1. TITLE PAGE

Title: Anesthesiology Control Tower: Feedback Alerts to Supplement Treatment (ACTFAST)

Principal Investigator: Michael Avidan, MBBCh

Key Team Members: Arbi Ben Abdallah, PhD; Yixin Chen, PhD; Brad Fritz, MD; Daniel Helston, MD; Sachin Kheterpal, MD; Teresa Murray, MD; Mary Politi, PhD; Anshuman Sharma, MD

Other Team Members: Thaddeus Budelier, Shreya Gowasni, Alex Kronzer, Sherry McKinnon

Organization: Washington University School of Medicine

Inclusive Dates of Project: 04/01/2017 – 03/31/2020

Federal Project Officer: Janey Hsiao, AHRQ

Acknowledgment of Agency Support: This project was supported by grant number R21HS24581 from the Agency for Healthcare Research and Quality.

Grant Number: R21HS24581

2. STRUCTURED ABSTRACT (250 word maximum). Include five headings:

Purpose: The purpose of this pilot study was to implement an Anesthesiology Control Tower (ACT). The aims were (i) to develop, refine, and validate forecasting algorithms for adverse outcomes, (ii) to assess the usability of an ACT for the operating suite, and (iii) to assess whether the ACT improves clinician compliance with standards of care and surrogate measures of patient outcomes.
Scope: Approximately 2% of surgical patients die within one month, and many more experience major morbidity. No previous study has adequately evaluated the potential to leverage health information technology to systematically address quality of care metrics and mitigate negative outcomes.
Methods: Software and workflows to support the ACT were refined through iterative usability tests, and a retrospective cohort of surgical patients was used to develop and validate machine learning algorithms for postoperative death, acute kidney injury, and respiratory failure. Finally, a pilot feasibility trial was conducted where adult surgical patients were randomized (clustered by operating room) to receive the ACT intervention or usual care. The primary outcomes were compliance with intraoperative temperature and blood glucose quality metrics.
Results: The software was refined seven times based on the usability results. Machine learning algorithms were successfully constructed for all three specified prediction targets. 27,704 patients were successfully enrolled in the feasibility trial. Data analysis is in progress.
Key Words: Telemedicine, Usability, Health information technology, Decision support

3. PURPOSE (Objectives of the study)

The purpose of this pilot study is to implement an Anesthesiology Control Tower (ACT), and will achieve the following: (i) develop, refine and validate forecasting algorithms for adverse outcomes; (ii) Assess the usability of an ACT for the operating suite; and (iii) assess whether the ACT improves clinician compliance with standards of care and surrogate measures of patient outcomes.

4. SCOPE (Background, Context, Settings, Participants, Incidence, Prevalence)

Although most patients who undergo elective surgery chose to do so in order to cure conditions or improve quality of life, approximately 1 in 50 surgical patients will die within one month of their operation. The incidence of major morbidity (e.g. myocardial infarction, stroke, renal failure) is even higher. Many factors contribute to this morbidity and mortality, some of which are not easily addressed (e.g. preexisting comorbid conditions, patient frailty, invasiveness of surgery). In contrast, there are modifiable intraoperative factors (e.g. late administration of antibiotics, hypothermia, hemodynamic instability) that are likely to be associated with increased postoperative complications. Yet, no study has evaluated the potential of leveraging information technology (IT) to systematically address candidate quality of care metrics to mitigate negative outcomes. Therefore, we sought to develop an air-traffic control-like command center for the operating suite and to conduct a pilot randomized trial to determine the feasibility of using such a command center to implement evidence-based approaches to modifiable perioperative risk factors.

The work took place at Barnes-Jewish Hospital, a 1,252-bed university-affiliated tertiary care facility in St. Louis, Missouri. On average, 125 surgeries take place every business day in the hospital's 48 ORs. Anesthetic care is provided by certified registered nurse anesthetists and by resident anesthesiologists under the direction of attending anesthesiologists.

A computer-based anesthesia information management system is in place at Barnes-Jewish Hospital—the hospital transitioned from MetaVision (iMDsoft, Needham, MA) to Epic EMR (Epic, Verona, WI) during the study period. Prior to initiation of this work, the FDA-approved AlertWatch® software (AlertWatch, LLC, Ann Arbor, MI) was also available for clinical use in the hospital's ORs. This baseline AlertWatch® software made pre-programmed in-room alerts available to anesthesiology clinicians who accessed the software on a computer in the OR.

