorithms are able to jointly learn complex feature interactions among different data types, ultimately improving predictions of adverse perioperative outcomes (such as in-hospital mortality, postoperative acute renal failure, postoperative respiratory failure). As our models accept raw data directly, high quality real-time prediction can be achieved.

The machine learning algorithms that will be used as part of the TECTONICS trial were developed, tested, and validated using multiple studies. We have tried many other algorithms, including in our ACTFAST2 study, and convolutional neural network (CNN) gives the best overall performance. We recently compared our CNN method with more than 10 other state-of-the-art time-series prediction algorithms and showed that CNN gives the best predictive performance. Our team members have been pioneers in the use of machine learning to predict adverse patient trajectories in the setting of critical illness. We have well-developed sophisticated machine learning techniques, and the application of our algorithms to the intraoperative setting – where changes in patient trajectory occur very rapidly – is a novel application of this technology.

4.3.3 Other Interventions

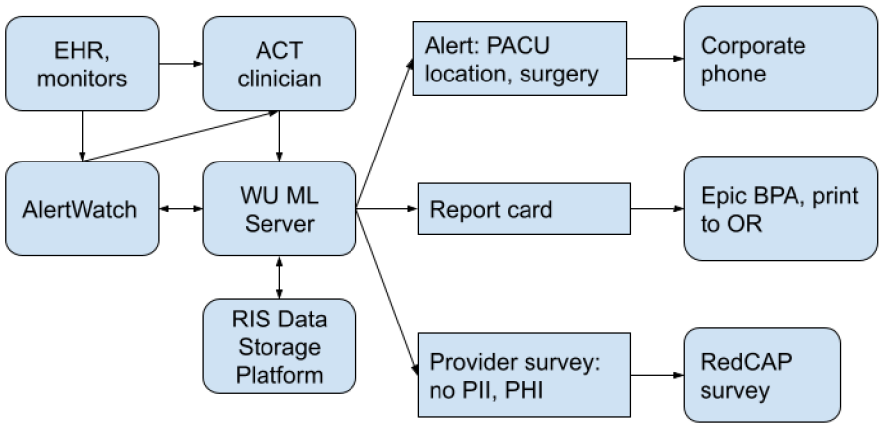
We will notify the PACU anesthesiology resident when report cards are generated for high-risk patients. This notification will occur by SMS to a BJH-issued phone which is passed between residents at each shift. The notification does not contain PHI or PII, only that a particular OR has a high-risk patient.

Discussed in the Outcomes section, the study also includes a survey for providers, which we mention here because of its patient-information security implications. The Epic BPA (and printout) will contain a patient-specific link and QR code to a RedCap survey. This survey contains no PII, only a randomized study ID.

Information flow and security. To minimize the risk of protected health information (PHI) or personally identifiable information (PII) leaking from the health-care environment, we will restrict communications to providers to secure and auditable delivery modes. A diagram of the flow of information is included in Figure 6. There are 3 distinct flows of information: from the ML server and ACT tower to the PACU provider notifying them that a high-risk patient is emerging (the alert), from the ML server and ACT to the healthcare provider in the body of the report card, and between the postoperative provider to the research coordinator with their survey responses (provider survey).

The alert is designed to include minimal potentially sensitive information. It will contain the location of the patient (operating room, PACU bay destination) and that the risk of patient decompensation is elevated. The name, MRN, date of birth, or other identifiers are not included.

Figure 6: Information flows and security



Flow of information and activities in study. EHR data to alertwatch via secure database replication (BJH controlled server), accessed by domain and wustl-key to WUIT server for ML calculations. Wustl-key plus VPN authentication for ACT clinician dashboard. RIS Data Storage accessible only by wustl-key plus VPN authentication. Message to PACU phone by Epic Secure Chat or SMS without any PII or PHI. Print by ML server via encrypted ipp (ipps) without PII to operating room. ML server-Epic communication secured by TLS and API-key authentication. RedCAP survey contains only study identifier, uses https.

The only individuals who can access the report card (which contains no PII) are (1) Epic users who access the patient's chart (2) pre-approved users in the ACT who log in with their wustl-key (3) individuals reading the patient's paper chart. The ML server will be behind the WashU-BJH firewall, and therefore only accessible from authenticated users on campus or logged in via VPN. Access to the report card will be logged by IP address.

