BMJ Open

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients: Protocol for a Retrospective Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031988
Article Type:	Protocol
Date Submitted by the Author:	28-May-2019
Complete List of Authors:	Cannesson, Maxime; UCLA, Anesthesiology Hofer, Ira; UCLA, Anesthesiology Rinehart, Joseph; UC Irvine, Anesthesiology Lee, Christine; UC Irvine, Bioengineering Subramaniam, Kathirvel; UPMC, Anesthesiology Baldi, Pierre; UC Irvine, Computer Sciences Dubrawski, Artur; Carnegie Mellon University, Computer Sciences Pinsky , MR; University of Pittsburgh School of Medicine,
Keywords:	ANAESTHETICS, BIOTECHNOLOGY & BIOINFORMATICS, CLINICAL PHYSIOLOGY, INTENSIVE & CRITICAL CARE

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients: Protocol for a Retrospective Study

Maxime Cannesson[1] MD PhD, Ira Hofer[1] MD, Joseph Rinehart[2] MD, Pierre Baldi[3] PhD, Christine Lee[4] PhD, Kathirvel Subramanian[5] MD, Artur Dubrawski[6] PhD, Michael R. Pinsky[7] MD PhD

Email for all authors: Maxime Cannesson: mcannesson@mednet.ucla.edu; Ira Hofer: IHofer@mednet.ucla.edu; Joseph Rinehart: jrinehar@uci.edu; Christine Lee: cklee219@gmail.com; Kathirvel Subramanian: subramaniamk@upmc.edu; Pierre Baldi: pfbaldi@ics.uci.edu; Artur Dubrawski: awd@cs.cmu.edu; Michael R. Pinsky: pinsky@pitt.edu.

Corresponding author: Maxime Cannesson, Department of Anesthesiology and Perioperative Medicine, UCLA David Geffen School of Medicine, 757 Westwood Plaza, Los Angeles CA 90095; +1-310-657-553, mcannesson@mednet.ucla.edu

From: 1. Department of Anesthesiology and Perioperative Medicine, UCLA David Geffen School of Medicine, 2. Department of Anesthesiology and Perioperative Care, UC Irvine Health, 3. Department of Computer Sciences, UC Irvine, 4. Department of Bioengineering, UC Irvine, 5. Department of Anesthesiology, University of Pittsburgh Medical Center, 6. Artur Dubrawski, Auton Lab, School of Computer Science, Carnegie Mellon University, 7. Department of Critical Care Medicine, University of Pittsburgh Medical Center.

Keywords: Machine learning, surgery, safety, haemodynamics, blood pressure, physiology

Word Count: 3,215

Abstract

Introduction: About 42 million surgeries are performed annually in the US. While the estimated postoperative mortality is less than 2%, 12% of all patients in the high-risk surgery group account for 80% of postoperative deaths. New onset of hemodynamic instability is common in patients undergoing surgery and its delayed treatment leads to increased morbidity, mortality and use of resources in surgical patients. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on clinical data and high-fidelity physiological waveforms to predict hemodynamic instability during surgery.

Methods and analysis: We propose to initiate our work using an already existing annotated intraoperative database (UCI) including clinical and high-fidelity waveform data. This data will be used for the initial training and development of the ML model (Carnegie Mellon University) that will then be tested on prospectively collected UCLA database. Simultaneously, we will use existing knowledge of hemodynamic instability patterns derived from our ICU cohorts, MIMIC II data, UCI data, and animal studies to create smart alarms and graphic user interface for a clinical decision support. Using machine learning analytics, we will extract a core dataset which best characterizes the signatures of normal intraoperative variability, various intraoperative hemodynamic instability aetiologies, and variable responses to intraoperative resuscitation. We will then employ clinician-driven iterative design to create a novel clinical decision support user interface, and evaluate its effect in simulated intraoperative high-risk surgeries.

Ethics and dissemination: Once the investigation has been completed, we intend to publish the results in a peer-reviewed publication. We also intend to present the results of this work at professional conferences for both the anaesthesiology and computer science communities. In accordance with the recent proposal from the International Committee of Medical Journal Editors, patient-level data will be made available within 6 months after publication of the primary manuscript.

Article Summary:

Strengths and limitations of this study: Our innovations are:

1) To impact surgical practice by better and earlier identifying patients at greatest risk for CRI and by providing point-of-care data-driven explanations and process-specific resuscitation using real-time data input and point-of-care management, potentially decreasing preventable surgical morbidity and mortality.

2) Being able to adjust for placement, level of monitoring needed, and pre-emptive therapies and response to therapy enabling personalized medicine. It will create a sensitive and specific means to predict which patients may or may not ever develop CRI, which has important implications for patient safety, surveillance, triage and care: needed frequency of monitoring, case load mixture, workload, staff allocation, patient triage to monitored or non-monitored units, higher cost vs. lower cost bed allocation, prevention of adverse events.

3) Few innovations have been introduced to improve technologic patient surveillance and management in decades. The modelling methods we propose to develop should drastically shift intraoperative paradigms and treatment protocols and they will be applicable to existing monitoring modalities.

4) Our ML analytics have potential to identify new monitoring parameters to improve prediction of instability and, in our exploratory specific aim, to reverse engineer our understanding of disease aetiology during surgery.

5) More than 42 million Americans undergo surgery each year and even though the perioperative complication rate is low, the absolute numbers are large. Although we will focus on high-risk surgery patients, we will collect data on the whole surgical population such that we will also develop a novel database on low-risk patients. They are an opportunity cohort to explore shared risk to compare with the high-risk patients. The main limitation of this study is that it is a retrospective study.

Introduction

About 42 million surgeries are performed annually in the US[1 2]. While the estimated postoperative mortality is less than 2%, 12% of all patients in the high-risk surgery group account for 80% of postoperative deaths[3 4]. To assist in guiding clinical decisions and prioritization of care, several perioperative clinical risk scores have been proposed[5-7]. The goal of these scores is to help planning clinical management and allocating resources to avoid postoperative complication and death. Recently, US hospitals have adopted electronic health report (EHR) documentation of patient care. Still, interoperability of these EHR systems remains an open issue, leading to challenges in data integration. As a result, the potential that hospital data offers in terms of understanding and improving care has not been realized. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on EHR data and high-fidelity physiological waveforms to predict cardiorespiratory instability (CRI). New onset of CRI is common in patients undergoing surgery. Delayed treatment of CRI leads to increased morbidity, mortality and use of resources in surgical patients. Even short periods of intraoperative hypotension (mean arterial pressure (MAP) < 65 mmHg) have been linked to postoperative complications such as myocardial infarction and kidney failure[8 9], and mortality[10]. Although anaesthesiologists can rescue patients with CRI - decreasing the incidence of cardiac arrests and inpatient mortality[11-13] - a more proactive approach would be to enable anaesthesiologists and nurse anaesthetists to recognize impending CRI before it happens. Specific Aim 1 (SA1). Development of the ML test set (University of California, Irvine (UCI)) and retrospective validation (UC Los Angles (UCLA)) in high-risk surgery patients to identify subsequent CRI. Using ML analytics, we will extract a core dataset which best characterizes the signatures of normal intraoperative variability, various intraoperative CRI aetiologies, and variable responses to intraoperative resuscitation using 1) existing high-granularity physiologic and clinical record data to define the structure of the database; 2) high-granularity intraoperative data from high-risk surgery patients for prospective input and further model development. Since

our approaches use agnostic prioritization of physiological signals, we will explore which processes underlie specific signatures. This reverse-engineering approach will give insight into cardiorespiratory homeostasis and intraoperative CRI.

Specific Aim 2 (SA2). Prospective Validation of a CDS tool. We will employ clinician-driven iterative design to create a novel CDS user interface, and evaluate its effect in simulated intraoperative high-risk surgeries.

Methods and Analysis

Study design

We propose to initiate our work using an already existing annotated intraoperative database (UCI) including EHR and high-fidelity waveform data. This data will be used for the initial training and development of the ML model that will then be tested on prospectively collected UCLA database (SA1). Simultaneously, we will use existing knowledge of CRI patterns derived from our SDU / ICU cohorts, MIMIC II data, UCI data, and animal studies (University of Pittsburgh) to create smart alarms and graphic user interface (GUI) for CDS (SA2).

Patient population and Data Acquisition

The UCI dataset is based on the Bernoulli® data collection system (Cardiopulmonary Corp., New Haven, CT), and as of February 2019 it includes high-fidelity physiological waveforms and EHR[14] data on more than 35,000 patients. All waveform and clinical data from surgical patients at UCI have been collected for research purposes since November 2015. The total UCI data collection as of February 2019 consists of >120,000 monitoring hours of waveform and clinical data, or >4,000 GBs of data. All waveform data is collected off of surgical patient monitors with the Bernoulli® software and equipment, and de-identified and synced retrospectively with clinical data (Figure 1). The sampling rates for EKG, Plethysmographic and Arterial waveforms are 300 Hz, 100 Hz, and 120 Hz. Clinical data is extracted from intraoperative EHR (Surgical Information Systems (Alpharetta, GA) and Epic (Verona, WI)) and synced with waveform data. These data are then linked to monitoring and clinical annotations, where adverse events are documented. At UCLA, we have established a perioperative data warehouse including all the EHR data and we plan to install an intraoperative data collection system similar to Bernoulli® for waveform collection. The EHR data at UCLA have already been analysed as proof of concepts[14 15]. The Bernoulli® and similar software (Bedmaster) provides a mission-critical application layer designed for multi-parameter data abstraction, fusion, remediation, time-synchronization, and

real-time processing. Bernoulli® provides an extensive distribution layer designed for export to third-party applications and EHR providers via HL7 or custom protocols.

Development thrusts.

Identification of potentially informative bio-signatures. In support of SA1 we plan to maximize the chances for ML-developed feature extraction methods (time-window-gated statistics such as averages, variances, entropies and trends of univariate time series, cross-correlations between pairs of series, or spectral decompositions followed by compression with e.g. PCA) with massive-scale comprehensive searches for change points in signal to add information useful in detecting and characterizing CRI. We have successfully applied such methods to problems involving large amounts of multivariate time series data[16]. Next, a ML algorithm uses a sample of annotated training data to identify empirically a subset of those change points that bring predictive value to models of CRI (supervised learning). In addition, we will take our quest for precursors of CRI outside of the usual single-signal or joint multi-signal modelling, by learning structures of multivariate correlations between pairs of signals, and tracking them over time. A novel use of Canonical Correlation Analysis (CCA) will be the first approach to try in this context, and we will extend it to adopt temporal regularization constraints to enable smooth transitions between consecutive timeframes. Our novel approach will let us discover pairs of combinations of features, extracted independently from two VS time series that highly correlate with each other. CCA identifies pairs of such “principal components” that could be learned from the same null space data as shown in the context of the PCA-based approach. If certain pairs of components are found to correlate consistently during stability but lose correlation at or before the onset of a CRI (or vice versa), they boost detectability of our target events. We will perform extensive searches for such CCA components and add them to the pool of factors worth tracking. We will use regularization and feature ablation to maintain parsimony of the resulting models, and to identify features of data that have key contributory effects on performance. We

expect to identify such features, which would not be visible to alternatives (either using many independent single-stream models or a single (fully combined) joint model for null-space PCA-like modelling of baseline variance, or via bivariate cross-correlations of pairs of time series).

