Title

Comparison of differences in cohort (forwards) and case control (backwards) methodological approaches for validation of the Hypotension Prediction Index.

Authors

Davies, Simon J

Sessler, Daniel I

Jian, Zhongping

et al.

Publication Date

2024-04-01

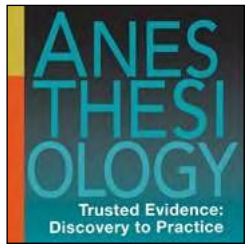
DOI

10.1097/aln.0000000000004989

Peer reviewed

ANESTHESIOLOGY
Trusted Evidence: Discovery to Practice®

Journal Pre-proof



Comparison of differences in cohort (forwards) and case control (backwards) methodological approaches for validation of the Hypotension Prediction Index

Simon J Davies, FRCA, MD, Daniel I Sessler, MD, Zhongping Jian, PhD, Neal W Fleming, MD, PhD, Monty Mythen, FRCA, MD, Kamal Maheshwari, MD, Denise P Veelo, MD, PhD, Alexander PJ Vlaar, MD, PhD, Jos Settels, MS, Thomas Scheeren , MD, PhD, BJP van der Ster, MD, PhD, Michael Sander, MD, PhD, Maxime Cannesson, MD, PhD, Feras Hatib, PhD.

DOI: https://doi.org/10.1097/ALN.0000000000004989

To appear in: Anesthesiology

Submitted for publication: April 12, 2023

Accepted for publication: March 18, 2024

Please cite this article as: Davies SJ, Sessler DI, Jian Z, Fleming NW, Mythen M, Maheshwari K, Veelo DP, Vlaar APJ, Settels J, Scheeren T, van der Ster BJP, Sander M, Cannesson M, Hatib F: Comparison of differences in cohort (forwards) and case control (backwards) methodological approaches for validation of the Hypotension Prediction Index. Anesthesiology. 2024; https://doi.org/10.1097/ALN.0000000000004989

Anesthesiology Publish Ahead of Print

DOI: 10.1097/ALN.0000000000004989

Comparison of differences in cohort (forwards) and case control (backwards) methodological approaches for validation of the Hypotension Prediction Index

Simon J Davies FRCA, MD[1,2], Daniel I Sessler MD[3], Zhongping Jian PhD[4], Neal W Fleming MD, PhD[5], Monty Mythen FRCA[4], MD[6], Kamal Maheshwari MD[3], Denise P Veelo MD, PhD[7], Alexander PJ Vlaar MD, PhD[7], Jos Settels MS[4], Thomas Scheeren MD, PhD[4,8], BJP van der Ster MD, PhD[7,9], Michael Sander MD, PhD[10], Maxime Cannesson MD, PhD[11], Feras Hatib PhD[4]

[1]Department of Anaesthesia, Critical Care and Perioperative Medicine, York and Scarborough Teaching Hospitals NHS foundation Trust, York, UK

[2]Centre for Health and Population Science, Hull York Medical School, York, UK.

[3]Department of Outcomes Research, Cleveland Clinic, Ohia, USA

[4]Edwards Lifesciences, Irvine, California, USA

[5]UC Davis School of Medicine, California, USA

[6]UCL/UCLH National Institute of Health Research Biomedical Research Centre, London, UK

[7]Departments of Anaesthesia and Intensive Care, Amsterdam University Medical Center, Amsterdam, The Netherlands

[8]Department of Anesthesiology, University Medical Centre Groningen, The Netherlands.

[9]ErasmusMc, Rotterdam, The Netherlands

[10]Department of Anaesthesiology, Intensive Care Medicine and Pain Medicine, University Hospital Giessen, Giessen, Germany.

[11]Department of Anesthesiology and Perioperative Medicine, UCLA, California, USA

Corresponding Author: Simon J Davies, Centre for Health and Population Science, Hull York Medical School, York, UK. simon.davies@york.ac.uk

Clinical trial number and registry URL" Not applicable.

Prior Presentation.: Not applicable.

Acknowledgements.: Not applicable.

The article processing charge was funded by Edwards Lifesciences.

Words and element count.

Abstract 280 words

Introduction 322 words

Discussion 1677 words

Number of figures 5

Number of tables 0

Supplementary digital files 4

Abbreviated title: Hypotension Prediction Index revalidation

Summary statement: Not applicable.

Funding Statement: Support was provided solely from institutional and department funds.

Conflicts of interests

SD, NF, DV, KM, DS, MS, MC consultancy fees, speaker fees and funded research from Edwards Lifesciences.

MM paid consultancy and speaker fees from Edwards Lifesciences. As of August 1st, 2023 MM is a paid employee of Edwards Lifesciences. MM has recently resigned as a founding Director of the Perioperative Quality Initiative, and is also a Board Member of Evidence Based Perioperative Medicine CiC and co-editor in chief of TopMedTalk.

FH, ZJ, TS and JS are paid employees of Edwards Lifesciences.

TS has received consultancy fees, speaker fees and funded research from Masimo Inc.

AP has received consulting fees from Edwards Lifesciences (paid to institution) and Philips.

MC is the founder of Sironis and Perceptive Medical and owns patents and receives royalties for closed loop hemodynamic management technologies that have been licensed to Edwards Lifesciences.

Data Availability Data will be available from Edwards Lifesciences upon reasonable written request including IRB approval. Analysis script can be found at https://github.com/InstabilityPrediction/Analysis2024.

Abstract

Background

The Hypotension Prediction Index (the index) software is a machine learning algorithm that detects physiological changes that may lead to hypotension. The original validation used a case control (backwards) analysis that has been suggested to be biased. We therefore conducted a cohort (forwards) analysis and compared this to the original validation technique.

Methods

We conducted a retrospective analysis of data from previously reported studies. All data were analysed identically with 2 different methodologies and receiver operating characteristic curves (ROC) constructed. Both backwards and forwards analyses were performed to examine differences in area under the ROC for HPI and other haemodynamic variables to predict a MAP < 65mmHg for at least 1 minute 5, 10 and 15 minutes in advance.