The survey is an outcome measurement, discussed in more detail below. The survey link is patient and date specific using the study ID, but does not contain the patient name or other identifiers. The survey link will be contained in the Epic BPA and the print-out in the form of a QR code; because study IDs are long and randomized they are functionally un-guessable. Versions of Android and iOS since 2018 include QR code and web-link identification from the

default camera application. Providers can also use BJH bedside computers, which are equipped with scanners that read QR codes.

4.3.4 Modifications

We anticipate that the form of the report template will vary as we conduct the study based on feedback from providers and qualitative observations of handoffs conducted as part of other research projects. The primary objective (comparison of contact versus noncontact) will be unaffected. The total number of ACT-postoperative provider contacts is limited by the staffing of the ACT and alert fatigue in the postoperative provider. We may adjust the screening threshold to avoid overwhelming either provider. We will record PACU participation, and if participation in handoff has been low, or the overall subjective utility is low we will revisit the method of ACT-postoperative provider communication. We anticipate small improvements and changes to the underlying risk prediction algorithm. This is required because of ongoing changes in the data sources that feed into the algorithm, changes in documentation practices, and periodic need to re-calibrate predictions based on changes in unmeasured risk factors. We will update the information provided with the report card to reflect any such changes.

4.3.5 Adherence

Crossover from non-contact to contact will not occur because report cards are not generated for these patients. Our outcome survey will detect if providers believe they used the report, but no other encouragement to use it will be provided. Cross-over into non-contact (failing to contact the correct provider) will be minimized by appropriate training to ACT clinicians.

4.3.6 Concomitant Care

ACT-intraoperative provider contact will be directed by clinical protocol. Intra- and postoperative care will be directed by the responsible physicians without any restrictions from this trial.

4.4 Outcomes

4.4.1 Primary Outcome Measures

The primary outcome will be self-reported handoff effectiveness.

The following survey questions will be asked of postoperative providers:

> Did you receive at handoff all the information you needed to safely take care of this patient?
> [Yes, No]
>
> Did you receive information from the team giving handoff to explicitly help you manage any potential complications or problems?
> [Yes, No]
>
> If so, please list what problems / topics were discussed?
> [Free response]
>
> [If with-intervention phase *] Did you look at or discuss the postoperative report card for this patient?
> [Yes, No]
>
> [If * yes] Did the report card contain any information that was notably misleading or incorrect?
> [Yes, No]

[If yes, please describe: text box]

[If * yes] Are there any important pieces of information that the computer should have included in the report card but didn't?
[Yes, No]
[If yes, please describe: text box]

[If * yes] How much did the ML report change questions asked or information transfer at handoff?
[Not applicable , Not at all , a little , a fair amount , a great amount]

[If * yes] How much effort was accessing the report card?
[Very easy, easy, some effort, hard, unable to get to it or had to ask for help]

[If * yes] Was the information presentation in the report card easy to understand?
[Easy, some effort, hard to understand, unable to use or had to ask for help]

[If * yes] Was the information in the report card relevant?
[a great amount, a fair amount, modestly, mostly not, not at all, not sure]

[If * yes] Was the information in the report card useful to decision making or helpful to your understanding of the patient?
[a great amount, a fair amount, modestly, mostly not, not at all, not sure]

The following questions will additionally be asked of PACU resident physicians:

Did you participate in handoff for this patient?
[Yes, No]

Did you visit this patient's bedside or speak to their nurse before being asked to sign them out?
[Yes, No]

Figure 7 illustrates the RedCAP survey for safety questions which will be included throughout the study.

Figure 7: RedCAP Safety Survey

Postoperative AI Risk Report Card

Resize font: ⊞ | ⊟ Returning?

This form is to collect reports of problems with the AI generated risk estimates for postoperative patients. You can contact the study PI at christopherking@wustl.edu if this doesn't cover your needs. The form doesn't track who you are (it is anonymous) but it does try to link to a patient so that we can figure out what caused the problem. This is an experiment (IRB: 202103210) and we appreciate all feedback. The study team will never try to figure out who submitted reports unless you ask us to follow up with you.