Learning predictive models of instability. We will use predictive ML models trained on the UCI annotated data to empirically identify the parsimonious set of predictors of the targeted pathologies, selected from features extracted as described above. In one instance of our preliminary work, primary observables (e.g. EKG signals, blood pressure, central venous and pulmonary arterial pressure, and pulse oximetry waveforms) are processed to generate beat-to-beat HR, as well as the various diastolic and systolic pressure and oximetry signals. This feature become input for Random Forest regression model that learns to predict the time since start of bleed (equivalent to the amount of blood lost given fixed bleed rate in the referred experiment), and in the process identifies a subset of all features that jointly yields optimal performance. We also create 1 and 5 minute moving averages of these derived variables, as well as often reported measures of heart rate variability along several domains (time and frequency domains, non-linear measures)[17]. The time windows determining the features and output classes will vary according to aims. In SA1, we will enrich the development set to include a significant proportion of time windows from patients in CRI, thus the model will predict either one of the CRI states, or absence of CRI. In SA1, models predictive of fluid resuscitation will include patients in CRI and the output (fluid responsiveness) is binary. We will use the AMOC analysis as described above to validate models' sensitivity and specificity. To take our CDS beyond capability to detect intraoperative bleeding at a single specific rate, we will collect human data, retrospectively annotated by OR and ICU physicians for CRI events and apply Bayesian Aggregation (BA) method[18]. It accumulates evidence from subsequent measurements and tracks multiple hypotheses as their posterior probability distribution evolves while new evidence becomes available. BA will be our foundational approach to characterization of detected and predicted events of interest whenever low signal to noise ratios in the available

data would yield more direct regressive models impractical. In particular, it will not only detect functional hypovolemia, but also estimate the rate of intravascular fluid loss.

Explanatory analysis. For our exploratory aim we will perform explanatory analysis of the learned predictive model asking what physiologic control qualities of the primary predictors of the models best explain their predictive power. In this way, we will explore foundations of autoregulation, adaptation and failure in the context of CRI. In ML experiments we will favour the types of models that avail explainability of generated results. We had success applying Random Forest classifiers and regressors in applications that require high level of user interaction and extensive explanations of predictions made[19]. We have also tried our RIPR algorithm[20] to support interactive adjudication of alerts as either true episodes of instability or artifacts[21]. It relies on point estimators for conditional entropy and recovers a desirably small set of projections of data which accurately classify test alerts while remaining intuitive to humans. New alerts can be adjudicated using one of the projections from the retrieved set. We will extend this approach to systematically include semi-supervised and active learning concepts to support semi-automated annotation of large-scale data sets given limited availability of qualified human experts. We have also shown combined predictive and explanatory utility of learning temporal association rules from asynchronous sequences of discrete events and continuous signals with TITARL algorithm. We have shown that it can be used to identify which of the potentially large number of patterns detected in data coincide or precede particular events of interest, and present the results in a readable and interpretable form of manageably simple logical statements. We will extend this approach so that the most predictive combinations of patterns and states that can asynchronously appear in multivariate clinical data, irrespective of the temporal resolution of their observation (waveforms, beat to beat, breath to breath, disparate clinical records and demographic data), or missingness of data (frequent in hemodynamic monitoring of human patients), will be revealed, validated by expert clinicians, and used to support predictive models.

Creation of a prototype CDS system bedside GUI showing real-time probability of impending instability, lead-time to event and determining factors (SA2).

Primary model development and initial simulation testing will be done using the initial UCI and UCLA databases, and then validated at UCLA simulator centre. Then the UCLA and UPMC datasets will serve as an external validation set for our predictive models in an iterative fashion. First to predict CRI, then to diagnose specific aetiologies, and then to guide resuscitation. Thus, our primary goal is to create a robust and generalizable approach that will readily expand across other healthcare information systems. We will finalize the ML analysis of the prospectively acquired high-risk surgery patient cohort dataset in year 5. We anticipate creating final prototype GUI incorporating both measures of present instability and our predictive model with descriptions as to the potential processes that will be causing instability. We will use clinician-driven iterative design to refine the GUI.

SimStage 1. We will enlist 12 experts (6 nurse anaesthetists and 6 physician anaesthesiologists) to provide feedback. The HFHS will be set to simulate the OR and using the prototype GUIs we have and will develop. The feedback sessions will be conducted in 2 groups of 6 clinicians to maximize variety in the input while benefitting from group dynamics. We will seek feedback regarding GUI (1) completeness or redundancy of content, (2) ease of interpretation, and (3) ergonomics. This feedback evaluation will be iterative, with a new generation of the GUI, which takes into consideration input from each group, to be rolled out sequentially. We anticipate 3-5 GUI iterations before subsequent larger group testing.

SimStage 2. We will move to larger-scale clinician-driven iterative design through evaluation of simulated clinical GUI use. In this stage the HFHS will be set as an OR and based on 10 scenarios. The patient's monitors will live stream the data collected in C.1 for 10 selected patient cases; clinicians will also have access to the other de-identified case data such as past medical history, laboratory and diagnostic test results, current medications, etc. Of the 10

simulated cases, 8 will show evidence of instability across a 2-hr interval. Two for volume loss, two for hemodynamically unstable arrhythmia, two for post-abdominal insufflation hemodynamic instability, two for post anaesthesia induction hypotension, while two will remain stable with half the patients monitored according to current standard practice, while half will use the additional GUI. Each scenario will be run by one anaesthesiologist and subsequently by one nurse anaesthetist. The clinicians will document patient assessment and care directives. We will then evaluate each scenario for effectiveness based on accuracy of diagnosis, time to correct diagnosis, accuracy of intervention choices, and time to intervention. These scenarios will be repeated four times, using different clinicians (i.e. total of 40 clinicians, half MD and half nurses). All clinicians will be debriefed to gain their feedback as in SimStage 1.

SimStage 3. Then we conduct a midterm analysis of effectiveness results that we will share with our original team of experts from SimStage 1, and make iterative design adjustments to the GUI, the alerting system, or the action algorithms based on SimStage 2. We will then proceed to SimStage 3, another simulated trial with the same simulated settings, scenarios and equipment, but 40 different volunteers (so as to avoid Hawthorne effect). We will collect the same information, and evaluate if the effectiveness indicators from SimStage 2 are improved upon in SimStage 3. The final refined design represents the deliverable CDS prototype for this project, and to estimate effect sizes for a future clinical trial.

Discussion

Implications and future directions

If one could accurately predict who, when and why patients develop cardiorespiratory instability (CRI) during surgery, then effective pre-emptive treatments could be given to improve postoperative outcome and more effectively use healthcare resources. But signs of shock often occur late once organ injury is already present. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on electronic health record (EHR) data and high-fidelity physiological waveforms to predict CRI and make the databases of intraoperative data and waveforms used for these developments freely accessible. This is extremely relevant because although 5.7 million Americans are admitted to an Intensive Care Units (ICU) in one year, more than 42 million undergo surgery annually. Previous and ongoing studies conducted in the ICU and in the step-down unit have built the architecture to collect real-time high-fidelity physiological waveform data streams and integrate them with patient demographics from the EHR to build large data sets, and derive actionable fused parameters based on machine learning (ML) analytics as well as display information in real time at the bedside to drive clinical decision support (CDS) in the critical care setting. The goal of this proposal is to apply these ML approaches to the complex and time compressed environment of high-risk surgery where greater patient and disease variability exist and shorter period of time is available to deliver truly personalized medicine approaches.

Strengths and Limitations

We will leverage our previous work and NIH/R01 funded projects in the SDU/ICU using similar methodologies to characterize CRI during surgery (R01-GM117622 and R01-NR013912). Our innovations are: 1) To impact surgical practice by better and earlier identifying patients at greatest risk for CRI and by providing point-of-care data-driven explanations and process-specific resuscitation using real-time data input and point-of-care management, potentially

decreasing preventable surgical morbidity and mortality. 2) Being able to adjust for placement, level of monitoring needed, and pre-emptive therapies and response to therapy enabling personalized medicine. It will create a sensitive and specific means to predict which patients may or may not ever develop CRI, which has important implications for patient safety, surveillance, triage and care: needed frequency of monitoring, case load mixture, workload, staff allocation, patient triage to monitored or non-monitored units, higher cost vs. lower cost bed allocation, prevention of adverse events. 3) Few innovations have been introduced to improve technologic patient surveillance and management in decades. The modelling methods we propose to develop should drastically shift intraoperative paradigms and treatment protocols and they will be applicable to existing monitoring modalities. 4) Our ML analytics also have potential to identify new monitoring parameters to improve prediction of instability and, in our exploratory specific aim, to reverse engineer our understanding of disease aetiology during surgery. 5) More than 42 million Americans undergo surgery each year and even though the perioperative complication rate is low, the absolute numbers are large. Although we will focus on high-risk surgery patients, we will collect data on the whole surgical population such that we will also develop a novel database on low-risk patients. They are an opportunity cohort to explore shared risk to compare with the high-risk patients.

Potential limitations: First, this proposal starts as a retrospective analysis of UCI data of patients with or without CRI. Limiting our variables to those available to us in retrospective analyses may limit the ultimate prediction compared to including more variables that we may find useful beyond this first pass. However, in our prospective UCLA and UPMC data collection interval, if specific parameters appear useful, we will prospectively use them in a non-protocol fashion in patients because we are those patients' providers. Second, the modelling of variables may not allow for discrimination of CRI caused by specific diagnoses, rather than by pathophysiologic processes. Still this would be an improvement over existing bedside monitoring analysis. Since we treat these surgical patients and CRI is a relatively uncommon event we will annotate

patients' records weekly as to CRI and specific aetiologies for retrospective analysis. Third, our medical centres are installing under our leadership high-density data collection systems (Bernoulli or Bedmaster) in all ORs. We have not used this data formatting/data synthesis platform before in the OR at UCLA and UPMC, but the system has been used at UCI and we have the expertise for its installation in SDU and ICU at UCLA and UPMC. We have planned for six months to reformat the data collection and secured query system as needed at UCLA and UPMC and budgeted for data collection and secured processing personnel for this support throughout the duration of the study. Based on the funding cycle criteria our system could be installed at UCLA (07/2018) before this project becomes active. Fourth, we planned for Honest Broker recording efforts for EHR review similar to ontology analysis already being done for another funded protocol (R01 NR013912). Accordingly, we have planned for a 6-month lead-in to insure she/he is cognizant of the UCLA and UPMC EHR idiosyncrasies.

Ethics and Dissemination

Once the investigation has been completed, we intend to publish the results in a peer-reviewed publication. We also intend to present the results of this work at professional conferences for both the anaesthesiology and computer science communities. In accordance with the recent proposal from the International Committee of Medical Journal Editors, patient-level data will be made available within 6 months after publication of the primary manuscript. Data will be provided to researchers who submit a methodologically sound research proposal including a protocol and statistical analysis plan. No patient identifying fields (including dates) will be included in the shared dataset. Age will be provided in years, unless the patient is older than 89 years. In this case, age will be reported as '>89 years.' Any dates will be presented as 'number of days since index surgery.'

Author contributions:
Maxime Cannesson: Draf the manuscript and final approval of the manuscript.
Ira Hofer: Draft the manuscript and final approval of the manuscript.
Joseph Rinehart: Draft the manuscript and final approval of the manuscript.
Pierre Baldi: Draft the manuscript and final approval of the manuscript.
Christine Lee: Draft the manuscript and final approval of the manuscript.
Kathirvel Subramanian: Draft the manuscript and final approval of the manuscript.
Artur Dubrawski: Draft the manuscript and final approval of the manuscript.
Michael R. Pinsky: Draft the manuscript and final approval of the manuscript.

Funding: This work is supported by NIH R01 HL144692.

Conflict of Interest: Dr. Cannesson is a consultant for Edwards Lifesciences and Masimo Corp, and has funded research from Edwards Lifesciences and Masimo Corp. He is also the founder of Sironis owns patents and receive royalties for closed loop hemodynamic management that have been licensed to Edwards Lifesciences. Dr. Cannesson's Department receives funding from the NIH (R01GM117622; R01 NR013012; U54HL119893; 1R01HL144692).
Dr. Rinehart is the founder of Sironis owns patents and receive royalties for closed loop hemodynamic management that have been licensed to Edwards Lifesciences.
Dr. Christine Lee is an engineer at Edwards Lifesciences.