5. METHODS (Study Design, Data Sources/Collection, Interventions, Measures, Limitations)

Aim 1: Develop, refine, and validate forecasting algorithms for adverse outcomes.

To achieve this aim, we first created a large database that incorporated both discrete and time series data from approximately 110,000 adult (age ≥ 18) patients who underwent surgery with anesthesia and Barnes-Jewish Hospital between June 2012 and August 2016. Data were integrated in SQL format from four separate electronic health data repositories: MetaVision (discrete and time series from the preoperative and intraoperative period), COMPASS (discrete and time series from the postoperative period), Center for Biomedical Informatics (discrete data from the preoperative period and discrete outcomes data from the postoperative period) and SATISFY-SOS (patient reported outcomes from the early and intermediate term postoperative period). This was a substantial undertaking, and we encountered some specific challenges. The source databases had different structures and different naming conventions for their included items. Integrating these required time and expertise, and quality assessment to ensure the fidelity of the resulting merged database.

Once the database had been built, we constructed novel machine learning algorithms to predict postoperative 30-day mortality, postoperative acute renal failure, and postoperative acute respiratory failure. The large database was randomly divided into training, validation, and testing sets in a ratio of approximately 7:1:2. Input features for each model included patient demographic characteristics, comorbid conditions, preoperative vital signs, selected preoperative laboratory values, intraoperative time series (e.g., vital signs, ventilator readings), and selected intraoperative medications (e.g., vasopressors, sedative agents, fluids, blood products). Because each of the prediction targets was a relatively rare outcome, upsampling of positive cases and inverse weighting of cases were used to address the imbalanced nature of the dataset.

To predict death, we constructed a multi-path convolutional neural network that extracted information from the raw time series features using a series of convolution blocks and using long short-term memory units. These extracted features were then concatenated with the discrete input features and entered into a neural network with a series of fully connected hidden layers and a softmax output layer. The validation set was used to identify optimal values of the model hyperparameters. Model performance was quantified in the testing set using area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPRC). Case-wise feature importance was quantified using a back-propagation-based method.

To predict acute kidney injury and respiratory failure, we designed a novel model architecture that we called a factored generalized additive model. The motivation behind this design was to provide interpretable, actionable information about how the predicted risk of the adverse outcome would change if each input feature were to take on a different value. In this model,

time-varying intraoperative features were transformed using deep and narrow neural networks. These transformed features were entered into a logistic regression where the feature weights were functions of the static patient characteristics. Model performance was quantified in the testing set using AUROC and AUPRC.

Aim 2: Assess the usability of an ACT for the operating suite

To achieve this aim, three phases of usability testing were conducted in series over the course of five months. The first two phases centered on how clinicians working in the ACT interacted with the AlertWatch® software and the ACT workflow. The third phrase centered on the clinicians in the OR who would be receiving communication from the ACT.

In phase one, attending anesthesiologists and resident anesthesiologists participated in a moderated 20-minute session in the ACT. The goal was to identify major surface-level usability problems with the software, equipment, and supporting documentation. The participant used the ACT workstation to load AlertWatch® and the hospital's electronic health record programs. Then they addressed as many AlertWatch® alerts as they could while voicing their thoughts and actions aloud. A structured debriefing was held at the end of each session. In addition, the participant completed a survey containing the System Usability Scale (SUS), Computer System Usability Questionnaire (CSUQ), and NASA Task Load Index (NASA-TLX).

In phase two, attending anesthesiologists and resident anesthesiologists spent the full day reviewing alerts in the ACT without researcher supervision, mimicking the intended workflow of the ACT. Participant responses to alerts were logged automatically by the software. The participants did not communicate with the clinicians in the OR. Participants completed the SUS, CSUQ, and NASA-TLX after participating.

In phase three, attending anesthesiologists, resident anesthesiologists, and certified registered nurse anesthetists participated in interviews. The participant was presented with six scenarios where they were to imagine being the OR anesthesia clinician for the case and receiving a specified message from the ACT. The research team documented the participant's spontaneous responses and also asked open-ended questions.

Aim 3: Assess whether the ACT improves clinician compliance with standards of care and surrogate measures of patient outcomes.

