

Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension

Caleb W. Hug, Ph.D.

Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA

Gari D. Clifford, Ph.D.

Institute of Biomedical Engineering, University of Oxford, UK;

Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA

Andrew T. Reisner, M.D.

Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA;

Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA

Work performed at:

Massachusetts Institute of Technology, Cambridge, MA

Contact Information:

Caleb Hug

(206) 399-0637

2048 154Ave SE

chug@csail.mit.edu

Bellevue, WA 98007

Financial Support:

This work was supported in part by the National Library of Medicine (NLM) Medical Informatics Traineeship (LM 07092), the U.S. National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institutes of Health (NIH) under Grant Number R01 EB001659, Philips Healthcare and the Information and Communication University (ICU), Korea.

Keywords: Hypotension; Intensive Care; Physiologic Monitoring; Electronic Medical Record; Digital Signal Processing; Automatic Data Processing

ABSTRACT

Objective: To compare invasive blood pressure (BP) measurements recorded using an automated archiving method against clinician-documented values from the same invasive monitor, and determine which method of recording BP is more highly associated with the subsequent onset of hypotension.

Design: Retrospective comparative analysis.

Setting: Intensive care patients in a university hospital.

Patients: Mixed medical/surgical patients.

Interventions: N/A

Measurements: Using intervals of hemodynamic stability from 2,320 patient records, we retrospectively compared paired sources of invasive BP data: (1) measurements documented by the nursing (RN) staff; and (2) measurements generated by an automated archiving method that intelligently excludes unreliable (e.g., noisy or excessively damped) BP values. The primary outcome was the occurrence of subsequent “consensus” hypotension, i.e., hypotension documented jointly by the RN and the automated archive.

Main Results: The automated method could be adjusted to alter its operating characteristics (sensitivity and specificity). At a matched level of specificity (96%), BP from the automated archiving method was more sensitive (28%) for subsequent “consensus” hypotension versus the RN documented values (21%). Likewise, at a matched level of sensitivity (21%), the automated method was more specific (99%) versus the RN documented values (96%). These significant findings ($p < 0.001$) were consistent in a set of sensitivity analyses which employed alternative criteria for patient selection and the clinical outcome definition.

Conclusions: During periods of hemodynamic stability in an ICU patient population, clinician-documented BP values were inferior to an intelligent automated archiving method, as early indicators of hemodynamic instability. Human oversight may not be necessary for creating a

valid archive of vital signs data within an electronic medical record. Moreover, if clinicians do have a tendency to disregard early indications of instability, then an automated archive may be a preferable source of data for so-called Early Warning Systems that identify patients at-risk of decompensation.

INTRODUCTION

Continual physiologic monitoring, e.g., within the Intensive Care Unit (ICU), poses a work burden on caregivers who must regularly document the data. Although recent computing capabilities make it technologically feasible to automatically record voluminous, continuous physiologic data, it may not be desirable: automatically archived data may be overly polluted with the measurement errors and artifacts that are known to corrupt physiologic data [1-10].

Today's typical documentation workflow relies on a clinician to filter vital sign data, applying judgment and bedside observation to document only values and patterns that are deemed clinically meaningful, while excluding erroneous or unrepresentative data. A possible alternative to using a "clinician filter" would be an automated algorithm to filter out unreliable physiologic data. We have previously described signal quality indices (SQIs) designed to automatically evaluate the reliability of a continuous arterial blood pressure (ABP) waveform that is measured from an indwelling radial artery catheter [11-14]. These algorithms compute parameters related to the shape of the waveforms, and, based on if these parameters are within normative ranges and similar to prior beats' values, the algorithms output a rating of the reliability of the ABP waveform.

In this present study, we retrospectively compared the validity of arterial pressures documented in an ICU nursing record (MAP-RN and SBP-RN, for the mean arterial pressure and systolic pressure, respectively) versus arterial pressures from an investigational archiving agent (MAP-AUTO and SBP-AUTO) that employed the aforementioned SQIs to automatically exclude unreliable (e.g., noisy or excessively damped) blood pressure measurements. As a measure of validity, we examined the association of those two sources of archival blood pressure data for subsequent hypotension, assuming that there would be the strongest associations if data were more valid and weaker associations if the data contained invalid, erroneous measurements.

We previously compared the distributional characteristics between nurse-charted blood pressure values versus data employing the automated SQI algorithm, and found a bias towards higher values for the clinician's documentation [15]. This preceding analysis, however, did not address the clinical validity of the automated algorithm. If the present investigation can demonstrate that automated algorithms are more valid than clinicians' documentation, then the

automated methodology might enhance healthcare, creating a more useful record of data with less human effort.¹

MATERIALS AND METHODS

Study data were retrospectively extracted from the MIMIC II database [16]. The MIMIC II database includes physiologic and wide-ranging clinical data from over 30,000 ICU patient visits (medical ICU, critical care unit, and surgical ICU) hospitalized at the Beth Israel Deaconess Medical Center, Boston, USA between 2001 and 2005. Additional details about the MIMIC II database are available in [16]. The data were collected and analyzed with institutional approval by the local IRB.