This form is to report safety problems with the AI postoperative risk report card. This form does not track who you are, but it does try to link to a patient so that we can determine what happened under the hood. This survey form (REDCap) is approved for HIPAA-protected information.

This field should auto-populate with a study ID from the link you followed to get here. Otherwise, (1) if you have a QR code with the email me link, and you are at a machine with a barcode scanner, click in the below field and scan the QR code (2) this is a HIPAA-protected portal, you can enter any of the Epic patient IDs and the study team will link it to the correct person.

* must provide value

d46e6e6bb147f7ebdf645e2e9be156172c32897e

Expand

Common problems:

- ☐ The report card is for the wrong patient
- ☐ The report card generated on a non-surgical patient
- ☐ The report card generated too early (preop)
- ☐ The report card generated too late (not immediately postop)
- ☐ The risk estimates are too low
- ☐ The risk estimates are too high
- ☐ The risk estimates were otherwise confusing
- ☐ The report card does have / use data that I think it should
- ☐ The report card has incorrect data
- ☐ The report card display / formatting was broken

This problem made the report card unsafe

* must provide value

- ○ Yes
- ○ No

reset

Tell us anything about the problem

Expand

If you have comments or suggestions that aren't a problem with the report card, include them here

4.4.2 Secondary Outcome Measures

A research assistant will directly observe up to 150 OR - PACU and PACU-Ward handoffs in the intervention phase. An assessment form is attached. After direct observation, the research assistant may ask unstructured clarifying questions and present the survey questions discussed above.

We will conduct 20-30 debriefing interviews with ward and PACU clinicians on their perception of handoff effectiveness and the usefulness of the report card. These will be audio recorded and transcribed.

We will extract exploratory clinical outcomes from the EHR: ICU admission, hospital length of stay, discharge disposition (including death), and changes in serum creatinine (acute kidney injury).

4.5 Sample Size

Precision on the absolute rate of clinician-rated inaccurate reports depends on the rate, but for 0/150 clinically inaccurate reports, the upper 95% confidence limit is 2.4%. Handoff surveys will be dichotomized into (un)acceptable. If the baseline rate of "acceptable" handoff is 70%, 50 patients without and 100 patients with intervention implies an 18% absolute increase as the minimum change with 80% power. The study is designed primarily to assess the perceived safety and workflow integration of the report card, not any patient outcomes.

4.7 Recruitment

Because of the nature of the study, recruitment of patients will be automatic.

PACU providers will be determined from the schedule and approached in the morning of their shifts for consent for participation.

4.8 Allocation

Allocation is determined solely by entry time into the study.

4.9 Blinding

Patients will not be directly informed of their involvement in ODIN-Pilot, nor are they likely to discover this as the intervention occurs while they are under or emerging from anesthesia. Clinicians will be unblinded by the nature of the intervention; however, they will not know if a non-contact patient is included in the study.

5 Data collection, management, analysis

5.1 Data Collection Methods

Pre- and intraoperative data will be extracted from the EHR and AlertWatch as specified in the TECTONICS (IRB# 201903026) protocol identically to the clinical operations of the ACT. Postoperative exploratory outcomes and patient demographics will be extracted from the EHR by an Institute for Informatics data broker. Survey instruments are included in the appendix and above. Because of the nature of study timing and extraction of data collected during the normal clinical process, no additional retention efforts of patient-level communication is planned.

5.2 Data Management
Preoperative and intraoperative data used for machine learning will not be retained for this study, but other studies and clinical operations may share the same hardware. Only the minimum necessary private patient information will be collected for the purposes of the study. Any protected health data will be kept in a secure digital environment that is password protected and limited to research team only as discussed above. De-identified data may be kept and used in future studies not pre-specified in the above protocol. The investigators are responsible for ensuring the accuracy, completeness, legibility, and timeliness of the data collected. Data storage and safety practices are discussed above. Surveys of postoperative providers will be stored on REDCap.

The ACT will use clinical applications available to all clinicians to monitor ongoing surgical procedures. These applications can only be accessed over the secure hospital network or by VPN. This arrangement meets and/or exceeds HIPAA standards Patient Health Information (PHI) security. AlertWatch® is an approved clinical application at Barnes-Jewish Hospital, and therefore maintains these same levels of protection. Access to the data collected in this study will be restricted to approved personnel. It is a strict policy that PHI content cannot be reviewed outside of this protected environment. When possible, extracts from this project will avoid the use of PHI. De-identified patients can be identified using a special, non-PHI primary key which will be stored on the RIS Data Storage Platform.