Acknowledgements: None

Data Statement:

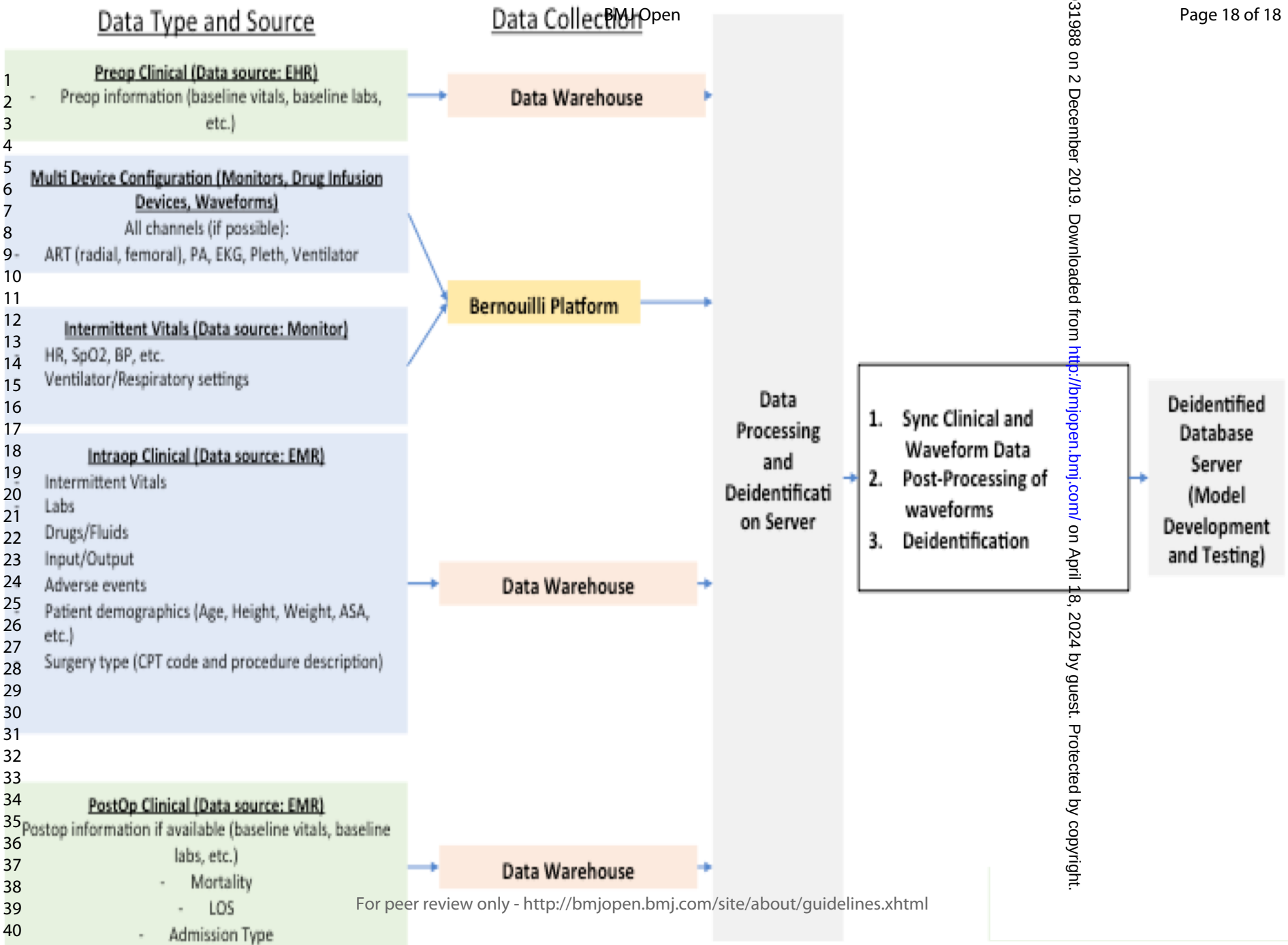
References

1. Weiser TG, Haynes AB, Molina G, et al. Size and distribution of the global volume of surgery in 2012. *Bull World Health Organ* 2016;94(3):201-09F. doi: 10.2471/BLT.15.159293
2. Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008;372(9633):139-44. doi: 10.1016/S0140-6736(08)60878-8
3. Pearse RM, Harrison DA, James P, et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. *Crit Care* 2006;10(3):R81.
4. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012;380(9847):1059-65. doi: 10.1016/S0140-6736(12)61148-9
5. Gawande AA, Kwaan MR, Regenbogen SE, et al. An Apgar score for surgery. *J Am Coll Surg* 2007;204(2):201-8. doi: 10.1016/j.jamcollsurg.2006.11.011
6. Le Manach Y, Collins G, Rodseth R, et al. Preoperative Score to Predict Postoperative Mortality (POSPOM). *Anesthesiology* 2016;124(3):570-79. doi: 10.1097/aln.0000000000000972
7. American Society of Anesthesiologists ASoAASoA. New classification of physical status. *Anesthesiology* 1963(24)
8. Salmasi V, Maheshwari K, Yang D, et al. Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. *Anesthesiology* 2017;126:47-65. doi: 10.1097/ALN.0000000000001432
9. Walsh M, Devereaux PJ, Garg AX, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. *Anesthesiology* 2013;119(3):507-15. doi: 10.1097/ALN.0b013e3182a10e26
10. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I, Devereaux PJ, Chan MT, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2012;307(21):2295-304. doi: 10.1001/jama.2012.5502
11. Winters BD, Pham JC, Hunt EA, et al. Rapid response systems: a systematic review. *Crit Care Med* 2007;35(5):1238-43. doi: 10.1097/01.CCM.0000262388.85669.68
12. Chan PS, Khalid A, Longmore LS, et al. Hospital-wide code rates and mortality before and after implementation of a rapid response team. *JAMA* 2008;300(21):2506-13. doi: 10.1001/jama.2008.715
13. Goodrich C. Endpoints of resuscitation: what should we be monitoring? *AACN Adv Crit Care* 2006;17(3):306-16.
14. Hofer IS, Gabel E, Pfeffer M, et al. A Systematic Approach to Creation of a Perioperative Data Warehouse. . *Anesth Analg* 2016;122:1880-4.
15. Gabel E, Hofer IS, Satou N, et al. Creation and Validation of an Automated Algorithm to Determine Postoperative Ventilator Requirements After Cardiac Surgery. *Anesth Analg* 2017;124(5):1423-30. doi: 10.1213/ANE.0000000000001997
16. Dubrawski A. Extracting Useful Information from Multivariate Temporal Data. In: Jay Liebovitz TF, ed. Big Data and Business Analytics2013.

17. Tandon P, Huggins P, Dubrawski A, et al. Source location via Bayesian aggregation of evidence with mobile sensor data. IEEE Symposium on Radiation Measurements and Applications. Oakland, CA, May 2012.
18. Donze A, Clermont G, Langmead CJ. Parameter synthesis in nonlinear dynamical systems: application to systems biology. *J Comput Biol* 2010;17(3):325-36. doi: 10.1089/cmb.2009.0172
19. Milbrandt EB, Clermont G, Martinez J, et al. Predicting late anemia in critical illness. *Crit Care* 2006;10(1):R39. doi: 10.1186/cc4847
20. Projection Retrieval for Classification, Neural Information Processing Systems. NIPS; December 2012; Lake Tahoe, NV.
21. Fiterau M, Dubrawski A, Chen L, et al. Automatic Identification of Artifacts in Monitoring Critically Ill Patients. *Intensive Care Med* 2013;39(2):S470.

Figure 1. Process, workflow, and architecture of the already existing UCI database.

ART: Arterial waveform, PA: Pulmonary artery, HR: Heart rate, BP: Blood pressure, LOS: Length of stay



BMJ Open

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients: Protocol for a Retrospective Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031988.R1
Article Type:	Protocol
Date Submitted by the Author:	20-Sep-2019
Complete List of Authors:	Cannesson, Maxime; UCLA, Anesthesiology Hofer, Ira; UCLA, Anesthesiology Rinehart, Joseph; UC Irvine, Anesthesiology Lee, Christine; UC Irvine, Bioengineering Subramaniam, Kathirvel; UPMC, Anesthesiology Baldi, Pierre; UC Irvine, Computer Sciences Dubrawski, Artur; Carnegie Mellon University, Computer Sciences Pinsky , MR; University of Pittsburgh School of Medicine,
Primary Subject Heading:	Anaesthesia
Secondary Subject Heading:	Surgery, Anaesthesia
Keywords:	ANAESTHETICS, BIOTECHNOLOGY & BIOINFORMATICS, CLINICAL PHYSIOLOGY, INTENSIVE & CRITICAL CARE, SURGERY

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients: Protocol for a Retrospective Study

Maxime Cannesson[1] MD PhD, Ira Hofer[1] MD, Joseph Rinehart[2] MD, Christine Lee[3] PhD, Kathirvel Subramaniam[4] MD, Pierre Baldi[5] PhD, Artur Dubrawski[6] PhD, Michael R. Pinsky[7] MD PhD

Email for all authors: Maxime Cannesson: mcannesson@mednet.ucla.edu; Ira Hofer: IHofer@mednet.ucla.edu; Joseph Rinehart: jrinehar@uci.edu; Christine Lee: cklee219@gmail.com; Kathirvel Subramanian: subramaniamk@upmc.edu; Pierre Baldi: pfbaldi@ics.uci.edu; Artur Dubrawski: awd@cs.cmu.edu; Michael R. Pinsky: pinsky@pitt.edu.

Corresponding author: Maxime Cannesson, Department of Anesthesiology and Perioperative Medicine, UCLA David Geffen School of Medicine, 757 Westwood Plaza, Los Angeles CA 90095; +1-310-657-553, mcannesson@mednet.ucla.edu

From: 1. Department of Anesthesiology and Perioperative Medicine, UCLA David Geffen School of Medicine, 2. Department of Anesthesiology and Perioperative Care, UC Irvine Health, 3. Department of Bioengineering, UC Irvine, 4. Department of Anesthesiology, University of Pittsburgh Medical Center, 5. Department of Computer Sciences, UC Irvine, 6. Artur Dubrawski, Auton Lab, School of Computer Science, Carnegie Mellon University, 7. Department of Critical Care Medicine, University of Pittsburgh Medical Center.

Keywords: Machine learning, surgery, safety, haemodynamics, blood pressure, physiology

Word Count: 3,215

Abstract

Introduction: About 42 million surgeries are performed annually in the US. While the postoperative mortality is less than 2%, 12% of all patients in the high-risk surgery group account for 80% of postoperative deaths. New onset of hemodynamic instability is common in surgical patients and its delayed treatment leads to increased morbidity and mortality. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on clinical data and high-fidelity physiological waveforms to predict hemodynamic instability during surgery.

Methods and analysis: We will initiate our work using an existing annotated intraoperative database from the University of California Irvine including clinical and high-fidelity waveform data. This data will be used for the training and development of the machine learning model (Carnegie Mellon University) that will then be tested on prospectively collected database (University of California Los Angeles). Simultaneously, we will use existing knowledge of hemodynamic instability patterns derived from our intensive care unit cohorts, medical information mart for intensive care II data, University of California Irvine data, and animal studies to create smart alarms and graphic user interface for a clinical decision support. Using machine learning, we will extract a core dataset which characterizes the signatures of normal intraoperative variability, various hemodynamic instability aetiologies, and variable responses to resuscitation. We will then employ clinician-driven iterative design to create a clinical decision support user interface, and evaluate its effect in simulated high-risk surgeries.

Ethics and dissemination: We will publish the results in a peer-reviewed publication and will present this work at professional conferences for the anaesthesiology and computer science communities. Patient-level data will be made available within 6 months after publication of the primary manuscript. The study has been approved by UCLA IRB (IRB #19-000354).

Article Summary:

Strengths and limitations of this study: Our innovations are:

1) To impact surgical practice by better and earlier identifying patients at greatest risk for cardiorespiratory instability and by providing point-of-care data-driven explanations and process-specific resuscitation using real-time data input and point-of-care management, potentially decreasing preventable surgical morbidity and mortality.

2) Our deleiverables will create a sensitive and specific means to predict which patients may or may not ever develop cardiorespiratory instability, which has important implications for patient safety, surveillance, triage and care: needed frequency of monitoring, case load mixture, workload, staff allocation, patient triage to monitored or non-monitored units, higher cost vs. lower cost bed allocation, prevention of adverse events.

3) The modelling methods we propose to develop should drastically shift intraoperative paradigms and treatment protocols and they will be applicable to existing monitoring modalities.

4) Our machine learning analytics have potential to identify new monitoring parameters to improve prediction of instability and, in our exploratory specific aim, to reverse engineer our understanding of disease aetiology during surgery.