Data inclusion/exclusion criteria

We analyzed all measurements of MAP-RN and MAP-AUTO for which two criteria were met:

- 1) MAP-RN and MAP-AUTO were both available at the same point in time. This enabled a paired comparison between the two sources of data.
- 2) The measurement pairs were preceded by ≥ 4 hours of consensus stable blood pressure (specifically, ≥ 4 hours in which all values of MAP-RN and MAP-AUTO were ≥ 70 mmHg).²

Investigational measurements

1. MAP-RN

The process for comparing MAP-RN versus MAP-AUTO is illustrated in Fig. 1 and Fig.2. First, we identified documented MAP-RN data. Blood pressures of patients with indwelling

¹Of course, human documentation also promotes awareness; this important matter is addressed in the Discussion.

² For determination of consensus stability, which is one inclusion criterion, MAP-AUTO was computed from blood pressure waveform data with a SQI ≥ 70 %; see MAP-AUTO section for details. As a practical matter, this meant that the antecedent hemodynamic state of each subject, upon comparison of MAP-RN versus MAP-AUTO, was unambiguous and consistent throughout our analyses. By contrast, if the patients' initial states had been ambiguous, i.e., there had been a discrepancy between MAP-RN and MAP-AUTO, it would have been challenging to conduct a fair comparison.

arterial catheters were documented by the nursing staff as part of routine clinical operations, at a sampling period of between 1 and 120 minutes (median of precisely 60 minutes; interquartile range of 30 minutes) using Philips CareVue (Philips Healthcare, Andover, MA) for electronic nursing documentation. The CareVue system pre-populates data fields with current blood pressure data transmitted directly from the ABP measurement device, and the clinician either accepts or edits these values, as per the discretion of the caregiver. These radial ABP waveform data, measured by the M1006B invasive pressure module, were sourced from Philips CMS bedside patient monitors (Philips Healthcare, Andover, MA). ABP waveforms were sampled at 125 Hz with 8 bit resolution. The ICU protocol called for rezeroing and the flush test at least once per shift, although the rates of protocol compliance during actual clinical operations are undocumented.

2. MAP-AUTO

For every MAP-RN value documented for a patient, a temporally-matched MAP-AUTO was sought. MAP-AUTO was computed from the continuous ABP signal sourced from the same Philips CMS bedside patient monitor. A MAP-AUTO value was therefore available for nearly any time, except for relatively uncommon episodes *without at least* 10 seconds of continuous, reliable ABP waveform data within the preceding 6 minutes, in which case the unmatched MAP-RN value was excluded from further analysis.

Computation of MAP-AUTO was comprised of two processing steps. First, unreliable ABP data with low SQI ($< 70\%$) were identified and excluded. Second, a representative value was extracted from the remaining *reliable* ABP waveform data. For the first step, unreliable ABP data were identified using an SQI algorithm that has been previously described in detail [**Error! Reference source not found., Error! Reference source not found.**], which combines functionality of two antecedent SQI algorithms [11,12]. Any segments of ABP waveform data with an SQI rating less than a threshold, i.e., $SQI \leq \text{threshold}$, were excluded from a given analysis. All analyses were repeated using the full spectrum of integer SQI cut-off thresholds, from 0 to 100.

The SQI algorithm computes parameters related to the shape of the waveforms, and, based on if these parameters are within normative ranges and are similar to prior beats' values, outputs a

rating of the reliability of the ABP waveform. SQI is expressed as an integer between 0 (poorest quality data) and 100 (highest quality data) for consecutive non-overlapping 10 seconds windows of ABP. The waveform features that are considered by the SQI algorithm include, for each pulse, the systolic pressure, diastolic pressure, mean pressure, pulse pressure (the difference between the SBP and the DBP in a beat), pulse-to-pulse interval (T), the maximum and minimum slope during the up-stroke, the duration of the up-stroke, the duration of the crest of the beat, and the average of all negative slopes (a metric of spiky, nonphysiologic noise in the waveform).

After excluding unreliable ABP data (i.e., $SQI < \text{threshold}$), for a given time t in a patient's chronologic record, MAP-AUTO and SBP-AUTO were computed using the median value from the remaining *reliable* ABP waveform in the most recent 6 minutes (A series of permutations on this primary methodology were computed; see the *Sensitivity Analysis* section below for details). As noted above, if at time t , there were less than 10 seconds of continuous reliable ABP waveform data within the past 6 minutes, we registered MAP-AUTO and SBP-AUTO as unavailable.