Data will be maintained for at least seven years after the close of the study in accordance with Washington University guidelines. Further study record retention will be at the discretion of the study investigator. Data may be used for future studies not mentioned in the protocol.

5.3 Statistical Methods
Differences in the rate of acceptable handoffs will be compared with chi-square contingency table tests.

Although the primary analysis is intention to treat (comparing the with- and without-intervention groups directly), we will report as exploratory analyses per-protocol analyses in which patients whose report card was never accessed (outside of the ACT) are excluded.

5.4 Education phase
Between the pre-intervention and intervention phases of the study, staff members of the study hospital units will be alerted to the function and location in the EHR of the report card. Written material including the same information and research team contact information will be left in the breakrooms of these units and disseminated by email. Research team members will answer any questions and record any concerns. All staff comments and concerns will be recorded completely anonymously. Research team members will review these comments and concerns to determine if any modifications to the intervention or clarification to the educational material is required. During this period, the report card will be marked as in-testing.

6 Monitoring
6.1 Adverse event and safety monitoring:
No data safety monitoring committee is planned for this sub-study due to the non-significant risk and pilot nature. No external safety monitor is planned, though all adverse events will be reported to NCATS (funder of the study).

6.2 Interim Analyses
No formal interim analyses are planned.

6.3 Harms
The postoperative provider survey will include a link and phone number to contact the research coordinator if they believe a serious adverse event or near miss occurred related to the ACT-postoperative communication. As an example, they may believe that ML risk estimates were sent for the incorrect patient. Our communications will all include the role of the ML predictions as advisory, and that the responsibility for decision making lies with the provider. These will be reported to the WUSTL IRB. Descriptive statistics related to uncommon potential harms will be included in the trial reporting.

6.4 Auditing
Auditing will consist of debriefing interviews and surveys described above.

7 Ethics and Dissemination
7.1 Ethics review approval
This study has been submitted to the Human Research Protection Office at Washington University (St. Louis, MO). This study will be registered at clinicaltrials.gov.

7.2 Protocol Amendments
Any modifications to the protocol which may impact on the conduct of the study, potential benefit of the patient or may affect patient safety, including changes of study objectives, study design, patient population, sample sizes, study procedures, or significant administrative aspects will require a formal amendment to the protocol. Such amendment will be agreed upon by the WUSTL IRB. As discussed above, small changes to the ACT report template are anticipated and will not be regarded as protocol modifications.

7.3 Consent
A waiver of informed consent <u>for patients</u> is requested.

1) The study presents non-significant risk to participants. The report card is not intended as an implant and presents no potential for serious risk to the health, safety or welfare of a subject. The report card is not purported or represented to be for use supporting or sustaining human life. The report card is not being used for substantial importance in diagnosing, curing, mitigating, or treating disease or otherwise preventing impairment of human health. Treating physicians retain control over all therapies, and no specific therapies are recommended. The warning may at most “drive” clinical management. The conditions predicted in PACU are not generally life-threatening and providers have time to correct any errors that they make. The ACT clinician will screen out grossly incorrect alerts. The educational sheet presented prior to the report card will remind providers that the report card is intended to cause them to seek out additional information and draw their own conclusions about the best course of action for a particular patient; it is not an alarm intended to provoke immediate clinical action.

   There is no obvious mechanism for physical, psychological, financial, or legal harm. Patients in the usual care (non-contact) period remain in the current standard of care with critical-care level oversight by physicians and nurses in phase-one recovery. It is unlikely that physicians would take harmful action based on the information in the contact-arm report as it does not recommend specific actions. Because the education for providers will include the ~50% sensitivity of the alert, they will be aware not be falsely

reassured by the absence of an alert
As documented above, the minimal necessary amount of information about subjects is collected, and best practices are used to secure it.