Results

Two thousand and twenty-two patients were included in the analysis, yielding 4,152,124 measurements taken at 20 second intervals. The area-under-the-curve for the index predicting hypotension analysed by backward and forward methodologies respectively was 0.957 (95% CI, 0.947–0.964) vs 0.923 (95% CI, 0.912–0.933) 5 minutes in advance, 0.933 (95% CI, 0.924–0.942) vs 0.923 (95% CI, 0.911–0.933) 10 minutes in advance , and 0.929 (95% CI, 0.918–0.938) vs. 0.926 (95% CI, 0.914–0.937) 15 minutes in advance. No other variable had an area-under-the-curve > 0.7 except for MAP. Area-under-the-curve using forward analysis for MAP predicting hypotension 5, 10, and 15 minutes in advance was 0.932 (95% CI, 0.920–0.940), 0.929 (95% CI, 0.918–0.938), and 0.932 (95% CI, 0.921–0.940). The R^2 for the variation in the index due to MAP was 0.77.

Conclusion

Using an updated methodology, we found the utility of the HPI index to predict future hypotensive events is high, with an area under the receiver-operating-characteristics curve similar to that of the original validation method.

Introduction

The Acumen Hypotension Prediction Index ('the index') software is a logistic regression machine-learning based algorithm that detects early hemodynamic instability that may lead to hypotensive events defined as MAP of < 65mmHg for at least 1 minute. The algorithm is a probabilistic model that analyses multiple arterial pressure waveform features and their inter-relationships to detect physiological alterations in central compensatory mechanisms that precede episodes of hypotension. The index has been shown to be an accurate predictor of future hypotensive events although studies linking hypotension reduction to improved post operative outcomes currently remain lacking.[1-5]

The methods used for the index validation[5] have been questioned by Enevoldsen and colleagues[6] who also suggested that the concurrent value MAP may predict hypotension as well as the index. Based on simulated data, they postulated that the performance may be overestimated due to a selection bias in the definition of hypotensive and normotension. A hypotensive event allowed for the full range of preceding MAP values to be used in model training dataset. The same did not apply for the normotension which was defined as the midpoint of a continuous 30-min episode where MAP was consistently greater than 75 mmHg. Enevoldsen concluded that a data sample with a MAP < 75mmHg will therefore always correspond to a future hypotensive event. The accompanying editorial[7] advocated for a validation including data for the full range of values for predictor variables, and that the analysis should mimic the flow of data seen in clinical practice i.e. a cohort design using a forward looking approach, rather than the backwards case control approach previously used.

Studies that previously used a forward methodology to validate the index[4,8,9] with similar results to the original validation, however, were not explicit in how data were selected. We therefore provide a direct comparison of the original backwards validation of the index software with the suggested forward methodology and address the issue if concurrent MAP values can predict future hypotension.

Methods

We conducted a retrospective analysis of prospectively gathered anonymized data. This manuscript adheres to the Enhancing the Quality and Transparency of Health Research guidelines[10]. Data were analysed from subjects from 9 previously reported studies[2-5,8,9,11-13]. Institutional review board approval was not required as we analysed pre-existing anonymised data. Analyses were based on invasive arterial waveforms via radial arterial cannulation, and non-invasive arterial waveforms derived from a finger cuff (ClearSight, Edwards Lifesciences, Irvine, CA, USA). Detailed inclusion and exclusion criteria are presented in the individual reports, and a summary is provided in supplemental material table S1 (https://links.lww.com/ALN/D526). Briefly, all studies enrolled adults over 18 years of age. Two studies collected non-invasive measurements in subjects undergoing non cardiac surgery[2,4]. Five studies analysed invasive blood pressure measurements in patients undergoing moderate to major surgery; one included elective total hip arthroplasties[12], two studies included elective major non cardiac surgery only[8,13], one study included elective cardiac surgery requiring cardiopulmonary bypass only[3], and one study included a mix of elective major non cardiac surgery and off pump cardiac surgery[1]. One study included patients in the intensive care unit (ICU) with a diagnosis of COVID[9] and one study included a mixed group containing predominately ICU patients but also mixed and surgery that included cardiac, neurosurgical, general, vascular and thoracic surgeries[5].

From the collected data, variables were computed including both the index and mean arterial pressure (MAP). All collected variables are shown in supplementary data file 1 (https://links.lww.com/ALN/D527) and supplementary data file 2 (https://links.lww.com/ALN/D528). Our analysis was based on 20-second averages of beat-to-beat values. Poor arterial waveform signals (e.g., line flushing, significantly damped waveforms, and other waveform artifacts) were detected by the arterial beat detection algorithm and excluded from the analysis.1.1% of data was categorized as unreliable for invasive A-line data, and 2.1% for non-invasive A-line data equally distributed across all studies.

Statistical Methods

Descriptive statistics are presented as mean (SD) for normally distributed continuous data, or median (25^{th}–75^{th} percentiles) for non-normally distributed data. Normality of data was assessed by using the Shapiro-Wilk statistic. The data set was analysed in its entirety to describe the total number of hypotensive events (MAP < 65 mmHg for at least a minute measured either invasively or non-invasively), absolute duration of hypotension, area under the threshold of 65 mmHg, and time-weighted average of area under the threshold. HPI and MAP data were plotted to assess their relationship, and the coefficient of determination was calculated to determine the proportion of the variation in the dependent variable that is predictable from the independent variable. In addition, a reclassification table was created to compare the performance of HPI and MAP for the prediction of future MAP < 65mmHg (full details in supplementary material, https://links.lww.com/ALN/D526).

ROC analysis was used to evaluate the performance of variables in predicting hypotension 5, 10, and 15 minutes prior to the hypotensive event. The ROC analysis was performed with both a backwards analysis with dual boundary conditions (grey zone), and a forward analysis using all

data points. Analysis was performed separately for invasive arterial waveform and non-invasive waveform data. We also performed ROC analysis on MAP, SV, and HR at various thresholds to assess the effect of ROC analysis on time series data to predict future values.

No data were available to determine if potential interventions occurred to treat or prevent hypotension, which could include surgical incision, change in position or vasopressor treatment amongst others. Treatment interventions were assumed when MAP increased more than 5 mmHg within 20 seconds (most probably caused by vasopressor or inotropic injections) or when MAP increased more than 8 mmHg within 2 minutes (change in vasopressor or inotropic infusion rate or fluid bolus) when the MAP was <75 mmHg. As sensitivity analyses for the effects of excluding presumed treatment periods, we also conducted a forward analysis with a threshold of 10 mmHg MAP change to define an intervention, as well as an analysis with no interventions censored. We used bootstrapping to account for repeated measurements from each subject for the calculation of confidence intervals.[14]

Analysis method 1 – Backwards analysis.

