PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	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
AUTHORS	Cannesson, Maxime; Hofer, Ira; Rinehart, Joseph; Lee, Christine; Subramaniam, Kathirvel; Baldi, Pierre; Dubrawski, Artur; Pinsky, MR

VERSION 1 – REVIEW

REVIEWER	Rishikesan Kamaleswaran University of Tennessee Health Science Center, Memphis, TN, USA
REVIEW RETURNED	29-Jun-2019

GENERAL COMMENTS	The authors provide a valuable protocol for the implementation of a machine learning pipeline to analyze a variety of clinical and physiological data sources to predict cardiorespiratory instability. The paper is well written and of high quality. However, a few recommendations are provided below to improve the quality of the manuscript. Introduction: 1. The introduction reads nicely, however, there should be more details about prior work involving high-frequency data streams for identifying cardiorespiratory deterioration, which in this version of the manuscript has not sufficiently been covered. While those similar approaches may not be within the intraoperative space, it is still important to acknowledge relevant methodologically similar work: - Ahmad, S., Ramsay, T., Huebsch, L., Flanagan, S., McDiarmid, S., Batkin, I., McIntyre, L., Sundaresan, S.R., Maziak, D.E., Shamji, F.M. and Hebert, P., 2009. Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. PloS one, 4(8), p.e6642. - Blount, M., McGregor, C., James, A., Sow, D., Kamaleswaran, R., Tuuha, S., Percival, J. and Percival, N., 2010, November. On the integration of an artifact system and a real-time healthcare analytics system. In Proceedings of the 1st ACM International Health Informatics Symposium (pp. 647-655). ACM. - Shashikumar, S.P., Stanley, M.D., Sadiq, I., Li, Q., Holder, A., Clifford, G.D. and Nemati, S., 2017. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. Journal of electrocardiology, 50(6), pp.739-743. - Kamaleswaran, R., Akbilgic, O., Hallman, M.A., West, A.N., Davis, R.L. and Shah, S.H., 2018. Applying artificial intelligence to

identify physiomarkers predicting severe sepsis in the PICU. Pediatric Critical Care Medicine, 19(10), pp.e495-e503.

Methods:

1. In the developmental thrusts section I would like to see a comparison to similar approaches that have become standardized for analyzing high-frequency data streams, please see above references for more detail.

2. It is still not clear whether these waveform features are extracted from peak annotations or the raw electrical signal, please confirm and expand as necessary. I.e. "… generate beat-tobeat HR, as well as the various diastolic and systolic pressure and oximetry signals" It is unclear whether the features are the numeric physiological values or some higher-dimensional abstraction.

3. Please review the manuscript for undefined acronyms. i.e. AMOC, RIPR etc..

4. A subsection on how you will validate your models for generalization is highly recommended. There is a scant reference to performance measures, stability principle checks, etc for reproducibility.

5. The GUI implementation is interesting, however, there should be some discussion about threats to external validity from using a small number of experts. Additional detail about these experts are also missing, i.e. are they from a single site, similar training background etc.?

6. Can you clarify whether experts required for the early phase of development? Could those initial developments be tested by a larger number of novices, i.e. mechanical turk etc.?

Discussions:

There is some ambiguity about the development pipeline introduced in the limitations section, p.g. 14 line 46, which seems to imply that the pipeline will be applied in some ad-hoc prospective manner. Would you expand/clarify this in the methods section?

REVIEWER	Daniel I. Sessler
	Michael Cudahy Professor and Chair
	Department of Outcomes Research
	Cleveland Clinic
	9500 Euclid Ave — P77
	Cleveland, OH 44195
	I am a consultant for Edwards Lifesciences
REVIEW RETURNED	04-Jul-2019

| GENERAL COMMENTS | The authors are authorities in physiological modeling and are well positioned to do the proposed research. If anyone can do the proposed research, they are the ones.

The general purpose of publishing protocols is to provide a detailed record of what is planned so reviewers of the eventual results report can confirm that authors are transparent about what was planned and what changed. My concern for this type of research is that it is inherently exploratory and it is obvious that the investigators will need to try many different approaches in the hopes of finding one that works. That is perfectly appropriate, but would seem to make publication of a protocol manuscript unnecessary. |

	As is, the manuscript doesn't contain the detail I would have expected such as detailed statistical methodology and defined primary and secondary outcomes, or any measures of success including what level of prediction will be considered clinically meaningful. I think it would be difficult for other investigators to replicate the proposed research based on the presented detail. The abstract seems brief and doesn't contain much detail about the planned analysis. The manuscript reads more like a funding application than a conventional protocol manuscript. For example, "We will finalize the ML analysis ... in year 5."