2) The project is not scientifically feasible without a waiver of consent. At the time when eligibility can be determined, the patient has undergone anesthesia or sedation, and would need to have been consented in preop (or preferably sooner). Difficulty with obtaining same-day consent would lead to a much more selected population and not answer the pragmatic scientific question. It would also lead to a much slower recruitment, making it likely that the trial would fail to complete in a timely fashion and "wasting" the effort.
3) Because of the need to link intraoperative risk models to patient locations and outcomes, it is not possible to conduct it without identifiable private information.
4) The waiver will not affect the rights or welfare of the participants. As the Secretary's Advisory Committee on Human Research Protections notes in 2008, this requirement related to any other legal consent requirements, which are not present in this case. Additionally, we should consider if the population would object to the waiver if informed of it or believe that it could cause them harm. We believe that due to the novel nature of the technology, standard-of-care control arm, minimal risk, and diversity of successful PACU staffing models participants would not object to participation.
5) There are no special circumstances necessitating post-facto information provision to the patient. This requirement is intended for research involving deception.

A participation document will be included at the beginning of the workday for PACU physicians and postoperative providers receiving messages describing the study.

A partial waiver of informed consent for providers is requested. Providers are exposed to the report card and may optionally complete the above survey reporting inaccurate data. Because (1) surveys do not collect identifying information and participation requires active effort, (2) predicting the applicable providers is impossible, separate consent forms will not be obtained and we request a waiver of consent. Study information is included in the survey link which includes that their participation is completely voluntary, that no communication is made to the institutional leadership, and that all stored information is anonymized to the extent possible.

1) The study presents non-significant risk to participants. Survey responses are securely stored and contain no identifying information. Providers can skip the survey without any harm.
2) The project is not scientifically feasible without a waiver of consent. Because the midlevel and nursing providers for this subset of patients are not knowable ahead of time, we would need to consent every possible nurse and midlevel provider.
3) No identifiable private information is requested.
4) There is no foreseeable effect on the rights or welfare of the participants. Because it is unidentified, there is no mechanism for (non) participation to affect the providers.
5) There are no special circumstances necessitating post-facto information provision to the patient. This requirement is intended for research involving deception.

For direct handoff observations, participants will be approached and provided with study information before handoff begins. We will identify the relevant OR cases and approach the OR staff during emergence from anesthesia. Because the participating PACU and ward nurse is impossible to identify ahead of time (these are assigned throughout the day), we will approach and ask for written consent shortly before handoff begins. We will attempt to approach the pool of PACU nurses with study information during the morning of days on which we plan direct

observations, with the goal of having discussed the study with participants before the patient arrives.

Because many PACU nurse to ward nurse handoffs occur over phone calls, having obtained consent to observed the handoff observation from the PACU nurse, we will disclose our observation of the PACU nurse to any party on the phone, allowing them to request privacy if needed.

No consent will be recorded for the education phase in which we notify staff members how to use the intervention as all information collected in that phase is completely anonymous.

Because of the nature of the intervention. no provisions are made for ancillary post-trial care or participants who are harmed. Participants will receive appropriate care at BJH prior to discharge.

7.4 Confidentiality
Identifying data will be protected, as described above.

7.5 Declaration of interests
The investigators have no financial conflicts of interest.

7.6 Dissemination
We will disseminate the results with presentations to relevant stakeholder groups, peer-reviewed publications, and national meetings. We will also use the BJC Collaborative as a vehicle for dissemination.

7.7 Data Sharing Policy
Data from the ODIN-Pilot trial will be made available for analysis in compliance with the recommendations of the ICMJE. Individual participant data that underlie the results of the trial will be made available after appropriate de-identification, along with the study protocol and statistical analysis plan. We plan to make this information accessible to researchers who provide a methodologically appropriate proposal for the purpose of achieving the aims of that proposal. Data will be available beginning 9 months and ending 36 months following trial publication at a third-party website. Data requestors will need to sign a data access agreement to gain access to trial data. Proposals should be directed to *avidanm@wustl.edu*.

This protocol was written in compliance with the SPIRIT Checklist Guidelines for Interventional Trials.

7.8 Pandemic Infectious Risks to Participants and Study Personnel

The study has very little to no implications for exposure to sars-cov-2 infection.