ROC analysis was performed to evaluate the performance of variables to predict a future hypotensive event including the index, mean arterial pressure, and ΔMAP at different blood pressure thresholds. ΔMAP analysis was performed in the MAP ranges of 75–85, and 85–95 mmHg. ΔMAP was the difference in 2 MAP measurements 3 minutes apart. Our methodology has been described previously[5]. In summary:

- A hypotensive event was identified as a section of at least 1-min duration, with a MAP < 65 mmHg for all three 20-second averages. A positive data point was chosen as the data point at 5, 10, or 15 min before the hypotensive event. All positive data points were included in the analysis regardless of their MAP values.

- A non-hypotensive event was identified as a 30-min continuous section of data points at least 20 min apart from any hypotensive event and with a MAP > 75 mmHg for all data points in that section. A negative data point was chosen as the centre data point of the non-hypotensive event to reduce intraclass correlation. Only one data point within a 30-min window was selected since selecting all the data points within the window would have introduced statistical bias since neighbouring 20-s data points are highly correlated.
- This approach incorporates a 'grey zone' between a MAP of 65 and 75 mmHg and values within this range were excluded. This is a zone of a quantitative test in which there is an area of values where the discriminatory performance is 'insufficient', in the sense that a value in the grey zone does not allow the target disease to be scored as either present or absent[15]. A flow diagram illustrates the backward methodology in figure 1.

Analysis method 2 – Forward analysis.

We used ROC analyses to evaluate the performance of the same variables as in the backwards analyses. As previously, a hypotensive event was defined by < 65 mmHg persisting at least 1 minute, with all other pressures being considered non-hypotensive events. In this analysis, every data point except those that were during hypotensive events or that had an unreliable arterial line trace were examined. Every prediction index data point was compared to a threshold (for example 85), and a window of 5, 10, or 15 minutes immediately following the data point was searched to identify if there were hypotensive events. No grey zones were used in this analysis. As in prior work of assessing the performance of real-time and continuous risk scores[4,9,16-19], the analysis sequence was as follows: A true positive (TP) prediction occurred when the index was above the designated threshold and there was a hypotensive event in the search window. A false positive (FP) prediction occurred when the index was above the designated threshold and there

was no hypotensive event in the search window. A true negative (TN) prediction occurred when the index was below the designated threshold and there was no hypotensive event in the search window. A false negative (FN) prediction occurred when there was a hypotensive event that the index failed to predict at least 2 minutes in advance. The above process was repeated so that the designated threshold covers all possible values of the index. Similar analyses were also performed for other variables. A flow diagram illustrates the forward methodology in figure 2, similar to published clinical analyses of other real-time and continuous risk scores[4,9,16-19].

The index vs. MAP analysis

The index vs. MAP analysis was performed to assess the relationship between the index and MAP. All paired data points of the index and MAP from all patients were analysed together, then for each integer value of MAP we calculated the mean, standard deviation, minimum, and maximum of the corresponding index values. Similarly, we calculated the mean, standard deviation, minimum, and maximum of corresponding MAP values for each integer value of the index.

All statistics were performed with MATLAB (version R2021a; The MathWorks Inc, Natick, MA). Data is available upon reasonable request and IRB approval. Analysis code can be found in supplemental material.

Results

Two thousand and twenty-two patients (913 women, 1109 men) with a mean age of 61 (14) years were included in the analysis. Based on 20-second recordings epochs, there were a total of 4,152,124 data points. Among the enrolled patients, 339 (17%) were from an Intensive Care Unit (ICU) and 1683 (83%) were intraoperative surgical patients. Data were obtained from an arterial catheter in 1210 patients (60%) and from a ClearSight non-invasive cuff in 812 patients (40%).

The median monitoring time per patient was 230 minutes (IQR 144 - 420 minutes), and 1683 of 2022 patients (83%) had at least 1 hypotensive event defined as a MAP <65 mmHg for ≥1 minute. In total, 24,654 hypotensive events were detected, and the median number of events per patient was 4 (1–10) with a median duration of 2 minutes (1–5) per event. The median cumulative duration of hypotension per patient was 13.3 minutes (2.7–44.0), which was 5.5% (1.0%–16.4%) of total monitoring time. The median area under the threshold of 65 mmHg was 79.9 mmHg min (15–271 mmHg min) and the median time-weighted average area under the threshold was 0.30 mmHg (0.05–0.96 mmHg).

A plot of all paired index and MAP data measured at the same time points is shown in figure 3. The R^2 for the variation in the index due to MAP was 0.77. Mean, SD, minimum and maximum values of the index for different MAP thresholds and vice versa are shown in supplemental material table S2 (https://links.lww.com/ALN/D526).

Receiver operating curve analysis.

i. Comparison of a backwards versus forwards analysis for the index to predict hypotension.

ROC curves for the index predicting hypotension 5, 10, and 15 minutes in advance by backwards and forwards analysis are shown in figure 4 separated for invasive and non-invasive arterial waveform analysis.

For invasive arterial waveform data, the AUC for the index predicting hypotension by backwards and forwards methodology respectively was 0.957 (95% CI, 0.947–0.964; sensitivity, 86%; specificity 95%) using vs. 0.923 (95% CI, 0.912–0.933; sensitivity, 87%; specificity 84%) 5 minutes prior to the event, 0.933 (95% CI, 0.924–0.942; sensitivity, 81%; specificity 93%) vs. 0.923 (95% CI 0.911-0.933; sensitivity, 88%; specificity 83%) 10 minutes prior to the event, and

0.929 (95% CI, 0.918–0.938; sensitivity, 81%; specificity 92%) vs. 0.926 (95% CI, 0.914–0.937; sensitivity, 87%; specificity 84%) 15 minutes prior to the event.

For non-invasive arterial waveform data, the AUC for the index predicting hypotension by backwards and forwards methodology respectively was 0.950 (95% CI, 0.939–0.961; sensitivity, 85%; specificity 94%) versus 0.917 (95% CI, 0.908–0.925; sensitivity, 88%; specificity 84%) 5 minutes prior to the event; 0.930 (95% CI, 0.914–0.944; sensitivity, 81%; specificity 93%) versus 0.918 (95% CI, 0.906–0.928; sensitivity, 87%; specificity 86%) 10 minutes prior to the event, and 0.912 (95% CI, 0.893–0.929; sensitivity, 79%; specificity 91%) versus 0.923 (95% CI, 0.910–0.933; sensitivity, 88%; specificity 86%) 15 minutes prior to hypotension.