5) Although we will focus on high-risk surgery patients, we will collect data on the whole surgical population such that we will also develop a novel database on low-risk patients but the main limitation of this study is that it is a retrospective study.

Introduction

About 42 million surgeries are performed annually in the United States (US)[1 2]. While the estimated postoperative mortality is less than 2%, 12% of all patients in the high-risk surgery group account for 80% of postoperative deaths[3 4]. To assist in guiding clinical decisions and prioritization of care, several perioperative clinical risk scores have been proposed[5-7]. The goal of these scores is to help planning clinical management and allocating resources to avoid postoperative complication and death. Recently, US hospitals have adopted electronic health report (EHR) documentation of patient care. Still, interoperability of these EHR systems remains an open issue, leading to challenges in data integration. As a result, the potential that hospital data offers in terms of understanding and improving care has not been realized. While physiological prediction tools have been developed in the critical care setting[8-11], the goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on EHR data and high-fidelity physiological waveforms to predict cardiorespiratory instability (CRI) in the perioperative / surgical setting. New onset of CRI is common in patients undergoing surgery. Delayed treatment of CRI leads to increased morbidity, mortality and use of resources in surgical patients. Even short periods of intraoperative hypotension (mean arterial pressure (MAP) < 65 mmHg) have been linked to postoperative complications such as myocardial infarction and kidney failure[12 13], and mortality[14]. Although anaesthesiologists can rescue patients with CRI - decreasing the incidence of cardiac arrests and inpatient mortality[15-17] - a more proactive approach would be to enable anaesthesiologists and nurse anaesthetists to recognize impending CRI before it happens.

Specific Aim 1 (SA1). Development of the machine learning (ML) test set (University of California, Irvine (UCI)) and retrospective validation (UC Los Angles (UCLA)) in high-risk surgery patients to identify subsequent CRI. Using ML analytics, we will extract a core dataset which best characterizes the signatures of normal intraoperative variability, various intraoperative CRI aetiologies, and variable responses to intraoperative resuscitation using 1)

existing high-granularity physiologic and clinical record data to define the structure of the database; 2) high-granularity intraoperative data from high-risk surgery patients for prospective input and further model development. Since our approaches use agnostic prioritization of physiological signals, we will explore which processes underlie specific signatures. This reverse-engineering approach will give insight into cardiorespiratory homeostasis and intraoperative CRI.

Specific Aim 2 (SA2). Prospective Validation of a clinical decision support system (CDS) tool. We will employ clinician-driven iterative design to create a novel CDS user interface, and evaluate its effect in simulated intraoperative high-risk surgeries.

Methods and Analysis

Study design

We propose to initiate our work using an already existing annotated intraoperative database (UCI) including EHR and high-fidelity waveform data. This data will be used for the initial training and development of the ML model that will then be tested on prospectively collected UCLA database (SA1). Simultaneously, we will use existing knowledge of CRI patterns derived from our step down and intensive care unit (SDU / ICU) cohorts, medical information mart for intensive care II (MIMIC II) data, UCI data, and animal studies (University of Pittsburgh) to create smart alarms and graphic user interface (GUI) for CDS (SA2).

Patient population and Data Acquisition

The UCI dataset is based on the Bernoulli® data collection system (Cardiopulmonary Corp., New Haven, CT), and as of February 2019 it includes high-fidelity physiological waveforms and EHR[18] data on more than 35,000 patients. All waveform and clinical data from surgical patients at UCI have been collected for research purposes since November 2015. The total UCI data collection as of February 2019 consists of >120,000 monitoring hours of waveform and clinical data, or >4,000 GBs of data. All waveform data is collected off of surgical patient monitors with the Bernoulli® software and equipment, and de-identified and synced retrospectively with clinical data (Figure 1). The sampling rates for EKG, Plethysmographic and Arterial waveforms are 300 Hz, 100 Hz, and 120 Hz. Clinical data is extracted from intraoperative EHR (Surgical Information Systems (Alpharetta, GA) and Epic (Verona, WI)) and synced with waveform data. These data are then linked to monitoring and clinical annotations, where adverse events are documented. At UCLA, we have established a perioperative data warehouse including all the EHR data and we plan to install an intraoperative data collection system similar to Bernoulli® for waveform collection. The EHR data at UCLA have already been analysed as proof of concepts[18 19]. The Bernoulli® and similar software (Bedmaster) provides a mission-critical application layer

designed for multi-parameter data abstraction, fusion, remediation, time-synchronization, and real-time processing. Bernoulli® provides an extensive distribution layer designed for export to third-party applications and EHR providers via HL7 or custom protocols. To create any predictive analytics, one must have a dataset free of significant artifacts. Alert artifacts greatly reduce accuracy of predictive models, which may misinform therapy and undermine response. Having experts manually annotate large amounts of data to identify all artifacts is impractical. We developed an active ML approach to identify real vital sign alerts from artefact. We found that increasing the amount of adjudicated training data improves accuracy of alert identification, but just 30% of labelled events are sufficient to confidently identify 77±11% of all the remaining artifacts[43]. We have also developed automated algorithms to automatically extract accurate clinical information from the EHR. In one project, we trained an algorithm to automatically extract duration of postoperative mechanical ventilation from the EHR after cardiac surgery. By incorporating three different data sources into our algorithm and by using preprogrammed clinical judgment to overcome common errors with data entry, our results proved to be more comprehensive, more accurate, and required a fraction of the computation time compared with manual review[32]. We will use these approaches to reduce the need for human expert annotation of monitoring events and to make our proposed scale data review realistic and manageable.

Development thrusts.

Identification of potentially informative bio-signatures. In support of SA1 we plan to maximize the chances for ML-developed feature extraction methods (time-window-gated statistics such as averages, variances, entropies and trends of univariate time series, cross-correlations between pairs of series, or spectral decompositions followed by compression with e.g. principal component analysis (PCA)) with massive-scale comprehensive searches for change points in signal to add information useful in detecting and characterizing CRI as a dichotomous variable. Some of these approaches have been used in the ICU setting to predict sepsis but have never

been used in the perioperative / surgical setting[8-11]. We have successfully applied such methods to problems involving large amounts of multivariate time series data[20]. Next, a ML algorithm uses a sample of annotated training data to identify empirically a subset of those change points that bring predictive value to models of CRI (supervised learning). In addition, we will take our quest for precursors of CRI outside of the usual single-signal or joint multi-signal modelling, by learning structures of multivariate correlations between pairs of signals, and tracking them over time. A novel use of Canonical Correlation Analysis (CCA) will be the first approach to try in this context, and we will extend it to adopt temporal regularization constraints to enable smooth transitions between consecutive timeframes. Our novel approach will let us discover pairs of combinations of features, extracted independently from two VS time series that highly correlate with each other. CCA identifies pairs of such "principal components" that could be learned from the same null space data as shown in the context of the PCA-based approach. If certain pairs of components are found to correlate consistently during stability but lose correlation at or before the onset of a CRI (or vice versa), they boost detectability of our target events. We will perform extensive searches for such CCA components and add them to the pool of factors worth tracking. We will use regularization and feature ablation to maintain parsimony of the resulting models, and to identify features of data that have key contributory effects on performance. We expect to identify such features, which would not be visible to alternatives (either using many independent single-stream models or a single (fully combined) joint model for null-space PCA-like modelling of baseline variance, or via bivariate cross-correlations of pairs of time series).

Learning predictive models of instability. We will use predictive ML models trained on the UCI annotated data to empirically identify the parsimonious set of predictors of the targeted pathologies, selected from features extracted as described above. In one instance of our preliminary work, primary observables (e.g. EKG signals, blood pressure, central venous and pulmonary arterial pressure, and pulse oximetry waveforms) are processed to generate beat-to-beat heart rate (HR), as well as the various diastolic and systolic pressure and oximetry signals.

This feature become input for Random Forest regression model that learns to predict the time since start of bleed (equivalent to the amount of blood lost given fixed bleed rate in the referred experiment), and in the process identifies a subset of all features that jointly yields optimal performance. We also create 1 and 5 minute moving averages of these derived variables, as well as often reported measures of heart rate variability along several domains (time and frequency domains, non-linear measures)[21]. The time windows determining the features from the raw electrical signal and output classes will vary according to aims. In SA1, we will enrich the development set to include a significant proportion of time windows from patients in CRI, thus the model will predict either one of the CRI states, or absence of CRI. In SA1, models predictive of fluid resuscitation will include patients in CRI and the output (fluid responsiveness) is binary. We will use the Activity Monitoring Operating Characteristic analysis as described above to validate models' sensitivity and specificity. To take our CDS beyond capability to detect intraoperative bleeding at a single specific rate, we will collect human data, retrospectively annotated by OR and ICU physicians for CRI events and apply Bayesian Aggregation (BA) method[22]. It accumulates evidence from subsequent measurements and tracks multiple hypotheses as their posterior probability distribution evolves while new evidence becomes available. BA will be our foundational approach to characterization of detected and predicted events of interest whenever low signal to noise ratios in the available data would yield more direct regressive models impractical. In particular, it will not only detect functional hypovolemia, but also estimate the rate of intravascular fluid loss.

Explanatory analysis. For our exploratory aim we will perform explanatory analysis of the learned predictive model asking what physiologic control qualities of the primary predictors of the models best explain their predictive power. In this way, we will explore foundations of autoregulation, adaptation and failure in the context of CRI. In ML experiments we will favour the types of models that avail explainability of generated results. We had success applying Random Forest classifiers and regressors in applications that require high level of user

interaction and extensive explanations of predictions made[23]. We have also tried our RIPR algorithm[24] to support interactive adjudication of alerts as either true episodes of instability or artifacts[25]. It relies on point estimators for conditional entropy and recovers a desirably small set of projections of data which accurately classify test alerts while remaining intuitive to humans. New alerts can be adjudicated using one of the projections from the retrieved set. We will extend this approach to systematically include semi-supervised and active learning concepts to support semi-automated annotation of large-scale data sets given limited availability of qualified human experts. We have also shown combined predictive and explanatory utility of learning temporal association rules from asynchronous sequences of discrete events and continuous signals with TITARL algorithm. We have shown that it can be used to identify which of the potentially large number of patterns detected in data coincide or precede particular events of interest, and present the results in a readable and interpretable form of manageably simple logical statements. We will extend this approach so that the most predictive combinations of patterns and states that can asynchronously appear in multivariate clinical data, irrespective of the temporal resolution of their observation (waveforms, beat to beat, breath to breath, disparate clinical records and demographic data), or missingness of data (frequent in hemodynamic monitoring of human patients), will be revealed, validated by expert clinicians, and used to support predictive models.

Creation of a prototype CDS system bedside GUI showing real-time probability of impending instability, lead-time to event and determining factors (SA2).

Primary model development and initial simulation testing will be done using the initial UCI and UCLA databases, and then validated at UCLA simulator centre. Then the UCLA and University of Pittsburgh datasets will serve as an external validation set for our predictive models in an iterative fashion (we will conduct external validation in years 3 to 5 of our project and will use area under the curve, sensitivity, specificity, positive predictive value, and negative predictive

value as well as precision recall to evaluate the performance of the model). First to predict CRI, then to diagnose specific aetiologies, and then to guide resuscitation based on our explanatory analyses and CDS developmenty. Thus, our primary goal is to create a robust and generalizable approach that will readily expand across other healthcare information systems. We will finalize the ML analysis of the prospectively acquired high-risk surgery patient cohort dataset in year 5. We anticipate creating final prototype GUI incorporating both measures of present instability and our predictive model with descriptions as to the potential processes that will be causing instability. We will use clinician-driven iterative design to refine the GUI.

SimStage 1. We will enlist 12 experts (6 nurse anaesthetists and 6 physician anaesthesiologists) to provide feedback. The high-fidelity simulator will be set to simulate the OR and using the prototype GUIs we have and will develop. The feedback sessions will be conducted in 2 groups of 6 clinicians to maximize variety in the input while benefitting from group dynamics. We will seek feedback regarding GUI (1) completeness or redundancy of content, (2) ease of interpretation, and (3) ergonomics. This feedback evaluation will be iterative, with a new generation of the GUI, which takes into consideration input from each group, to be rolled out sequentially. We anticipate 3-5 GUI iterations before subsequent larger group testing.