The only study personnel on-site are those doing direct observations of handoffs. The study personnel will be screened by universal policy before entry to BJH, use appropriate face coverings, and practice physical distancing. During obtaining consent to observe, hand hygiene will be observed with disposable pens and paper records stored in a daily "used documents" cardboard folder to minimize accidental fomites. During observation of handoff the assistant will maximize physical distance while being able to overhear the conversation. This activity will not continue during any time of increased community transmission (e.g. cancellation of elective surgery).

The personnel in the ACT, PACU providers, and wards providers are all present at BJH as part of their usual clinical activity with no modification. Individuals are contacted only over telephone. Electronic informed consent is used, creating no additional risk of exposure.

Patients are exposed to no increased risk of exposure; no study personnel directly interact with the patient.

8. Works Cited

Chetta, A., Carrington, J.M., and Forbes, A.G. (2015). Augmenting EHR interfaces for enhanced nurse communication and decision making. In Proceedings of the 2015 Workshop on Visual Analytics in Healthcare, (Chicago, Illinois: Association for Computing Machinery), pp. 1–6.

Hines, R., Barash, P., Watrous, G., and O'Connor, T. (1992). Complications Occurring in the Postanesthesia Care Unit. Anesth. Analg. *74*, 503–509.

Maddox, T.M., Rumsfeld, J.S., and Payne, P.R.O. (2018). Questions for Artificial Intelligence in Health Care. JAMA.

Pearse, R.M., Harrison, D.A., James, P., Watson, D., Hinds, C., Rhodes, A., Grounds, R.M., and Bennett, E.D. (2006). Identification and characterisation of the high-risk surgical population in the United Kingdom. Crit. Care *10*, R81.

Sanchez-Pinto, L.N., Luo, Y., and Churpek, M.M. (2018). Big Data and Data Science in Critical Care. Chest *154*, 1239–1248.

Segall, N., Bonifacio, A.S., Schroeder, R.A., Barbeito, A., Rogers, D., Thornlow, D.K., Emery, J., Kellum, S., Wright, M.C., Mark, J.B., et al. (2012). Can We Make Postoperative Patient Handovers Safer? A Systematic Review of the Literature. Anesth. Analg. *115*, 102.

Story, D.A., Leslie, K., Myles, P.S., Fink, M., Poustie, S.J., Forbes, A., Yap, S., Beavis, V., Kerridge, R., and REASON Investigators, Australian and New Zealand College of Anaesthetists Trials Group (2010). Complications and mortality in older surgical patients in Australia and New Zealand (the REASON study): a multicentre, prospective, observational study. Anaesthesia *65*, 1022–1030.

Xiao, C., Choi, E., and Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J. Am. Med. Inform. Assoc. JAMIA *25*, 1419–1428.

Weinger, M.B., Slagle, J.M., Kuntz, A.H., Schildcrout, J.S., Banerjee, A., Mercaldo, N.D., Bills, J.L., Wallston, K.A., Speroff, T., Patterson, E.S., et al. (2015). A Multimodal Intervention Improves Postanesthesia Care Unit Handovers. Anesth. Analg. *121*, 957–971.

Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in Advances in Neural Information Processing Systems 4765–4774 (2017).

Appendix: Handoff Direct Observation Guide

Global Rating of Handover Effectiveness

[ ] Not at all effective

[ ] Somewhat effective

[ ] Moderately effective

[ ] Very effective

[ ] Extremely effective

Content and organization of handoff was

[ ] Very disorganized and unclear

[ ] Somewhat disorganized and or irrelevant

[ ] Moderate relevance and organization

[ ] Mostly relevant and organized

[ ] Very well organized and very effective delivery

Content and Completeness

[ ] No aspects of SBAR adequately covered

[ ] Some but not all critical content

[ ] All critical content (minimally sufficient)

[ ] All critical and most SBAR items

[ ] All SBAR and pertinent info, very thorough

Confirmation and comprehension

[ ] No questions or concerns raised or addressed

[ ] Questions or concerns not adequately addressed

[ ] Moderate confirmation of comprehension

[ ] Adequately addressed concerns and questions

[ ] All questions and issues fully addressed

Level of Engagement

[ ] None

[ ] Brief engagement by one or both sides, but mostly inattentive

[ ] Moderate: no negative impact on quality

[ ] Both sides engaged throughout handover

[ ] Ideal engagement