Full details of all analyses for all haemodynamic variables including area under the curve, area under the precision recall curve, sensitivity, specificity, positive predictive value, negative predictive value, and optimal cut off values are shown in supplemental data file 1 (backwards analysis, https://links.lww.com/ALN/D527), supplemental data file 2 (forwards analysis, https://links.lww.com/ALN/D528).

ii. ROC analysis using hemodynamic variables to predict future values in time series data.

The AUC using forward analysis for MAP predicting a MAP of < 65mmHg for at least 1 minute 5, 10, and 15 minutes in the future was 0.932 (95% CI, 0.920–0.940; sensitivity, 86%; specificity 86%), 0.929 (95% CI, 0.918–0.938; sensitivity, 87%; specificity 85%), and 0.932 (95% CI, 0.921–0.940; sensitivity, 88%; specificity 84%).

Figure 5 shows ROC curves for SV, HR and MAP predicting a decrease in future values below various thresholds 10 minutes into the future for invasive arterial waveform data only. All variables have a high AUC in predicting their future values regardless of the threshold chosen.

MAP predicted its future value 10 minutes in advance of being less than 75 mmHg with an AUC of 0.908 (95% CI, 0.898-0.917), less than 70 mmHg with an AUC of 0.921 (95% CI, 0.911-0.929) and less than 65 mmHg with an AUC 0.929 (95% CI, 0.911-0.929). SV predicted its future value being less than 60 ml with an AUC of 0.924 (95% CI, 0.917 – 0.932), and less than 50 ml with an AUC of 0.929 (95% CI, 0.920 – 0.936). HR predicted its future value 10 minutes in advance, with an AUC of being less than 60 bpm of 0.973 (95% CI, 0.967 – 0.977) and an AUC for less than 50 bpm of 0.980 (95% CI, 0.968 – 0.987).

Area under the curve, sensitivity, specificity, positive predictive value, negative predictive value, and optimal cut off for SV, HR and MAP predicting their future values 10 minutes into the future at different thresholds are shown in supplemental data file 3 (https://links.lww.com/ALN/D529).

A net reclassification improvement is shown in supplementary material table S3 (https://links.lww.com/ALN/D526) with a 45% net improvement in false positives compared to false negatives with the index compared to MAP assigning a 10:1 ratio to emphasise the importance of a non-event and a 7.6% improvement assuming equal importance[20].

Discussion

Using a forward methodology in the analysis of the index to predict hypotension defined as a MAP < 65mmHg for at least 1 minute, there was a significant reduction in the AUC compared to our previously reported backwards methodology at 5 minutes only, however the AUC remained greater than 0.9[1,5]. Thus, the utility of the index to predict future hypotensive events remained excellent.

The original validation used a backward (case-control methodology) in which hypotensive events were identified, and then looked backwards in time to see if a variable predicted the event. This method means that, unlike the forwards approach, interventions that avoid hypotension do

not need to be identified. Using a forward approach, due to the lack of contemporaneous information on clinical treatment, potential interventions must be accounted for by predefined rapid changes in MAP which is a potential weakness. Future studies aimed at validating predictive algorithms will require highly annotated data on interventions to avoid this confounder. However, it is a fair criticism that using a backwards analysis does not represent how clinicians utilise data, and that a forward methodology better mimics the flow of data in clinical practice.

As part of the backwards methodology a 'grey zone' or zone of uncertainty was used in which hypotensive events were defined as a MAP < 65mmHg, and non-hypotensive events as > 75mmHg. These definitions were selected because they span the most common range of reported harm thresholds for acute kidney injury (AKI) and myocardial injury after non-cardiac surgery i.e. have a higher prevalence below the lower threshold and an increased threshold above the higher threshold.

The use of an uncertainty zone is common in clinical care because there is often ambiguity about disease states, especially at early stages. This type of ambiguity exists in many clinical areas with most measured clinical parameters since physiology is not a binary process, but rather an ensemble of complex and fuzzy processes.[21] For example, such areas exist in labelling of definitive sepsis and definitive absence of sepsis[22], between clearly normal and clearly abnormal levels of blood pressure, and in the labelling of brain natriuretic peptide data for diagnosis of heart failure[15]. Using a zone of uncertainty was proposed to evaluate biomarkers[23]. This concept is common in machine learning, and as machine learning applications expand more in the medical field, using a zone of uncertainty concept in their development and validation will

become critical for the clinical adoption of such technologies. In a sensitivity analysis, exclusion of the uncertainty zone did not significantly alter the predictive ability of HPI.

Enevoldsen,[6] in a simulated data set, suggested that a selection bias was present due to the definition of a non-hypotensive event, and that this bias would cause an overestimation of the predictive ability of the index. In the original validation study, hypotension was defined by a MAP of less than 65 mmHg for a minute using the full range of preceding MAP values. A non-hypotensive event was identified at the midpoint of a 30-min continuous section of data points at least 20 min away from any hypotensive event and with a MAP > 75 mmHg for all data points in that section. Enevoldsen also concluded that the index was therefore taught that if the MAP is < 75 mmHg then the only possible future event is hypotension. This is not how the index was trained, and hence it does not perform as proposed. At a MAP of 75mmHg the index can take any value from 7 (low probability of hypotension) to 98 (high probability of hypotension); therefore, it does not assume hypotension is the only possible outcome. Using all data points (no selection bias) resulted in only a small reduction in the AUC, however there is the possibility that the predictive ability is overestimated or underestimated by using points that have high intraclass correlation, hence the original methodology that avoided this, also allowing for a balanced ROC analysis. Area under the precision recall curve (AUPRC) can be used for analysis of unbalanced classes, however the methodology uses the Positive Predictive Value (PPV). PPV is not a direct metric to demonstrate performance of algorithms as it is primary impacted by the prevalence of the hypotension. If prevalence is low, then PPV is low even if the algorithm has extraordinarily high performance of sensitivity and specificity.