SimStage 2. We will move to larger-scale clinician-driven iterative design through evaluation of simulated clinical GUI use. In this stage the HFHS will be set as an operating room and based on 10 scenarios. The patient's monitors will live stream the data collected in C.1 for 10 selected patient cases; clinicians will also have access to the other de-identified case data such as past medical history, laboratory and diagnostic test results, current medications, etc. Of the 10 simulated cases, 8 will show evidence of instability across a 2-hr interval. Two for volume loss, two for hemodynamically unstable arrhythmia, two for post-abdominal insufflation hemodynamic instability, two for post anaesthesia induction hypotension, while two will remain stable with half the patients monitored according to current standard practice, while half will use the additional

GUI. Each scenario will be run by one anaesthesiologist and subsequently by one nurse anaesthetist. The clinicians will document patient assessment and care directives. We will then evaluate each scenario for effectiveness based on accuracy of diagnosis, time to correct diagnosis, accuracy of intervention choices (based on predefined scenarios and proposed correct interventions based on expert development), and time to intervention. These scenarios will be repeated four times, using different clinicians (i.e. total of 40 clinicians, half MD and half nurses). All clinicians will be debriefed to gain their feedback as in SimStage 1.

SimStage 3. Then we conduct a midterm analysis of effectiveness results that we will share with our original team of experts from SimStage 1, and make iterative design adjustments to the GUI, the alerting system, or the action algorithms based on SimStage 2. We will then proceed to SimStage 3, another simulated trial with the same simulated settings, scenarios and equipment, but 40 different volunteers (so as to avoid Hawthorne effect). We will collect the same information, and evaluate if the effectiveness indicators from SimStage 2 are improved upon in SimStage 3. The final refined design represents the deliverable CDS prototype for this project, and to estimate effect sizes for a future clinical trial.

Patients and Public Involvement: The choice of the outcome was guided by recent large studies showing the relationship between intraoperative hemodynamic instability and postoperative outcome. Patients were not involved in the design of the study and will not be involved in the recruitment of study participants. The results of this study will be disseminated viua publication in medical journals.

Discussion

Implications and future directions

If one could accurately predict who, when and why patients develop CRI during surgery, then effective pre-emptive treatments could be given to improve postoperative outcome and more effectively use healthcare resources. But signs of shock often occur late once organ injury is already present. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on EHR data and high-fidelity physiological waveforms to predict CRI and make the databases of intraoperative data and waveforms used for these developments freely accessible. This is extremely relevant because although 5.7 million Americans are admitted to an ICU in one year, more than 42 million undergo surgery annually. Previous and ongoing studies conducted in the ICU and in the step-down unit have built the architecture to collect real-time high-fidelity physiological waveform data streams and integrate them with patient demographics from the EHR to build large data sets, and derive actionable fused parameters based on ML analytics as well as display information in real time at the bedside to drive CDS in the critical care setting. The goal of this proposal is to apply these ML approaches to the complex and time compressed environment of high-risk surgery where greater patient and disease variability exist and shorter period of time is available to deliver truly personalized medicine approaches.

Strengths and Limitations

We will leverage our previous work and NIH/R01 funded projects in the SDU/ICU using similar methodologies to characterize CRI during surgery (R01-GM117622 and R01-NR013912). Our innovations are: 1) To impact surgical practice by better and earlier identifying patients at greatest risk for CRI and by providing point-of-care data-driven explanations and process-specific resuscitation using real-time data input and point-of-care management, potentially decreasing preventable surgical morbidity and mortality. 2) Being able to adjust for placement,

level of monitoring needed, and pre-emptive therapies and response to therapy enabling personalized medicine. It will create a sensitive and specific means to predict which patients may or may not ever develop CRI, which has important implications for patient safety, surveillance, triage and care: needed frequency of monitoring, case load mixture, workload, staff allocation, patient triage to monitored or non-monitored units, higher cost vs. lower cost bed allocation, prevention of adverse events. 3) Few innovations have been introduced to improve technologic patient surveillance and management in decades. The modelling methods we propose to develop should drastically shift intraoperative paradigms and treatment protocols and they will be applicable to existing monitoring modalities. 4) Our ML analytics also have potential to identify new monitoring parameters to improve prediction of instability and, in our exploratory specific aim, to reverse engineer our understanding of disease aetiology during surgery. 5) More than 42 million Americans undergo surgery each year and even though the perioperative complication rate is low, the absolute numbers are large. Although we will focus on high-risk surgery patients, we will collect data on the whole surgical population such that we will also develop a novel database on low-risk patients. They are an opportunity cohort to explore shared risk to compare with the high-risk patients.

Potential limitations: First, this proposal starts as a retrospective analysis of UCI data of patients with or without CRI. Limiting our variables to those available to us in retrospective analyses may limit the ultimate prediction compared to including more variables that we may find useful beyond this first pass. However, in our prospective UCLA and UPMC data collection interval, if specific parameters appear useful, we will prospectively use them in a non-protocol fashion in patients because we are those patients' providers. Second, the modelling of variables may not allow for discrimination of CRI caused by specific diagnoses, rather than by pathophysiologic processes. Still this would be an improvement over existing bedside monitoring analysis. Since we treat these surgical patients and CRI is a relatively uncommon event we will annotate patients' records weekly as to CRI and specific aetiologies for retrospective analysis. Third, our

medical centres are installing under our leadership high-density data collection systems (Bernoulli or Bedmaster) in all ORs. We have not used this data formatting/data synthesis platform before in the OR at UCLA and UPMC, but the system has been used at UCI and we have the expertise for its installation in SDU and ICU at UCLA and UPMC. We have planned for six months to reformat the data collection and secured query system as needed at UCLA and UPMC and budgeted for data collection and secured processing personnel for this support throughout the duration of the study. Based on the funding cycle criteria our system could be installed at UCLA (07/2018) before this project becomes active. Fourth, we planned for Honest Broker recording efforts for EHR review similar to ontology analysis already being done for another funded protocol (R01 NR013912). Accordingly, we have planned for a 6-month lead-in to insure she/he is cognizant of the UCLA and UPMC EHR idiosyncrasies. Finally, the CDS and GUI development will be inherently limited. We plan on only relying on board certified anaesthesiologists and certified nurse anaesthetists from UCLA for the iterative development of the CDS and GUI and this may limit the external validity of this system. Whether trainees could use this syetm appropriately would still have to be studied.

Ethics and Dissemination

Once the investigation has been completed, we intend to publish the results in a peer-reviewed publication. We also intend to present the results of this work at professional conferences for both the anaesthesiology and computer science communities. In accordance with the recent proposal from the International Committee of Medical Journal Editors, patient-level data will be made available within 6 months after publication of the primary manuscript. Data will be provided to researchers who submit a methodologically sound research proposal including a protocol and statistical analysis plan. No patient identifying fields (including dates) will be included in the shared dataset. Age will be provided in years, unless the patient is older than 89

years. In this case, age will be reported as '>89 years.' Any dates will be presented as 'number of days since index surgery.'

Author contributions:
Maxime Cannesson: Draf the manuscript and final approval of the manuscript.
Ira Hofer: Draft the manuscript and final approval of the manuscript.
Joseph Rinehart: Draft the manuscript and final approval of the manuscript.
Pierre Baldi: Draft the manuscript and final approval of the manuscript.
Christine Lee: Draft the manuscript and final approval of the manuscript.
Kathirvel Subramanian: Draft the manuscript and final approval of the manuscript.
Artur Dubrawski: Draft the manuscript and final approval of the manuscript.
Michael R. Pinsky: Draft the manuscript and final approval of the manuscript.
Patients/Public were not involved in the design of this study.

Funding: This work is supported by NIH R01 HL144692.

Conflict of Interest: Dr. Cannesson is a consultant for Edwards Lifesciences and Masimo Corp, and has funded research from Edwards Lifesciences and Masimo Corp. He is also the founder of Sironis owns patents and receive royalties for closed loop hemodynamic management that have been licensed to Edwards Lifesciences. Dr. Cannesson's Department receives funding from the NIH (R01GM117622; R01 NR013012; U54HL119893; 1R01HL144692).
Dr. Rinehart is the founder of Sironis owns patents and receive royalties for closed loop hemodynamic management that have been licensed to Edwards Lifesciences.
Dr. Christine Lee is an engineer at Edwards Lifesciences.

Acknowledgements: None

Data Statement: Data will be provided to researchers who submit a methodologically sound research proposal including a protocol and statistical analysis plan. No patient-iden- tifying fields (including dates) will be included in the shared dataset.