To achieve this aim, a pilot randomized trial was carried out in 48 ORs at Barnes-Jewish Hospital (BJH, South Campus), St. Louis, MO between 4/1/2017 and 6/30/2019. On every business day, each OR was randomized to ACT intervention or to usual care. In effect, this means that patients were cluster-randomized, with the OR serving as the cluster. Patients under age 18 or with > 50% of their surgery outside the ACT's hours of operation were excluded. Patients were included via a waiver of informed consent.

In the ACT intervention group, clinicians in the ACT viewed automated alerts generated by the AlertWatch® software throughout surgery. These alerts corresponded to either lapses in monitoring/documentation (e.g., no blood pressure documented, blood glucose not checked in a diabetic patient), physiologic derangements (e.g., prolonged hypotension, low concentration of inhaled anesthetic agent), or deviations from standard practice (e.g., antibiotic has not been redosed). The clinicians in the ACT evaluated each alert to determine its relevance and importance. If they felt the alert was important, they communicated the alert to the patient's

attending anesthesiologist. Clinicians were not forced to follow any recommendations sent by the ACT. In the usual care group, clinicians in the ACT did not communicate with the patient's attending anesthesiologist.

The primary outcomes of this trial were intraoperative compliance with temperature and blood glucose management metrics. Compliance with the temperature metric was defined as the proportion of cases with a final recorded intraoperative core temperature > 36°C. Compliance with the blood glucose metric was defined as the proportion of cases with blood glucose < 180 mg/dl upon arrival to the post-anesthesia care unit.

Secondary outcomes included intraoperative process measures and postoperative surrogate measures. Process measures included time spent with mean arterial pressure < 60 mmHg, proportion of cases lasting greater than 1 hour with documented temperature, proportion of procedures with appropriate administration of repeat doses of antibiotics, proportion of cases where blood glucose > 180 mg/dl was treated with insulin, proportion of cases where blood glucose was measured intraoperatively if the case lasted > 1 hour (type 1 diabetics) or > 2 hours (type 2 diabetes), proportion of cases where train of four was documented prior to extubation if a nondepolarizing neuromuscular blocking agent was administered, proportion of cases with median tidal volume < 10 ml/kg ideal body weight, and mean fresh gas flow rate for cases volatile anesthetic use for > 80% of case duration. Postoperative surrogate measures included acute kidney injury, atrial fibrillation, respiratory failure, delirium, awareness with recall, surgical site infection, 30-day readmission, and 30-day mortality.

6. RESULTS (Principal Findings, Outcomes, Discussion, Conclusions, Significance, Implications)

Aim 1: Develop, refine, and validate forecasting algorithms for adverse outcomes.

Machine learning algorithms for postoperative death, acute kidney injury, and acute respiratory failure were successfully developed and validated using the database of approximately 110,000 surgical patients.

The incidence of postoperative death was 2.4%. In the testing set, the multi-path convolutional neural network model predicted postoperative death with AUROC of 0.91 (95% CI 0.90-0.92) and AUPRC of 0.32 (95% CI 0.28-0.37). The AUPRC compares favorably to the incidence of death. Performance was also excellent in several comparator models, including a dense neural network, random forest, support vector machine, and linear regression. These results have been published in the *British Journal of Anaesthesia* (Fritz, et al., 2019).

The incidence of postoperative acute kidney injury was 6.1%. In the testing set, the factored generalized additive model predicted postoperative acute kidney injury with AUROC of 0.82 (95% CI 0.81-0.84) and AUPRC of 0.26 (95% CI 0.26-0.28). The incidence of postoperative acute respiratory failure was 3.1%. In the testing set, the factored generalized additive model predicted postoperative respiratory failure with AUROC of 0.72 (95% CI 0.66-0.78) and AUPRC of 0.11 (95% CI 0.09-0.13). These results have been presented at the American Medical Informatics Association annual symposium and published in the conference proceedings (Cui, et al., 2020).

These results demonstrate that our algorithms can predict postoperative adverse outcomes with a high degree of accuracy. Making these accurate predictive algorithms available to clinicians in

the ACT may enable those clinicians to spend a greater proportion of their time actively monitoring patients who are at the highest risk for postoperative adverse outcomes. Increasing the proportion of their time spent monitoring the highest risk patients gives the clinicians in the ACT more time to identify ways to mitigate the risk for these patients. This chain of events would maximize the likelihood that the ACT intervention can reduced postoperative adverse outcomes.