Various investigators have asserted that the index is simply a mirror of MAP[24,25], and that MAP itself can be used to predict and treat hypotension. Firstly, whilst there is an average non-linear relationship between MAP and the index, there is a difference when looking at the individual samples. The 2 variables convey different information and the relationship between the index and MAP not only changes within individuals at different time points, but also varies between different individuals [26]. Neither MAP, systolic pressure, nor diastolic pressure are direct features in the index model. All 23 features of the index model are mathematically combinatorial features and represent combinatorial interaction effects and not static effects. There is no single dominant feature that influences the index, as the median influence values are similar.

Using current MAP to avoid hypotension is clearly an obvious solution and represents current practice. However, MAP has been available for hemodynamic management for a considerable period and yet hypotension remains common[27]. Whilst it is appealing to think the associations of IOH and poor outcomes has changed practice, recent studies have shown that the prevalence of hypotension remans high across all settings, surgery types and patient groups[28,29]. It is however a reasonable question to ask whether current MAP predicts future hypotension as well as a complex proprietary algorithm.

A simple linear extrapolation of MAP predicts future hypotension better than the change in MAP in patients having high-risk non-cardiac surgery,[30] although the index was not compared in that analysis. Using the absolute value of MAP to predict future hypotension, rather than linear extrapolation or delta MAP, gives a high ROC curve AUC as in this analysis. However, this reflects a statistical relationship that arises when using any variable as both the independent and dependent variable as is shown in this analysis (figure 4) where SV, HR and MAP all predict future values at any threshold in their range with a high AUC. Mathematical coupling of data in

statistical analyses has been previously described[31]. Due to succession of numbers in time-series, any variable has high ROC when compared to a fixed threshold representing a number in the range of the times-series variable. When using the current MAP to predict a MAP < 65mmHg in the future, the ROC analysis simply illustrates the likelihood that a MAP value of 65 mmHg will be preceded by other MAP values (e.g., a MAP of 67 or 70 mmHg), which is intuitive since to reach a hypotensive event from a higher MAP, the blood pressure must pass through these values.

If it is assumed that MAP can predict itself, then there are additional clinical considerations. The optimal cut off to predict a MAP <65 mmHg would be 72 mmHg, therefore this value should be always treated to avoid future hypotension as has been suggested.[6,25] This is an extrapolation of a statistical analysis that does not translate into reasonable clinical practice. Clinicians do not treat an absolute value of MAP without examining the prior trajectories. If a threshold of 72 mmHg is taken as an absolute treatment threshold this would lead to over treatment. In our dataset taking every MAP of 72 mmHg in the 1210 patients monitored with an arterial catheter, only 18% led to a MAP<65 mmHg within the next 15 minutes. Therefore 82% of the episodes would have been treated inappropriately as hypotension would not have occurred. This is of potential concern as increasing doses of vasopressors have been associated with worse outcomes[32,33]. The index should be used in combination with MAP to distinguish patients likely to subsequently develop hypotension from those who are unlikely to, as it is a probalistic model that provides that information.

The use of the index in clinical practice reduces hypotension compared to using a target MAP[12,13,34] in some although not all trials[8]. Education of clinicians on the importance of maintaining a target MAP threshold did not result in a reduction in hypotension despite them believing they had sufficient knowledge and skills. The use of the index reduced hypotension and had perceived clinical utility[35].

There are several limitations to this analysis. Firstly, this is a retrospective analysis and data on interventions were not available. Interventions such as certain drug administration, change of ventilation or blood loss will predict hypotensive events, and the analysis does not account for these interventions in the operating room or the ICU. Future studies should include these data and need to evaluate the effect of using the index to minimise hypotension focusing on clinical outcomes such as AKI, MINS or a composite outcome rather than numerical measurements of hypotension depth and duration. We thus had to make assumptions about interventions based on MAP patterns. However, results were similar over a wide range of assumptions including no adjustment for any interventions. Secondly, some clinicians had access to the index data which presumably resulted in some potential future hypotensive episodes being treated before they occurred. However, treatment would generate false positive predictions, making the index perform worse. Fully addressing this issue will require a dataset, presumably obtained in the context of a trial, in which all interventions are well documented.

As predictive tools become more prevalent in clinical practice, consensus will be needed regarding the best methodology for validation of complex algorithms that considers the impact of both statistical bias and clinical utility. But in the meantime, our analysis shows that the index software accurately predicts hypotensive events regardless of the validation methodology used.

Supplemental Digital Content

Supplementary material: Additional analyses and tables, https://links.lww.com/ALN/D526

Supplementary table 1: Patient characteristics

Supplementary table 2: Corresponding HPI and MAP values at different thresholds.

Supplementary table 3: Reclassification Table

Supplementary data file 1: Backwards methodology full analysis,

https://links.lww.com/ALN/D527

Supplementary data file 2: Forwards methodology full analysis,

https://links.lww.com/ALN/D528

Supplementary data file 3: Analysis of HR, SV, and MAP to predict future values,

https://links.lww.com/ALN/D529

References

1. Davies SJ, Vistisen ST, Jian Z, Hatib F, Scheeren TWL: Ability of an Arterial Waveform Analysis-Derived Hypotension Prediction Index to Predict Future Hypotensive Events in Surgical Patients. Anesth Analg 2020; 130: 352-359

2. Maheshwari K, Buddi S, Jian Z, Settels J, Shimada T, Cohen B, Sessler DI, Hatib F: Performance of the Hypotension Prediction Index with non-invasive arterial pressure waveforms in non-cardiac surgical patients. J Clin Monit Comput 2021; 35: 71-78

3. Shin B, Maler SA, Reddy K, Fleming NW: Use of the Hypotension Prediction Index During Cardiac Surgery. J Cardiothorac Vasc Anesth 2021; 35: 1769-1775

4. Wijnberge M, van der Ster BJP, Geerts BF, de Beer F, Beurskens C, Emal D, Hollmann MW, Vlaar APJ, Veelo DP: Clinical performance of a machine-learning algorithm to predict intra-operative hypotension with noninvasive arterial pressure waveforms: A cohort study. Eur J Anaesthesiol 2021; 38: 609-615

5. Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, Rinehart J, Cannesson M: Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis. Anesthesiology 2018; 129: 663-674

6. Enevoldsen J, Vistisen ST: Performance of the Hypotension Prediction Index May Be Overestimated Due to Selection Bias. Anesthesiology 2022; 137: 283-289