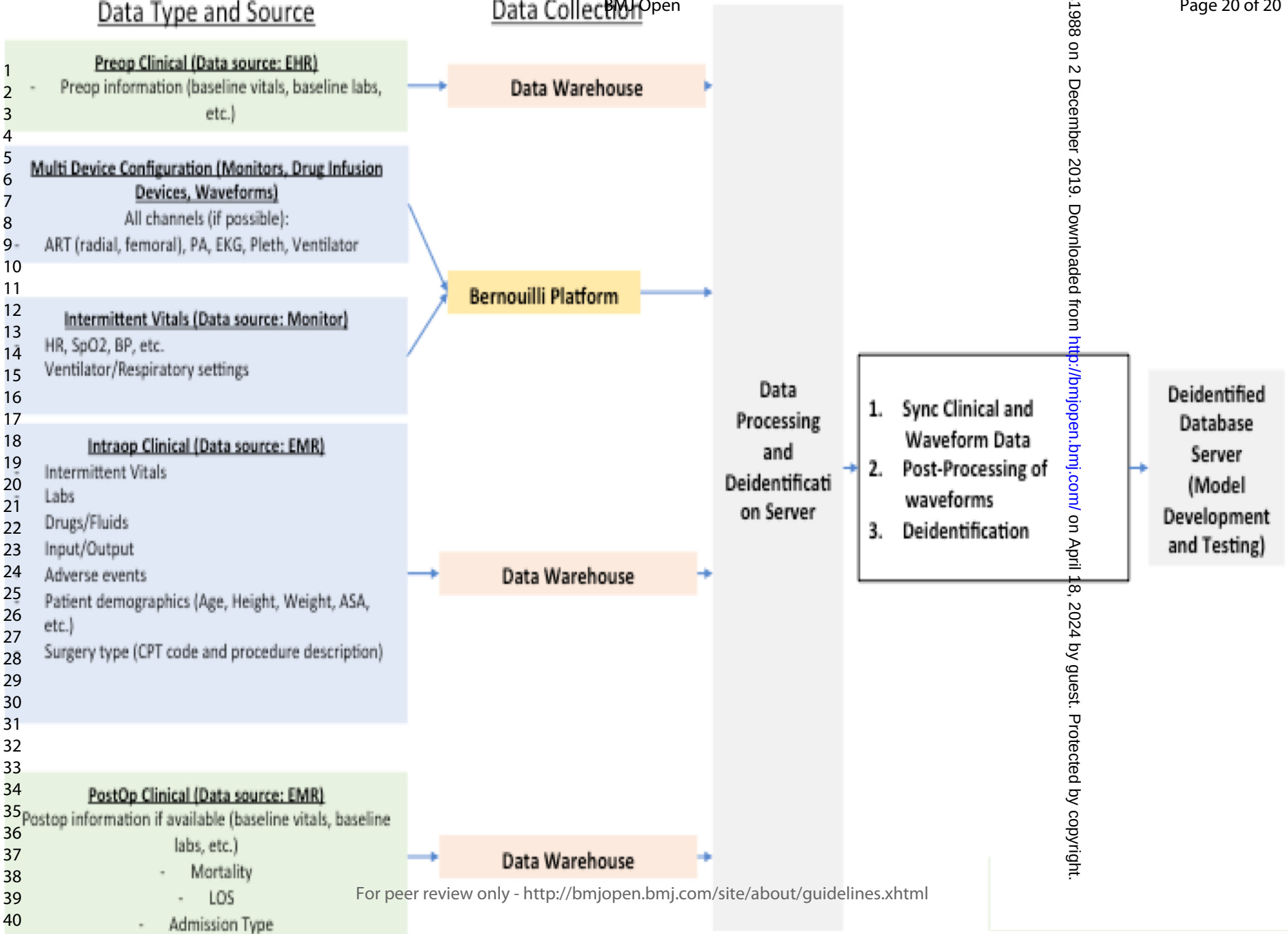
References

1. Weiser TG, Haynes AB, Molina G, et al. Size and distribution of the global volume of surgery in 2012. *Bull World Health Organ* 2016;94(3):201-09F. doi: 10.2471/BLT.15.159293
2. Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008;372(9633):139-44. doi: 10.1016/S0140-6736(08)60878-8
3. Pearse RM, Harrison DA, James P, et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. *Crit Care* 2006;10(3):R81.
4. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012;380(9847):1059-65. doi: 10.1016/S0140-6736(12)61148-9
5. Gawande AA, Kwaan MR, Regenbogen SE, et al. An Apgar score for surgery. *J Am Coll Surg* 2007;204(2):201-8. doi: 10.1016/j.jamcollsurg.2006.11.011
6. Le Manach Y, Collins G, Rodseth R, et al. Preoperative Score to Predict Postoperative Mortality (POSPOM). *Anesthesiology* 2016;124(3):570-79. doi: 10.1097/aln.0000000000000972
7. American Society of Anesthesiologists ASoAASoA. New classification of physical status. *Anesthesiology* 1963(24)
8. Ahmad S, Ramsay T, Huebsch L, et al. Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. *PLoS One* 2009;4(8):e6642. doi: 10.1371/journal.pone.0006642 [published Online First: 2009/08/15]
9. Shashikumar SP, Li Q, Clifford GD, et al. Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas* 2017;38(12):2235-48. doi: 10.1088/1361-6579/aa9772 [published Online First: 2017/11/02]
10. Shashikumar SP, Stanley MD, Sadiq I, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J Electrocardiol* 2017;50(6):739-43. doi: 10.1016/j.jelectrocard.2017.08.013 [published Online First: 2017/09/17]
11. Kamaleswaran R, Akbilgic O, Hallman MA, et al. Applying Artificial Intelligence to Identify Physiomarkers Predicting Severe Sepsis in the PICU. *Pediatr Crit Care Med* 2018;19(10):e495-e503. doi: 10.1097/PCC.0000000000001666 [published Online First: 2018/07/28]
12. Salmasi V, Maheshwari K, Yang D, et al. Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. *Anesthesiology* 2017;126:47-65. doi: 10.1097/ALN.0000000000001432
13. Walsh M, Devereaux PJ, Garg AX, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. *Anesthesiology* 2013;119(3):507-15. doi: 10.1097/ALN.0b013e3182a10e26
14. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I, Devereaux PJ, Chan MT, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2012;307(21):2295-304. doi: 10.1001/jama.2012.5502

15. Winters BD, Pham JC, Hunt EA, et al. Rapid response systems: a systematic review. *Crit Care Med* 2007;35(5):1238-43. doi: 10.1097/01.CCM.0000262388.85669.68
16. Chan PS, Khalid A, Longmore LS, et al. Hospital-wide code rates and mortality before and after implementation of a rapid response team. *JAMA* 2008;300(21):2506-13. doi: 10.1001/jama.2008.715
17. Goodrich C. Endpoints of resuscitation: what should we be monitoring? *AACN Adv Crit Care* 2006;17(3):306-16.
18. Hofer IS, Gabel E, Pfeffer M, et al. A Systematic Approach to Creation of a Perioperative Data Warehouse. . *Anesth Analg* 2016;122:1880-4.
19. Gabel E, Hofer IS, Satou N, et al. Creation and Validation of an Automated Algorithm to Determine Postoperative Ventilator Requirements After Cardiac Surgery. *Anesth Analg* 2017;124(5):1423-30. doi: 10.1213/ANE.0000000000001997
20. Dubrawski A. Extracting Useful Information from Multivariate Temporal Data. In: Jay Liebovitz TF, ed. Big Data and Business Analytics2013.
21. Tandon P, Huggins P, Dubrawski A, et al. Source location via Bayesian aggregation of evidence with mobile sensor data. IEEE Symposium on Radiation Measurements and Applications. Oakland, CA, May 2012.
22. Donze A, Clermont G, Langmead CJ. Parameter synthesis in nonlinear dynamical systems: application to systems biology. *J Comput Biol* 2010;17(3):325-36. doi: 10.1089/cmb.2009.0172
23. Milbrandt EB, Clermont G, Martinez J, et al. Predicting late anemia in critical illness. *Crit Care* 2006;10(1):R39. doi: 10.1186/cc4847
24. Projection Retrieval for Classification, Neural Information Processing Systems. NIPS; December 2012; Lake Tahoe, NV.
25. Fiterau M, Dubrawski A, Chen L, et al. Automatic Identification of Artifacts in Monitoring Critically Ill Patients. *Intensive Care Med* 2013;39(2):S470.

Figure 1. Process, workflow, and architecture of the already existing UCI database.

ART: Arterial waveform, PA: Pulmonary artery, HR: Heart rate, BP: Blood pressure, LOS: Length of stay



BMJ Open

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients: Protocol for a Retrospective Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-031988.R2
Article Type:	Protocol
Date Submitted by the Author:	31-Oct-2019
Complete List of Authors:	Cannesson, Maxime; UCLA, Anesthesiology Hofer, Ira; UCLA, Anesthesiology Rinehart, Joseph; UC Irvine, Anesthesiology Lee, Christine; UC Irvine, Bioengineering Subramaniam, Kathirvel; UPMC, Anesthesiology Baldi, Pierre; UC Irvine, Computer Sciences Dubrawski, Artur; Carnegie Mellon University, Computer Sciences Pinsky , MR; University of Pittsburgh School of Medicine,
Primary Subject Heading:	Anaesthesia
Secondary Subject Heading:	Surgery, Anaesthesia
Keywords:	ANAESTHETICS, BIOTECHNOLOGY & BIOINFORMATICS, CLINICAL PHYSIOLOGY, INTENSIVE & CRITICAL CARE, SURGERY

Machine Learning of Physiological Waveforms and Electronic Health Record Data to Predict, Diagnose, and Treat Hemodynamic Instability in Surgical Patients: Protocol for a Retrospective Study

Maxime Cannesson[1] MD PhD, Ira Hofer[1] MD, Joseph Rinehart[2] MD, Christine Lee[3] PhD, Kathirvel Subramaniam[4] MD, Pierre Baldi[5] PhD, Artur Dubrawski[6] PhD, Michael R. Pinsky[7] MD PhD

Email for all authors: Maxime Cannesson: mcannesson@mednet.ucla.edu; Ira Hofer: IHofer@mednet.ucla.edu; Joseph Rinehart: jrinehar@uci.edu; Christine Lee: cklee219@gmail.com; Kathirvel Subramanian: subramaniamk@upmc.edu; Pierre Baldi: pfbaldi@ics.uci.edu; Artur Dubrawski: awd@cs.cmu.edu; Michael R. Pinsky: pinsky@pitt.edu.

Corresponding author: Maxime Cannesson, Department of Anesthesiology and Perioperative Medicine, UCLA David Geffen School of Medicine, 757 Westwood Plaza, Los Angeles CA 90095; +1-310-657-553, mcannesson@mednet.ucla.edu

From: 1. Department of Anesthesiology and Perioperative Medicine, UCLA David Geffen School of Medicine, 2. Department of Anesthesiology and Perioperative Care, UC Irvine Health, 3. Department of Bioengineering, UC Irvine, 4. Department of Anesthesiology, University of Pittsburgh Medical Center, 5. Department of Computer Sciences, UC Irvine, 6. Artur Dubrawski, Auton Lab, School of Computer Science, Carnegie Mellon University, 7. Department of Critical Care Medicine, University of Pittsburgh Medical Center.

Keywords: Machine learning, surgery, safety, haemodynamics, blood pressure, physiology

Word Count: 3,215

Abstract

Introduction: About 42 million surgeries are performed annually in the US. While the postoperative mortality is less than 2%, 12% of all patients in the high-risk surgery group account for 80% of postoperative deaths. New onset of hemodynamic instability is common in surgical patients and its delayed treatment leads to increased morbidity and mortality. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on clinical data and high-fidelity physiological waveforms to predict hemodynamic instability during surgery.

Methods and analysis: We will initiate our work using an existing annotated intraoperative database from the University of California Irvine including clinical and high-fidelity waveform data. This data will be used for the training and development of the machine learning model (Carnegie Mellon University) that will then be tested on prospectively collected database (University of California Los Angeles). Simultaneously, we will use existing knowledge of hemodynamic instability patterns derived from our intensive care unit cohorts, medical information mart for intensive care II data, University of California Irvine data, and animal studies to create smart alarms and graphic user interface for a clinical decision support. Using machine learning, we will extract a core dataset which characterizes the signatures of normal intraoperative variability, various hemodynamic instability aetiologies, and variable responses to resuscitation. We will then employ clinician-driven iterative design to create a clinical decision support user interface, and evaluate its effect in simulated high-risk surgeries.

Ethics and dissemination: We will publish the results in a peer-reviewed publication and will present this work at professional conferences for the anaesthesiology and computer science communities. Patient-level data will be made available within 6 months after publication of the primary manuscript. The study has been approved by UCLA IRB (IRB #19-000354).

Article Summary:

Strengths and limitations of this study: Our innovations are:

1) To impact surgical practice by better and earlier identifying patients at greatest risk for cardiorespiratory instability and by providing point-of-care data-driven explanations and process-specific resuscitation using real-time data input and point-of-care management, potentially decreasing preventable surgical morbidity and mortality.

2) Our deliverables will create a sensitive and specific means to predict which patients may or may not ever develop cardiorespiratory instability, which has important implications for patient safety, surveillance, triage and care.

3) The modelling methods we propose to develop should drastically shift intraoperative paradigms and treatment protocols and they will be applicable to existing monitoring modalities.

4) Our machine learning analytics have potential to identify new monitoring parameters to improve prediction of instability and, in our exploratory specific aim, to reverse engineer our understanding of disease aetiology during surgery.

5) Although we will focus on high-risk surgery patients, we will collect data on the whole surgical population such that we will also develop a novel database on low-risk patients but the main limitation of this study is that it is a retrospective study.

Introduction

About 42 million surgeries are performed annually in the United States (US)[1 2]. While the estimated postoperative mortality is less than 2%, 12% of all patients in the high-risk surgery group account for 80% of postoperative deaths[3 4]. To assist in guiding clinical decisions and prioritization of care, several perioperative clinical risk scores have been proposed[5-7]. The goal of these scores is to help planning clinical management and allocating resources to avoid postoperative complication and death. Recently, US hospitals have adopted electronic health report (EHR) documentation of patient care. Still, interoperability of these EHR systems remains an open issue, leading to challenges in data integration. As a result, the potential that hospital data offers in terms of understanding and improving care has not been realized. While physiological prediction tools have been developed in the critical care setting[8-11], the goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on EHR data and high-fidelity physiological waveforms to predict cardiorespiratory instability (CRI) in the perioperative / surgical setting. New onset of CRI is common in patients undergoing surgery. Delayed treatment of CRI leads to increased morbidity, mortality and use of resources in surgical patients. Even short periods of intraoperative hypotension (mean arterial pressure (MAP) < 65 mmHg) have been linked to postoperative complications such as myocardial infarction and kidney failure[12 13], and mortality[14]. Although anaesthesiologists can rescue patients with CRI - decreasing the incidence of cardiac arrests and inpatient mortality[15-17] - a more proactive approach would be to enable anaesthesiologists and nurse anaesthetists to recognize impending CRI before it happens.

Specific Aim 1 (SA1). Development of the machine learning (ML) test set (University of California, Irvine (UCI)) and retrospective validation (UC Los Angles (UCLA)) in high-risk surgery patients to identify subsequent CRI. Using ML analytics, we will extract a core dataset which best characterizes the signatures of normal intraoperative variability, various intraoperative CRI aetiologies, and variable responses to intraoperative resuscitation using 1)

existing high-granularity physiologic and clinical record data to define the structure of the database; 2) high-granularity intraoperative data from high-risk surgery patients for prospective input and further model development. Since our approaches use agnostic prioritization of physiological signals, we will explore which processes underlie specific signatures. This reverse-engineering approach will give insight into cardiorespiratory homeostasis and intraoperative CRI.