REVIEWER	Bradley Fritz Washington University School of Medicine, United States
REVIEW RETURNED	14-Jul-2019

GENERAL COMMENTS	I am excited to read this protocol for a study where a prediction model for intraoperative cardiorespiratory instability will be trained using a retrospective cohort and then validated prospectively. The authors also propose to create a user interface and to use clinician focus groups and simulated operating room scenarios to refine these interfaces. This is an exciting project, but I do have some questions that might make the protocol easier for the reader to understand. 1. My major question is how cardiorespiratory instability is defined. The authors mention that the UC Irvine dataset includes "clinical annotations;" is that where they plan to obtain the labels? Who created these annotations (the anesthesia team taking care of the patient, previous research teams, hospital administration?) and what criteria were used for making annotations? How will label information for the target variable be obtained in the prospective phase? 2. Are the authors planning a model to predict the dichotomous target of cardiorespiratory instability, or will they be predicting a multi-class target of different types (causes?) of cardiorespiratory instability? If the latter is true, then how will the classes be defined? 3. Page 7 includes several concrete references to "time since start of bleed" (line 23) and "intraoperative bleeding" (line 45). Is this one of the classes for the target? 4. What will be the methods for the external validation of the model at UCLA and UPMC? In what time period will data be collected? How will model performance be measured? 5. In what way will the model be used to "guide resuscitation" (page 9, line 16)? 6. In the simulations, what does "accuracy of intervention choices" mean? There are usually many appropriate ways to respond to a given clinical scenario.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Rishikesan Kamaleswaran
Institution and Country: University of Tennessee Health Science Center, Memphis, TN, USA
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The authors provide a valuable protocol for the implementation of a machine learning pipeline to analyze a variety of clinical and physiological data sources to predict cardiorespiratory instability. The paper is well written and of high quality. However, a few recommendations are provided below to improve the quality of the manuscript.

Thank you very much for these positive comments.

Introduction:
1. The introduction reads nicely, however, there should be more details about prior work involving high-frequency data streams for identifying cardiorespiratory deterioration, which in this version of the manuscript has not sufficiently been covered. While those similar approaches may not be within the intraoperative space, it is still important to acknowledge relevant methodologically similar work:
- Ahmad, S., Ramsay, T., Huebsch, L., Flanagan, S., McDiarmid, S., Batkin, I., McIntyre, L., Sundaresan, S.R., Maziak, D.E., Shamji, F.M. and Hebert, P., 2009. Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. PloS one, 4(8), p.e6642.
- Blount, M., McGregor, C., James, A., Sow, D., Kamaleswaran, R., Tuuha, S., Percival, J. and Percival, N., 2010, November. On the integration of an artifact system and a real-time healthcare analytics system. In Proceedings of the 1st ACM International Health Informatics Symposium (pp. 647-655). ACM.
- Shashikumar, S.P., Stanley, M.D., Sadiq, I., Li, Q., Holder, A., Clifford, G.D. and Nemati, S., 2017. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. Journal of electrocardiology, 50(6), pp.739-743.
- Kamaleswaran, R., Akbilgic, O., Hallman, M.A., West, A.N., Davis, R.L. and Shah, S.H., 2018. Applying artificial intelligence to identify physiomarkers predicting severe sepsis in the PICU. Pediatric Critical Care Medicine, 19(10), pp.e495-e503.
Thank you very much for pointing this out. These references have been added.

Methods:
1. In the developmental thrusts section I would like to see a comparison to similar approaches that have become standardized for analyzing high-frequency data streams, please see above references for more detail.
This hgas been added.

2. It is still not clear whether these waveform features are extracted from peak annotations or the raw electrical signal, please confirm and expand as necessary. I.e. "… generate beat-tobeat HR, as well as the various diastolic and systolic pressure and oximetry signals" It is unclear whether the features are the numeric physiological values or some higher-dimensional abstraction.
This has now been clarified.