7. Wijeysundera DN, McIsaac DI, London MJ: The Promise and Challenges of Predictive Analytics in Perioperative Care. Anesthesiology 2022; 137: 275-279

8. Maheshwari K, Shimada T, Yang D, Khanna S, Cywinski JB, Irefin SA, Ayad S, Turan A, Ruetzler K, Qiu Y, Saha P, Mascha EJ, Sessler DI: Hypotension Prediction Index for

Prevention of Hypotension during Moderate- to High-risk Noncardiac Surgery. Anesthesiology 2020

9. van der Ven WH, Terwindt LE, Risvanoglu N, Ie ELK, Wijnberge M, Veelo DP, Geerts BF, Vlaar APJ, van der Ster BJP: Performance of a machine-learning algorithm to predict hypotension in mechanically ventilated patients with COVID-19 admitted to the intensive care unit: a cohort study. J Clin Monit Comput 2022; 36: 1397-1405

10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP: Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bmj 2007; 335: 806-8

11. Davies S, Jian Z, Hatib F, Gomes A, Mythen M: Indicators of haemodynamic instability and left ventricular function in a porcine model of esmolol induced negative inotropy. Journal of Clinical Monitoring and Computing 2022: 1-9

12. Schneck E, Schulte D, Habig L, Ruhrmann S, Edinger F, Markmann M, Habicher M, Rickert M, Koch C, Sander M: Hypotension Prediction Index based protocolized haemodynamic management reduces the incidence and duration of intraoperative hypotension in primary total hip arthroplasty: a single centre feasibility randomised blinded prospective interventional trial. Journal of Clinical Monitoring and Computing 2019; 124: 1 - 10

13. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, Schenk J, Terwindt LE, Hollmann MW, Vlaar AP, Veelo DP: Effect of a Machine Learning–Derived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery. JAMA: The Journal of the American Medical Association 2020: 1 - 9

14. Liu H, Li G, Cumberland WG, Wu T: Testing Statistical Significance of the Area under a Receiving Operating Characteristics Curve for Repeated Measures Design with Bootstrapping. Journal of Data Science 2022; 3: 257-278

15. Coste J, Jourdain P, Pouchot J: A gray zone assigned to inconclusive results of quantitative diagnostic tests: Application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspneic patients. Clin Chem 2006; 52: 2229-35

16. Bedoya AD, Futoma J, Clement ME, Corey K, Brajer N, Lin A, Simons MG, Gao M, Nichols M, Balu S, Heller K, Sendak M, O'Brien C: Machine learning for early detection of sepsis: an internal and temporal validation study. JAMIA Open 2020; 3: 252-260

17. Hyland SL, Faltys M, Huser M, Lyu X, Gumbsch T, Esteban C, Bock C, Horn M, Moor M, Rieck B, Zimmermann M, Bodenham D, Borgwardt K, Ratsch G, Merz TM: Early prediction of circulatory failure in the intensive care unit using machine learning. Nat Med 2020; 26: 364-373

18. Tomasev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, Ledsam JR, Mohamed S: A clinically applicable approach to continuous prediction of future acute kidney injury. Nature 2019; 572: 116-119

19. Wiens J, Guttag J, Horvitz E: Patient risk stratification with time-varying parameters: a multitask learning approach. J Mach Learn Res 2016; 17: 2797-819

20. Pencina MJ, D'Agostino RB, Vasan RS: Statistical methods for assessment of added usefulness of new biomarkers. Clin Chem Lab Med 2010; 48: 1703-11

21. Pinsky MR: Complexity modeling: identify instability early. Crit Care Med 2010; 38: S649-55

22. Vincent JL: The Clinical Challenge of Sepsis Identification and Monitoring. PLoS Med 2016; 13: e1002022

23. Ray P, Le Manach Y, Riou B, Houle TT: Statistical evaluation of a biomarker. Anesthesiology 2010; 112: 1023-40

24. Michard F, Futier E: Predicting intraoperative hypotension: from hope to hype and back to reality. Br J Anaesth 2023; 131: 199-201

25. Michard F, Biais M, Futier E, Romagnoli S: Mirror, mirror on the wall, who is going to become hypotensive? Eur J Anaesthesiol 2023; 40: 72-74

26. Ripollés-Melchor J, Ruiz-Escobar A, Fernández-Valdes-Bango P, Lorente JV, Jiménez-López I, Abad-Gurumeta A, Carrasco-Sánchez L, Monge-García MI: Hypotension prediction index: From reactive to predictive hemodynamic management, the key to maintaining hemodynamic stability. Frontiers in Anesthesiology 2023; 2: 1138175

27. Panjasawatwong K, Sessler DI, Stapelfeldt WH, Mayers DB, Mascha EJ, Yang D, Kurz A: A Randomized Trial of a Supplemental Alarm for Critically Low Systolic Blood Pressure. Anesth Analg 2015; 121: 1500-7

28. Saasouh W, Christensen AL, Chappell D, Lumbley J, Woods B, Xing F, Mythen M, Dutton RP: Intraoperative hypotension in ambulatory surgery centers. J Clin Anesth 2023; 90: 111181

29. Saasouh W, Christensen AL, Xing F, Chappell D, Lumbley J, Woods B, Mythen M, Dutton RP: Incidence of intraoperative hypotension during non-cardiac surgery in community

anesthesia practice: a retrospective observational analysis. Perioperative Medicine 2023; 12: 1-10

30. Jacquet-Lagrèze M, Larue A, Guilherme E, Schweizer R, Portran P, Ruste M, Gazon M, Aubrun F, Fellahi J-L: Prediction of intraoperative hypotension from the linear extrapolation of mean arterial pressure. European Journal of Anaesthesiology 2022; 39: 574-581

31. Archie JP: Mathematic Coupling of Data. Annals of Surgery 1981; 193: 296-303

32. Ariyarathna D, Bhonsle A, Nim J, Huang CK, Wong GH, Sim N, Hong J, Nan K, Lim AK: Intraoperative vasopressor use and early postoperative acute kidney injury in elderly patients undergoing elective noncardiac surgery. Renal Failure 2022; 44: 648-659

33. Chiu C, Fong N, Lazzareschi D, Mavrothalassitis O, Kothari R, Chen L-l, Pirracchio R, Kheterpal S, Domino KB, Mathis M: Fluids, vasopressors, and acute kidney injury after major abdominal surgery between 2015 and 2019: a multicentre retrospective analysis. British Journal of Anaesthesia 2022; 129: 317-326

34. Solares GJ, Garcia D, Monge Garcia MI, Crespo C, Rabago JL, Iglesias F, Larraz E, Zubizarreta I, Rabanal JM: Real-world outcomes of the hypotension prediction index in the management of intraoperative hypotension during non-cardiac surgery: a retrospective clinical study. J Clin Monit Comput 2023; 37: 211-220

35. de Keijzer IN, Vos JJ, Yates D, Reynolds C, Moore S, Lawton RJ, Scheeren TWL, Davies SJ: Impact of clinicians' behavior, an educational intervention with mandated blood pressure and the hypotension prediction index software on intraoperative hypotension: a mixed methods study. Journal of Clinical Monitoring and Computing 2023

Figure legends

Figure 1. Flow diagram to illustrate the methodology and assignment of true and false positive and negative data points for the backwards analysis.