Specific Aim 2 (SA2). Prospective Validation of a clinical decision support system (CDS) tool. We will employ clinician-driven iterative design to create a novel CDS user interface, and evaluate its effect in simulated intraoperative high-risk surgeries.

Methods and Analysis

Study design

We propose to initiate our work using an already existing annotated intraoperative database (UCI) including EHR and high-fidelity waveform data. This data will be used for the initial training and development of the ML model that will then be tested on prospectively collected UCLA database (SA1). Simultaneously, we will use existing knowledge of CRI patterns derived from our step down and intensive care unit (SDU / ICU) cohorts, medical information mart for intensive care II (MIMIC II) data, UCI data, and animal studies (University of Pittsburgh) to create smart alarms and graphic user interface (GUI) for CDS (SA2).

Patient population and Data Acquisition

The UCI dataset is based on the Bernoulli® data collection system (Cardiopulmonary Corp., New Haven, CT), and as of February 2019 it includes high-fidelity physiological waveforms and EHR[18] data on more than 35,000 patients. All waveform and clinical data from surgical patients at UCI have been collected for research purposes since November 2015. The total UCI data collection as of February 2019 consists of >120,000 monitoring hours of waveform and clinical data, or >4,000 GBs of data. All waveform data is collected off of surgical patient monitors with the Bernoulli® software and equipment, and de-identified and synced retrospectively with clinical data (Figure 1). The sampling rates for EKG, Plethysmographic and Arterial waveforms are 300 Hz, 100 Hz, and 120 Hz. Clinical data is extracted from intraoperative EHR (Surgical Information Systems (Alpharetta, GA) and Epic (Verona, WI)) and synced with waveform data. These data are then linked to monitoring and clinical annotations, where adverse events are documented. At UCLA, we have established a perioperative data warehouse including all the EHR data and we plan to install an intraoperative data collection system similar to Bernoulli® for waveform collection. The EHR data at UCLA have already been analysed as proof of concepts[18 19]. The Bernoulli® and similar software (Bedmaster) provides a mission-critical application layer

designed for multi-parameter data abstraction, fusion, remediation, time-synchronization, and real-time processing. Bernoulli® provides an extensive distribution layer designed for export to third-party applications and EHR providers via HL7 or custom protocols. To create any predictive analytics, one must have a dataset free of significant artifacts. Alert artifacts greatly reduce accuracy of predictive models, which may misinform therapy and undermine response. Having experts manually annotate large amounts of data to identify all artifacts is impractical. We developed an active ML approach to identify real vital sign alerts from artefact. We found that increasing the amount of adjudicated training data improves accuracy of alert identification, but just 30% of labelled events are sufficient to confidently identify 77±11% of all the remaining artifacts[43]. We have also developed automated algorithms to automatically extract accurate clinical information from the EHR. In one project, we trained an algorithm to automatically extract duration of postoperative mechanical ventilation from the EHR after cardiac surgery. By incorporating three different data sources into our algorithm and by using preprogrammed clinical judgment to overcome common errors with data entry, our results proved to be more comprehensive, more accurate, and required a fraction of the computation time compared with manual review[32]. We will use these approaches to reduce the need for human expert annotation of monitoring events and to make our proposed scale data review realistic and manageable.

Development thrusts.

Identification of potentially informative bio-signatures. In support of SA1 we plan to maximize the chances for ML-developed feature extraction methods (time-window-gated statistics such as averages, variances, entropies and trends of univariate time series, cross-correlations between pairs of series, or spectral decompositions followed by compression with e.g. principal component analysis (PCA)) with massive-scale comprehensive searches for change points in signal to add information useful in detecting and characterizing CRI as a dichotomous variable. The definition of CRI is defined both by exceedances of vital sign parameters (HR, RR, BP,

SpO2) beyond thresholds, EHR defined resuscitative actions such as bolus of luids, infusion of vasopressors, and inotropes. Initially, we will strive for 15 minutes advanced warning for volume loss, 15 minutes for hemodynamically unstable arrhythmia, 10 minutes for post abdominal insufflation hemodynamic instability, and 5 minutes for post anesthesia induction hypotension, prior to overt clinical signs and symptoms of CRI.

Some of these approaches have been used in the ICU setting to predict sepsis but have never been used in the perioperative / surgical setting[8-11]. We have successfully applied such methods to problems involving large amounts of multivariate time series data[20]. Next, a ML algorithm uses a sample of annotated training data to identify empirically a subset of those change points that bring predictive value to models of CRI (supervised learning). In addition, we will take our quest for precursors of CRI outside of the usual single-signal or joint multi-signal modelling, by learning structures of multivariate correlations between pairs of signals, and tracking them over time. A novel use of Canonical Correlation Analysis (CCA) will be the first approach to try in this context, and we will extend it to adopt temporal regularization constraints to enable smooth transitions between consecutive timeframes. Our novel approach will let us discover pairs of combinations of features, extracted independently from two VS time series that highly correlate with each other. CCA identifies pairs of such “principal components” that could be learned from the same null space data as shown in the context of the PCA-based approach. If certain pairs of components are found to correlate consistently during stability but lose correlation at or before the onset of a CRI (or vice versa), they boost detectability of our target events. We will perform extensive searches for such CCA components and add them to the pool of factors worth tracking. We will use regularization and feature ablation to maintain parsimony of the resulting models, and to identify features of data that have key contributory effects on performance. We expect to identify such features, which would not be visible to alternatives (either using many independent single-stream models or a single (fully combined) joint model for null-space PCA-like modelling of baseline variance, or via bivariate cross-correlations of pairs of time series).

Learning predictive models of instability. We will use predictive ML models trained on the UCI annotated data to empirically identify the parsimonious set of predictors of the targeted pathologies, selected from features extracted as described above. In one instance of our preliminary work, primary observables (e.g. EKG signals, blood pressure, central venous and pulmonary arterial pressure, and pulse oximetry waveforms) are processed to generate beat-to-beat heart rate (HR), as well as the various diastolic and systolic pressure and oximetry signals. This feature become input for Random Forest regression model that learns to predict the time since start of bleed (equivalent to the amount of blood lost given fixed bleed rate in the referred experiment), and in the process identifies a subset of all features that jointly yields optimal performance. We also create 1 and 5 minute moving averages of these derived variables, as well as often reported measures of heart rate variability along several domains (time and frequency domains, non-linear measures)[21]. The time windows determining the features from the raw electrical signal and output classes will vary according to aims. In SA1, we will enrich the development set to include a significant proportion of time windows from patients in CRI, thus the model will predict either one of the CRI states, or absence of CRI. In SA1, models predictive of fluid resuscitation will include patients in CRI and the output (fluid responsiveness) is binary. We will use the Activity Monitoring Operating Characteristic analysis as described above to validate models' sensitivity and specificity. To take our CDS beyond capability to detect intraoperative bleeding at a single specific rate, we will collect human data, retrospectively annotated by OR and ICU physicians for CRI events and apply Bayesian Aggregation (BA) method[22]. It accumulates evidence from subsequent measurements and tracks multiple hypotheses as their posterior probability distribution evolves while new evidence becomes available. BA will be our foundational approach to characterization of detected and predicted events of interest whenever low signal to noise ratios in the available data would yield more direct regressive models impractical. In particular, it will not only detect functional hypovolemia, but also estimate the rate of intravascular fluid loss.

Explanatory analysis. For our exploratory aim we will perform explanatory analysis of the learned predictive model asking what physiologic control qualities of the primary predictors of the models best explain their predictive power. In this way, we will explore foundations of autoregulation, adaptation and failure in the context of CRI. In ML experiments we will favour the types of models that avail explainability of generated results. We had success applying Random Forest classifiers and regressors in applications that require high level of user interaction and extensive explanations of predictions made[23]. We have also tried our RIPR algorithm[24] to support interactive adjudication of alerts as either true episodes of instability or artifacts[25]. It relies on point estimators for conditional entropy and recovers a desirably small set of projections of data which accurately classify test alerts while remaining intuitive to humans. New alerts can be adjudicated using one of the projections from the retrieved set. We will extend this approach to systematically include semi-supervised and active learning concepts to support semi-automated annotation of large-scale data sets given limited availability of qualified human experts. We have also shown combined predictive and explanatory utility of learning temporal association rules from asynchronous sequences of discrete events and continuous signals with TITARL algorithm. We have shown that it can be used to identify which of the potentially large number of patterns detected in data coincide or precede particular events of interest, and present the results in a readable and interpretable form of manageably simple logical statements. We will extend this approach so that the most predictive combinations of patterns and states that can asynchronously appear in multivariate clinical data, irrespective of the temporal resolution of their observation (waveforms, beat to beat, breath to breath, disparate clinical records and demographic data), or missingness of data (frequent in hemodynamic monitoring of human patients), will be revealed, validated by expert clinicians, and used to support predictive models.

Creation of a prototype CDS system bedside GUI showing real-time probability of impending instability, lead-time to event and determining factors (SA2).

Primary model development and initial simulation testing will be done using the initial UCI and UCLA databases, and then validated at UCLA simulator centre. Then the UCLA and University of Pittsburgh datasets will serve as an external validation set for our predictive models in an iterative fashion (we will conduct external validation in years 3 to 5 of our project and will use area under the curve, sensitivity, specificity, positive predictive value, and negative predictive value as well as precision recall to evaluate the performance of the model). First to predict CRI, then to diagnose specific aetiologies, and then to guide resuscitation based on our explanatory analyses and CDS developmenty. Thus, our primary goal is to create a robust and generalizable approach that will readily expand across other healthcare information systems. We will finalize the ML analysis of the prospectively acquired high-risk surgery patient cohort dataset in year 5. We anticipate creating final prototype GUI incorporating both measures of present instability and our predictive model with descriptions as to the potential processes that will be causing instability. We will use clinician-driven iterative design to refine the GUI.

SimStage 1. We will enlist 12 experts (6 nurse anaesthetists and 6 physician anaesthesiologists) to provide feedback. The high-fidelity simulator will be set to simulate the OR and using the prototype GUIs we have and will develop. The feedback sessions will be conducted in 2 groups of 6 clinicians to maximize variety in the input while benefitting from group dynamics. We will seek feedback regarding GUI (1) completeness or redundancy of content, (2) ease of interpretation, and (3) ergonomics. This feedback evaluation will be iterative, with a new generation of the GUI, which takes into consideration input from each group, to be rolled out sequentially. We anticipate 3-5 GUI iterations before subsequent larger group testing.

SimStage 2. We will move to larger-scale clinician-driven iterative design through evaluation of simulated clinical GUI use. In this stage the HFHS will be set as an operating room and based

on 10 scenarios. The patient's monitors will live stream the data collected in C.1 for 10 selected patient cases; clinicians will also have access to the other de-identified case data such as past medical history, laboratory and diagnostic test results, current medications, etc. Of the 10 simulated cases, 8 will show evidence of instability across a 2-hr interval. Two for volume loss, two for hemodynamically unstable arrhythmia, two for post-abdominal insufflation hemodynamic instability, two for post anaesthesia induction hypotension, while two will remain stable with half the patients monitored according to current standard practice, while half will use the additional GUI. Each scenario will be run by one anaesthesiologist and subsequently by one nurse anaesthetist. The clinicians will document patient assessment and care directives. We will then evaluate each scenario for effectiveness based on accuracy of diagnosis, time to correct diagnosis, accuracy of intervention choices (based on predefined scenarios and proposed correct interventions based on expert development), and time to intervention. These scenarios will be repeated four times, using different clinicians (i.e. total of 40 clinicians, half MD and half nurses). All clinicians will be debriefed to gain their feedback as in SimStage 1.