To facilitate investigation of this potential benefit, we have made it a priority to make these machine learning predictions available to the clinicians working in the ACT. To date, we have established a near-live stream of data from the electronic health record onto the server where the machine learning calculations can be performed. This involved several logistical challenges. As next steps, we will employ a user-centered design framework to create a display interface for showing the predictive algorithm output to the clinicians in the ACT.

Aim 2: Assess the usability of an ACT for the operating suite

In phase one, 8 attending anesthesiologists and 7 resident anesthesiologists participated. During these 20-minute sessions, participants evaluated an average of 11.5 alerts from 7.25 patients. A total of 155 usability problems were identified, most related to software functionality and alert content. The AlertWatch® software was iteratively updated seven times based on feedback from these sessions. Most of the changes related to the visual display as well as the content and prioritization of the alerts. With these changes, the fraction of software-generated alerts that were felt to be clinically significant or potentially significant by the participants increased from 27% in the first iteration to 56-73% in later iterations.

In phase two, 6 attending anesthesiologists and 8 resident anesthesiologists participated. During these daylong sessions, participants evaluated an average of 176 alerts from 54.9 patients. Across the first two phases, satisfaction scores measured by the SUS were higher and workload scores measured by the NASA-TLX were lower among resident anesthesiologists than among attending anesthesiologists. In addition, both scores improved when a participant returned to the ACT for a repeat testing session, as compared to the initial session.

In phase three, 4 attending anesthesiologists and 6 certified registered nurse anesthetists participated. The interviews provided a wealth of information on the perceived barriers and facilitators to the implementation of the ACT and enabled us to make alterations to our intervention prior to the conducting the pilot randomized trial. Some participants felt the ACT might limit their autonomy as anesthesia clinicians or be redundant with the care already being provided at the bedside. Others were concerned that the usefulness of alert communications would be limited if they were poorly timed and distracted the clinician from meaningful patient care tasks. However, all participants were able to identify specific instances when they perceived the ACT could be useful, such as when an anesthesiologist is covering multiple busy ORs.

Through our thorough and iterative usability testing process and stakeholder assessment of barriers and facilitators, we have enhanced the acceptability of our novel intraoperative telemedicine intervention. This accomplishment directly contributed to our ability to incorporate the ACT into the pilot randomized trial in aim 3. In the longer term, it will also improve our ability to implement this innovation in routine practice outside the experimental setting.

Aim 3: Assess whether the ACT improves clinician compliance with standards of care and surrogate measures of patient outcomes.

A total of 27,704 patients were enrolled in the pilot feasibility trial. Of these, 13,572 patients were randomized to the ACT intervention and 14,132 patients were randomized to usual care. We are currently in the process of finalizing outcomes data for this large patient population. We plan to perform data analysis and disseminate the findings via publications and conference presentations. The results of this aim will be important because an intraoperative telemedicine intervention must result in improvements in one or more intraoperative process measures if any effect on postoperative outcomes is to be expected.

By successfully completing enrollment in the pilot trial, we have demonstrated the feasibility of operating the ACT and of using the ACT as the basis for a randomized trial. Because the clinicians in the ACT regularly communicate with anesthesiologists in the OR, the anesthesiologists in the OR have begun to view the clinicians in the ACT as valuable collaborators. Such cultural acceptance of the ACT by the OR team is necessary for the ACT intervention to have any impact on process measures or patient clinical outcomes. Because we have achieved this cultural acceptance, our statistical analysis for this aim will yield informative results.

Unfortunately, we have not yet managed to complete Aim 3. There are two key reasons for this. The first is that we switched electronic health records and the second has been the COVID-19 pandemic. Most of the difficulty that we have had is related specifically to the transition from the electronic health records MetaVision and Compass to the EPIC electronic health record, as well as data retrieval delays related to COVID-19; our informatics team had to divert their effort to pandemic related activities. Last June (2019) we planned to extract all EPIC era data from the AlertWatch Server (a server that aggregates our data from different sources). We had to wait until our hospital migrated their servers to a new platform, which was delayed numerous times and was out of our control. The migration finally occurred right before COVID-19 hit, which tied up our database analysts since then, preventing our data extraction. We were eventually able to extract the EPIC era data directly from Hyperspace, but this was a challenging task, and it wasn't until this June (2020) that our database expert was successful in this regard. At this point we now have all the EPIC data we need together with the data we had already acquired prior to the EHR migration. Thus, we are now embarking on the analyses specified under Aim 3, and hope to have all aspect of the study completed within the next few months. We regret the delays to completing the study, and are grateful for the support and understanding from our colleagues at the AHRQ. We remain optimistic that we will complete all the Aims specified, and that our R21 study will provide value and insights. As soon as we have analyzed the data, we will provide a further update.