Figure 2 Flow diagram to illustrate the methodology and assignment of true and false positive and negative data points for the forwards analysis.

Figure 3. Index versus MAP plot for all paired data points. Each circle is a paired data point, and the solid white line represents the mean of the index values for each distinct value of MAP, and whiskers the standard deviation. The red line indicates the region where the index has values greater than 85.

Figure 4.ROC curves for the index predicting hypotension 5, 10 and 15 minutes apart by backward (dual boundary grey zone) and forward analysis (single boundary condition no grey zone) separated for invasive and non-invasive arterial waveform analysis.

Figure 5.ROC curves for SV, HR and MAP predicting a decrease below various thresholds 10 minutes into the future using cohort analysis.

Figure 1

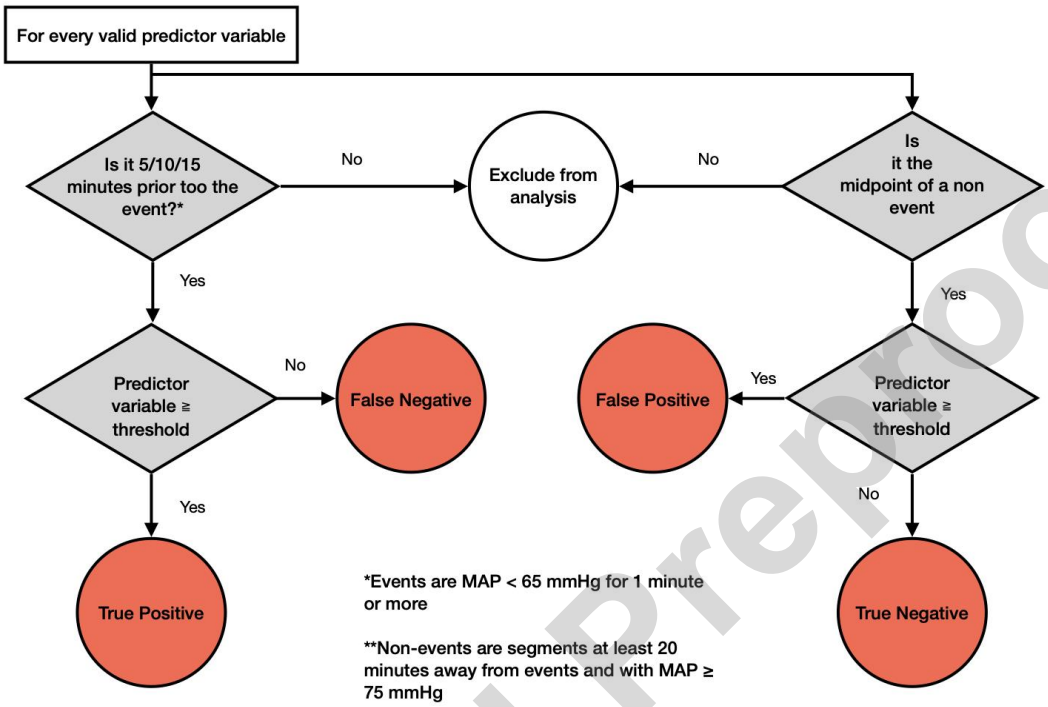


Figure 2

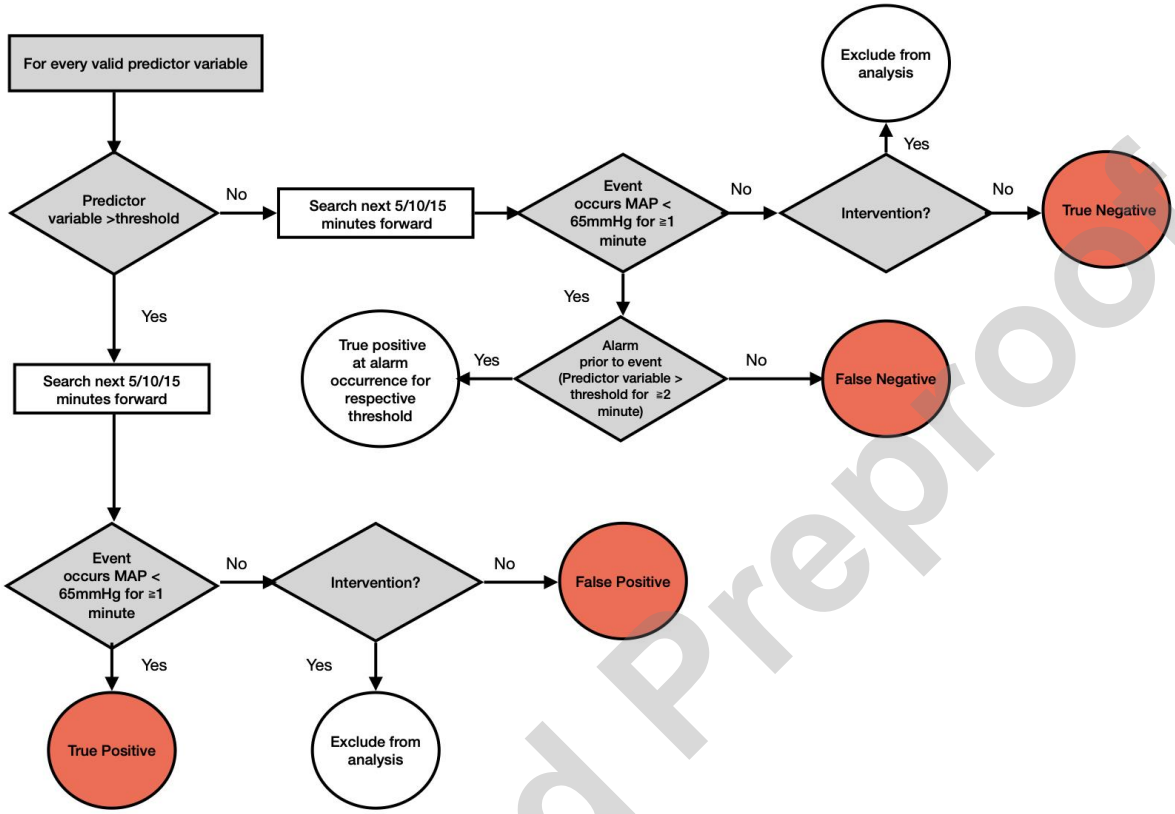


Figure 3

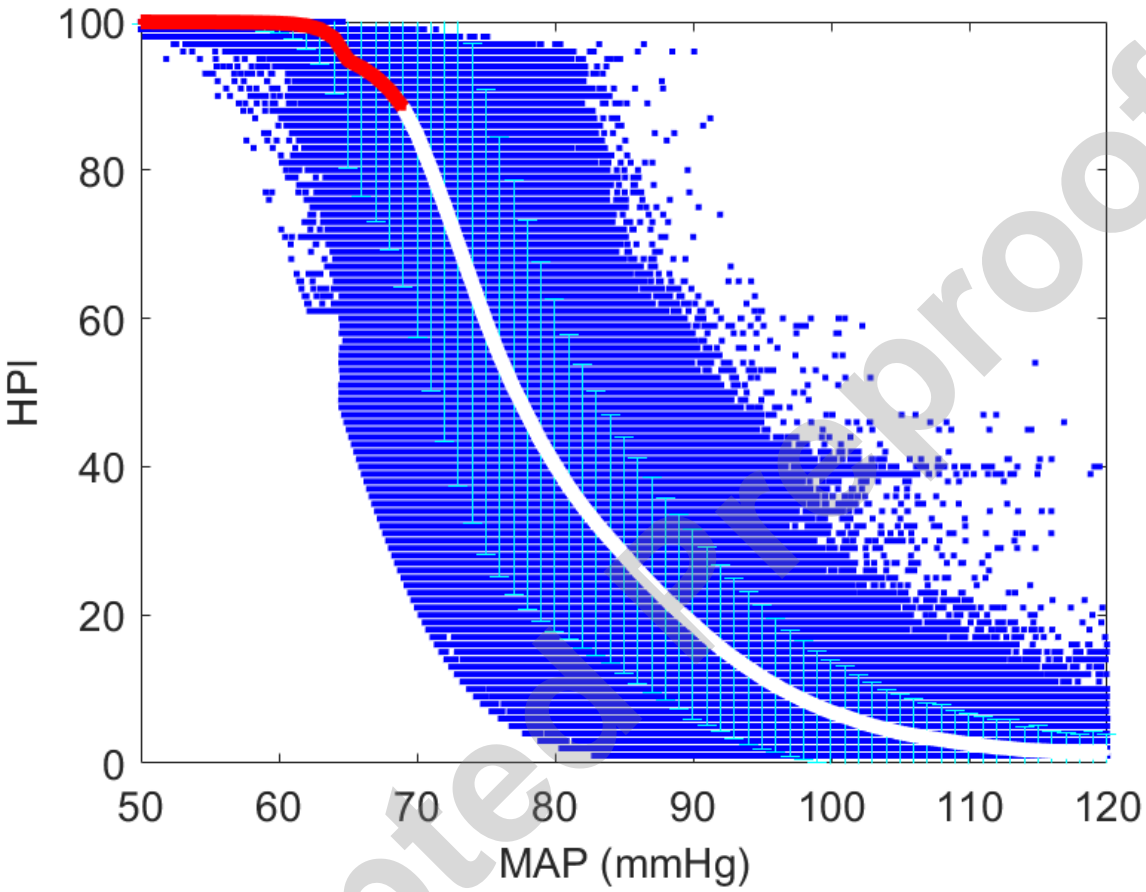


Figure 4

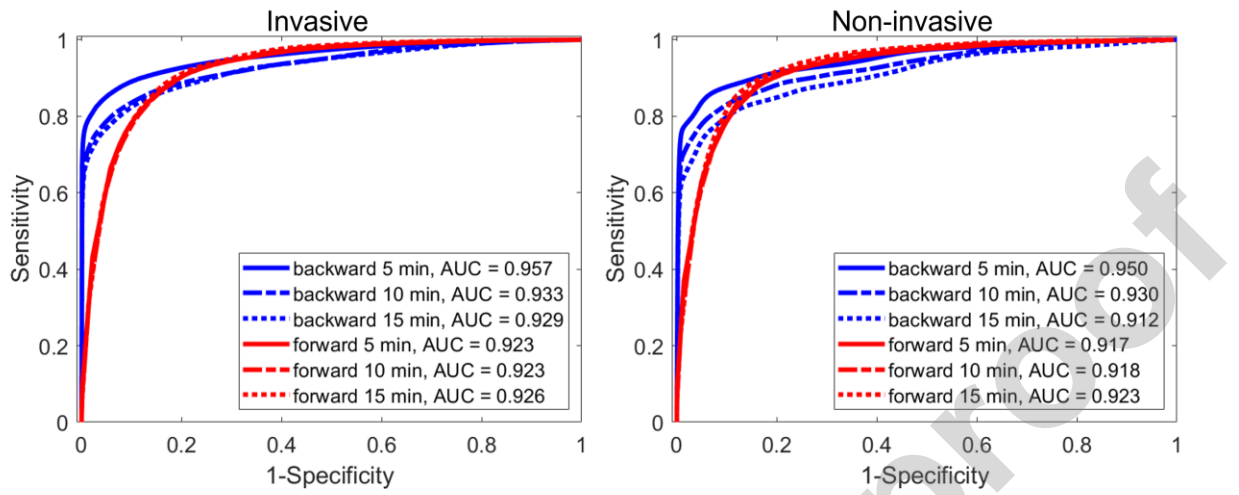


Figure 5