SimStage 3. Then we conduct a midterm analysis of effectiveness results that we will share with our original team of experts from SimStage 1, and make iterative design adjustments to the GUI, the alerting system, or the action algorithms based on SimStage 2. We will then proceed to SimStage 3, another simulated trial with the same simulated settings, scenarios and equipment, but 40 different volunteers (so as to avoid Hawthorne effect). We will collect the same information, and evaluate if the effectiveness indicators from SimStage 2 are improved upon in SimStage 3. The final refined design represents the deliverable CDS prototype for this project, and to estimate effect sizes for a future clinical trial.

Patients and Public Involvement: The choice of the outcome was guided by recent large studies showing the relationship between intraoperative hemodynamic instability and postoperative outcome. Patients were not involved in the design of the study and will not be involved in the

recruitment of study participants. The results of this study will be disseminated viua publication in medical journals.

Discussion

Implications and future directions

If one could accurately predict who, when and why patients develop CRI during surgery, then effective pre-emptive treatments could be given to improve postoperative outcome and more effectively use healthcare resources. But signs of shock often occur late once organ injury is already present. The goal of this proposal is to develop, validate, and test real-time intraoperative risk prediction tools based on EHR data and high-fidelity physiological waveforms to predict CRI and make the databases of intraoperative data and waveforms used for these developments freely accessible. This is extremely relevant because although 5.7 million Americans are admitted to an ICU in one year, more than 42 million undergo surgery annually. Previous and ongoing studies conducted in the ICU and in the step-down unit have built the architecture to collect real-time high-fidelity physiological waveform data streams and integrate them with patient demographics from the EHR to build large data sets, and derive actionable fused parameters based on ML analytics as well as display information in real time at the bedside to drive CDS in the critical care setting. The goal of this proposal is to apply these ML approaches to the complex and time compressed environment of high-risk surgery where greater patient and disease variability exist and shorter period of time is available to deliver truly personalized medicine approaches.

Strengths and Limitations

We will leverage our previous work and NIH/R01 funded projects in the SDU/ICU using similar methodologies to characterize CRI during surgery (R01-GM117622 and R01-NR013912). Our innovations are: 1) To impact surgical practice by better and earlier identifying patients at greatest risk for CRI and by providing point-of-care data-driven explanations and process-specific resuscitation using real-time data input and point-of-care management, potentially decreasing preventable surgical morbidity and mortality. 2) Being able to adjust for placement,

level of monitoring needed, and pre-emptive therapies and response to therapy enabling personalized medicine. It will create a sensitive and specific means to predict which patients may or may not ever develop CRI, which has important implications for patient safety, surveillance, triage and care: needed frequency of monitoring, case load mixture, workload, staff allocation, patient triage to monitored or non-monitored units, higher cost vs. lower cost bed allocation, prevention of adverse events. 3) Few innovations have been introduced to improve technologic patient surveillance and management in decades. The modelling methods we propose to develop should drastically shift intraoperative paradigms and treatment protocols and they will be applicable to existing monitoring modalities. 4) Our ML analytics also have potential to identify new monitoring parameters to improve prediction of instability and, in our exploratory specific aim, to reverse engineer our understanding of disease aetiology during surgery. 5) More than 42 million Americans undergo surgery each year and even though the perioperative complication rate is low, the absolute numbers are large. Although we will focus on high-risk surgery patients, we will collect data on the whole surgical population such that we will also develop a novel database on low-risk patients. They are an opportunity cohort to explore shared risk to compare with the high-risk patients.

Potential limitations: First, this proposal starts as a retrospective analysis of UCI data of patients with or without CRI. Limiting our variables to those available to us in retrospective analyses may limit the ultimate prediction compared to including more variables that we may find useful beyond this first pass. However, in our prospective UCLA and UPMC data collection interval, if specific parameters appear useful, we will prospectively use them in a non-protocol fashion in patients because we are those patients' providers. Second, the modelling of variables may not allow for discrimination of CRI caused by specific diagnoses, rather than by pathophysiologic processes. Still this would be an improvement over existing bedside monitoring analysis. Since we treat these surgical patients and CRI is a relatively uncommon event we will annotate patients' records weekly as to CRI and specific aetiologies for retrospective analysis. Third, our

medical centres are installing under our leadership high-density data collection systems (Bernoulli or Bedmaster) in all ORs. We have not used this data formatting/data synthesis platform before in the OR at UCLA and UPMC, but the system has been used at UCI and we have the expertise for its installation in SDU and ICU at UCLA and UPMC. We have planned for six months to reformat the data collection and secured query system as needed at UCLA and UPMC and budgeted for data collection and secured processing personnel for this support throughout the duration of the study. Based on the funding cycle criteria our system could be installed at UCLA (07/2018) before this project becomes active. Fourth, we planned for Honest Broker recording efforts for EHR review similar to ontology analysis already being done for another funded protocol (R01 NR013912). Accordingly, we have planned for a 6-month lead-in to insure she/he is cognizant of the UCLA and UPMC EHR idiosyncrasies. Finally, the CDS and GUI development will be inherently limited. We plan on only relying on board certified anaesthesiologists and certified nurse anaesthetists from UCLA for the iterative development of the CDS and GUI and this may limit the external validity of this system. Whether trainees could use this syetm appropriately would still have to be studied.

Ethics and Dissemination

Once the investigation has been completed, we intend to publish the results in a peer-reviewed publication. We also intend to present the results of this work at professional conferences for both the anaesthesiology and computer science communities. In accordance with the recent proposal from the International Committee of Medical Journal Editors, patient-level data will be made available within 6 months after publication of the primary manuscript. Data will be provided to researchers who submit a methodologically sound research proposal including a protocol and statistical analysis plan. No patient identifying fields (including dates) will be included in the shared dataset. Age will be provided in years, unless the patient is older than 89

years. In this case, age will be reported as '>89 years.' Any dates will be presented as 'number of days since index surgery.'

Author contributions:
Maxime Cannesson: Draf the manuscript and final approval of the manuscript.
Ira Hofer: Draft the manuscript and final approval of the manuscript.
Joseph Rinehart: Draft the manuscript and final approval of the manuscript.
Pierre Baldi: Draft the manuscript and final approval of the manuscript.
Christine Lee: Draft the manuscript and final approval of the manuscript.
Kathirvel Subramanian: Draft the manuscript and final approval of the manuscript.
Artur Dubrawski: Draft the manuscript and final approval of the manuscript.
Michael R. Pinsky: Draft the manuscript and final approval of the manuscript.
Patients/Public were not involved in the design of this study.

Funding: This work is supported by NIH R01 HL144692.

Conflict of Interest: Dr. Cannesson is a consultant for Edwards Lifesciences and Masimo Corp, and has funded research from Edwards Lifesciences and Masimo Corp. He is also the founder of Sironis owns patents and receive royalties for closed loop hemodynamic management that have been licensed to Edwards Lifesciences. Dr. Cannesson's Department receives funding from the NIH (R01GM117622; R01 NR013012; U54HL119893; 1R01HL144692).
Dr. Rinehart is the founder of Sironis owns patents and receive royalties for closed loop hemodynamic management that have been licensed to Edwards Lifesciences.
Dr. Christine Lee is an engineer at Edwards Lifesciences.

Acknowledgements: None

Data Statement: Data will be provided to researchers who submit a methodologically sound research proposal including a protocol and statistical analysis plan. No patient-iden- tifying fields (including dates) will be included in the shared dataset.

References

1. Weiser TG, Haynes AB, Molina G, et al. Size and distribution of the global volume of surgery in 2012. *Bull World Health Organ* 2016;94(3):201-09F. doi: 10.2471/BLT.15.159293
2. Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008;372(9633):139-44. doi: 10.1016/S0140-6736(08)60878-8
3. Pearse RM, Harrison DA, James P, et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. *Crit Care* 2006;10(3):R81.
4. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012;380(9847):1059-65. doi: 10.1016/S0140-6736(12)61148-9
5. Gawande AA, Kwaan MR, Regenbogen SE, et al. An Apgar score for surgery. *J Am Coll Surg* 2007;204(2):201-8. doi: 10.1016/j.jamcollsurg.2006.11.011
6. Le Manach Y, Collins G, Rodseth R, et al. Preoperative Score to Predict Postoperative Mortality (POSPOM). *Anesthesiology* 2016;124(3):570-79. doi: 10.1097/aln.0000000000000972
7. American Society of Anesthesiologists ASoAASoA. New classification of physical status. *Anesthesiology* 1963(24)
8. Ahmad S, Ramsay T, Huebsch L, et al. Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. *PLoS One* 2009;4(8):e6642. doi: 10.1371/journal.pone.0006642 [published Online First: 2009/08/15]
9. Shashikumar SP, Li Q, Clifford GD, et al. Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas* 2017;38(12):2235-48. doi: 10.1088/1361-6579/aa9772 [published Online First: 2017/11/02]
10. Shashikumar SP, Stanley MD, Sadiq I, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J Electrocardiol* 2017;50(6):739-43. doi: 10.1016/j.jelectrocard.2017.08.013 [published Online First: 2017/09/17]
11. Kamaleswaran R, Akbilgic O, Hallman MA, et al. Applying Artificial Intelligence to Identify Physiomarkers Predicting Severe Sepsis in the PICU. *Pediatr Crit Care Med* 2018;19(10):e495-e503. doi: 10.1097/PCC.0000000000001666 [published Online First: 2018/07/28]
12. Salmasi V, Maheshwari K, Yang D, et al. Relationship between Intraoperative Hypotension, Defined by Either Reduction from Baseline or Absolute Thresholds, and Acute Kidney and Myocardial Injury after Noncardiac Surgery: A Retrospective Cohort Analysis. *Anesthesiology* 2017;126:47-65. doi: 10.1097/ALN.0000000000001432
13. Walsh M, Devereaux PJ, Garg AX, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. *Anesthesiology* 2013;119(3):507-15. doi: 10.1097/ALN.0b013e3182a10e26
14. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I, Devereaux PJ, Chan MT, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA* 2012;307(21):2295-304. doi: 10.1001/jama.2012.5502

15. Winters BD, Pham JC, Hunt EA, et al. Rapid response systems: a systematic review. *Crit Care Med* 2007;35(5):1238-43. doi: 10.1097/01.CCM.0000262388.85669.68
16. Chan PS, Khalid A, Longmore LS, et al. Hospital-wide code rates and mortality before and after implementation of a rapid response team. *JAMA* 2008;300(21):2506-13. doi: 10.1001/jama.2008.715
17. Goodrich C. Endpoints of resuscitation: what should we be monitoring? *AACN Adv Crit Care* 2006;17(3):306-16.
18. Hofer IS, Gabel E, Pfeffer M, et al. A Systematic Approach to Creation of a Perioperative Data Warehouse. . *Anesth Analg* 2016;122:1880-4.
19. Gabel E, Hofer IS, Satou N, et al. Creation and Validation of an Automated Algorithm to Determine Postoperative Ventilator Requirements After Cardiac Surgery. *Anesth Analg* 2017;124(5):1423-30. doi: 10.1213/ANE.0000000000001997
20. Dubrawski A. Extracting Useful Information from Multivariate Temporal Data. In: Jay Liebovitz TF, ed. Big Data and Business Analytics2013.
21. Tandon P, Huggins P, Dubrawski A, et al. Source location via Bayesian aggregation of evidence with mobile sensor data. IEEE Symposium on Radiation Measurements and Applications. Oakland, CA, May 2012.
22. Donze A, Clermont G, Langmead CJ. Parameter synthesis in nonlinear dynamical systems: application to systems biology. *J Comput Biol* 2010;17(3):325-36. doi: 10.1089/cmb.2009.0172
23. Milbrandt EB, Clermont G, Martinez J, et al. Predicting late anemia in critical illness. *Crit Care* 2006;10(1):R39. doi: 10.1186/cc4847
24. Projection Retrieval for Classification, Neural Information Processing Systems. NIPS; December 2012; Lake Tahoe, NV.
25. Fiterau M, Dubrawski A, Chen L, et al. Automatic Identification of Artifacts in Monitoring Critically Ill Patients. *Intensive Care Med* 2013;39(2):S470.

Figure 1. Process, workflow, and architecture of the already existing UCI database.

ART: Arterial waveform, PA: Pulmonary artery, HR: Heart rate, BP: Blood pressure, LOS: Length of stay