Overall Conclusion

We have completed two of the three aims of this project, and we have laid the foundational groundwork for future investigations of the effect of intraoperative telemedicine interventions on postoperative clinical outcomes. We regret that we have been unable to complete Aim 3, but as we have outlined above, we have a clear plan for its completion. However, we have established the technological infrastructure and departmental culture to facilitate the ACT concept. We have maximized the usability of the physiologic alerts through usability tests, and we have developed machine learning algorithms that will further enhance intraoperative decision making once they are fully integrated into the ACT. We have also demonstrated the feasibility of efficiently enrolling large numbers of patients into a randomized trial using the ACT as an intervention. These accomplishments uniquely position our team to investigate the effect of the ACT on

postoperative clinical outcomes. Importantly, the success of this AHRQ-funded R21 project has provided the necessary foundation for a 5-year NINR-funded R01. The primary clinical outcome of our AHRQ-funded study was surrogate markers of quality of care. The next logical step is a much larger study focused on clinically relevant outcomes. Thus, we have now successfully embarked on the Telemedicine Control Tower for the OR: Navigating Information, Care and Safety (TECTONICS) trial (NCT03923699). This exciting and burgeoning research program would not have been possible without the foundational support for the ACTFAST studies, provided to Washington University by the AHRQ.

7. LIST OF PUBLICATIONS and PRODUCTS

1. Murray-Torres TM, Wallace F, Bollini M, Avidan MS, Politi MC. Anesthesiology Control Tower: Feasibility Assessment to Support Translation (ACT-FAST)—a feasibility study protocol. Pilot Feasibility Study. 2018;4:38. PMID 29416871. Doi: 10.1186/s40814-018-0233-4

2. Gregory S, Murray-Torres TM, Fritz BA, Ben Abdallah A, Helsten DL, Wildes TS, Sharma A, Avidan MS. Study protocol for the Anesthesiology Control Tower-Feedback Alerts to Supplement Treatments (ACTFAST-3) trial: A pilot randomized controlled trial in intraoperative telemedicine. F1000Research. 2018;7:623. PMID 30026931. Doi: 10.12688/f1000research.14897.2

3. Fritz BA, Chen Y, Murray-Torres TM, Gregory S, Ben Abdallah A, Kronzer A, McKinnon SL, Budelier T, Helsten DL, Wildes TS, Sharma A, Avidan MS. Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. BMJ Open. 2018;8:e020124. PMID 29643160. Doi: 10.1136/bmjopen-2017-020124

4. Murray-Torres TM, Casarella A, Bollini M, Wallace F, Avidan MS, Politi MC. Anesthesiology Control Tower—Feasibility Assessment to Support Translation (ACTFAST): Mixed-methods study of a novel telemedicine-based support system for the operating room. JMIR Hum Factors 2019;6:e12155. PMID 31012859. Doi: 10.2196/12155

5. Fritz BA, Cui Z, Zhang M, He Y, Chen Y, Kronzer A, Ben Abdallah A, King CR, Avidan MS. Deep-learning model for predicting 30-day postoperative mortality. Br J Anaesth. 2019;123:688-95. PMID 31558311. Doi: 10.1016/j.bja.2019.07.025

6. Cui Z, Fritz BA, King CR, Avidan MS, Chen Y. A factored generalized additive model for clinical decision support in the operating room. AMIA Annu Symp Proc. 2020 Mar 4;2019:343-52. PMID 32308827.

7. Fritz BA, Abdelhack M, King CR, Chen Y, Avidan MS. Update to 'Deep-learning model for predicting 30-day mortality' (Br J Anaesth 2019; 123: 688-95). Br J Anaesth. Published online May 7, 2020. PMID 32389391. Doi: 10.1016/j.bja.2020.04.010