3. Please review the manuscript for undefined acronyms. i.e. AMOC, RIPR etc..
Modified as suggested.

4

4. A subsection on how you will validate your models for generalization is highly recommended. There is a scant reference to performance measures, stability principle checks, etc for reproducibility.
This was also requested by reviewer #3 and has now been clarified.

5. The GUI implementation is interesting, however, there should be some discussion about threats to external validity from using a small number of experts. Additional detail about these experts are also missing, i.e. are they from a single site, similar training background etc.?
6. Can you clarify whether experts required for the early phase of development? Could those initial developments be tested by a larger number of novices, i.e. mechanical turk etc.?
These two questions have now been discussed in our revised limitations section.

Reviewer: 2
Reviewer Name: Daniel I. Sessler
Institution and Country:
Michael Cudahy Professor and Chair
Department of Outcomes Research
Cleveland Clinic
9500 Euclid Ave — P77
Cleveland, OH 44195
Please state any competing interests or state 'None declared': I am a consultant for Edwards Lifesciences.

Please leave your comments for the authors below
The authors are authorities in physiological modeling and are well positioned to do the proposed research. If anyone can do the proposed research, they are the ones.
Thank you for these positive comments.

The general purpose of publishing protocols is to provide a detailed record of what is planned so reviewers of the eventual results report can confirm that authors are transparent about what was planned and what changed. My concern for this type of research is that it is inherently exploratory and it is obvious that the investigators will need to try many different approaches in the hopes of finding one that works. That is perfectly appropriate, but would seem to make publication of a protocol manuscript unnecessary.
We agree with the assessment from Dr. Sessler. We believe that is an intrinsic characteristic of this kind of big data / bioinformatics research. Other protocol focusing on machine learning development have been published in the BMJ Open which is why we were encouraged to submit our NIH funded project to the journal. Obviously, we leave the decision up to the editors and editorial office to decide whether this submissioi in line with the journal focus. We assume the major revision status implies it is.

Reviewer: 3
Reviewer Name: Bradley Fritz
Institution and Country: Washington University School of Medicine, United States
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
I am excited to read this protocol for a study where a prediction model for intraoperative cardiorespiratory instability will be trained using a retrospective cohort and then validated

prospectively. The authors also propose to create a user interface and to use clinician focus groups and simulated operating room scenarios to refine these interfaces. This is an exciting project, but I do have some questions that might make the protocol easier for the reader to understand.
Thank you very much for these positive comments.

1. My major question is how cardiorespiratory instability is defined. The authors mention that the UC Irvine dataset includes "clinical annotations;" is that where they plan to obtain the labels? Who created these annotations (the anesthesia team taking care of the patient, previous research teams, hospital administration?) and what criteria were used for making annotations? How will label information for the target variable be obtained in the prospective phase?
Thank you very much for this comment. This has now been explained extensively in the revised application.

2. Are the authors planning a model to predict the dichotomous target of cardiorespiratory instability, or will they be predicting a multi-class target of different types (causes?) of cardiorespiratory instability? If the latter is true, then how will the classes be defined?
For this specific proposal we e plan on predicting dichotomous targets of CRI. This has been clarified.

3. Page 7 includes several concrete references to "time since start of bleed" (line 23) and "intraoperative bleeding" (line 45). Is this one of the classes for the target?
No, these are examples of our preliminary work and approaches that we plan on using to predict CRI. This has been clarified.

4. What will be the methods for the external validation of the model at UCLA and UPMC? In what time period will data be collected? How will model performance be measured?
This has been added to the new revised version.

5. In what way will the model be used to "guide resuscitation" (page 9, line 16)?
This has been clarified.

6. In the simulations, what does "accuracy of intervention choices" mean? There are usually many appropriate ways to respond to a given clinical scenario.
This has been explained and clairifed.

VERSION 2 – REVIEW

REVIEWER	Rishi Kamaleswaran Emory University, USA
REVIEW RETURNED	11-Oct-2019

GENERAL COMMENTS	The revised manuscript is acceptable for publication.

REVIEWER	Bradley Fritz Washington University School of Medicine United States
REVIEW RETURNED	24-Sep-2019

GENERAL COMMENTS	I am pleased to read this revised protocol. This study promises to be very interesting. The authors have addressed my minor comments in this revision, but they have not addressed my major comment. The revised manuscript still does not provide a definition

	for the target variable, cardiorespiratory instability. Please define this term.