In this investigation, the automated archive was created retrospectively by processing the ABP waveform data using the investigational algorithm, but there is no technical reason why this could not function in real-time (i.e., the processing is unsupervised by humans and it is not computationally intensive). Therefore, for this study, we refer to the set of MAP-AUTO and SBP-AUTO data as an archive.

Study outcomes

Our study outcome was “consensus hypotension” within the subsequent 4 hours. Consensus hypotension was defined by $MAP \leq 70$ mmHg, documented at the same time by both MAP-RN and MAP-AUTO. By using a definition that required the simultaneous agreement of both MAP-RN and MAP-AUTO, we limited bias in our results. Each MAP measurement was evaluated as follows (also see examples in Fig. 1 and Fig. 2):

- True Positive: $MAP \leq 70$ mmHg; within the next 4 hours, is followed by consensus hypotension

- False Positive: $MAP \leq 70$ mmHg; within the next 4 hours, *is not* followed by consensus hypotension
- True Negative: $MAP > 70$ mmHg; within the next 4 hours, *is not* followed by consensus hypotension
- False Negative: $MAP > 70$ mmHg; within the next 4 hours, *is* followed by consensus hypotension

The fact that the nurse and the algorithm could disagree about blood pressure at a given point in time may be counter-intuitive. Briefly, discrepancies occurred because the nurse and algorithm used different "filters" to exclude unreliable data and then to summarize blood pressure over an observation interval, so in fact, there were many instances when there was disagreement about what blood pressure value to document. Using a fixed threshold for hypotension, i.e., $MAP \leq 70$ mmHg, there are four possible observation combinations from the two documentation sources. In most cases the MAP-RN and MAP-AUTO are both documented as stable. The most likely interpretation of this combination is that the patient is in fact hemodynamically stable, but there is also the possibility that both documentation sources are errant. Table 1 details the possible clinical interpretations for each of the four paired measurement scenarios for MAP-AUTO and MAP-RN.

[Table 1 about here]

The key assumption in this study design is that, *on average*, a more valid MAP measurement will have a higher association with future hemodynamic states. The association of MAP-AUTO and MAP-RN with future consensus hypotension was statistically compared using McNemar's test on contingency tables from matched pairs. We compared MAP-RN versus MAP-AUTO first by using whatever SQI threshold gave equal sensitivities (but potentially unequal specificities), and then by using whatever SQI threshold gave equal specificities (but potentially unequal sensitivities).

[Figure 1 about here]

Sensitivity Analysis

We re-analyzed our data using several permutations of the primary methodology, investigating a different definition of hypotension, different methods of algorithmically processing the data, and different definitions of hemodynamic instability.

We explored the following alternatives to our primary methodology (summarized in Table 2):

- Examined *SBP* instead of *MAP*. For *SBP* we changed our hypotension definition to be < 90 mmHg
- Computed *MAP-AUTO* as the *minimum* value from all reliable *ABP* waveform data from the most recent six minutes (instead of using the *median* value)
- Computed *MAP-AUTO* as the median value from all reliable *ABP* waveform data from the most recent *60 minutes* (instead of six minutes)
- Altered the outcome to consensus hypotension *or* an increase of at least 100% in vasopressor infusion rate (Levophed, Vasopressin, Neosynephrine, Dopamine, or Epinephrine)
- Altered the inclusion criteria: when determining whether or not there were 4 hours of antecedent “consensus stability”, *MAP-AUTO* was computed using a lower ($SQI \geq 0$) and a higher ($SQI \geq 90$) threshold

[Table 2 about here]

[Figure 2 about here]

RESULTS

Primary Findings

Working from the beginning of each record in the 2,320 adult ICU visits with archived *ABP* waveform data, we found a total of 35,659 valid 8-hour episodes (episodes that begin with 4 hours of consensus stability) from 757 unique ICU visits. Subject characteristics are summarized in Table 3.

[Table 3 about here]

The receiver operating characteristic (ROC) curve for the association between any archived MAP value and subsequent hypotension is shown in Fig.3. For all blood pressure data intervals that were analyzed, RN's documented a single blood pressure value, and their summary sensitivity and (1-specificity) are plotted. By contrast, the MAP-AUTO was adjustable depending on how the data reliability criteria were set: the SQI (from an automated algorithm) provided a rating of the reliability of the ABP waveform, from 0-100. As the SQI cut-off approached 100, the reliability criteria grew more stringent, i.e., only the most pristine blood pressure measurements were archived. As a result, the specificity increased, the sensitivity decreased, and the positive predictive value for subsequent hypotension approached 80%. Conversely, as the SQI cut-off approached 0, the reliability criteria grew more relaxed until eventually all blood pressure measurements were used. As a result, the specificity decreased, the sensitivity increased, and the positive predictive value for subsequent hypotension eventually degraded to below 5%.

In Table 4, data from four points of the ROC curve in Fig.3 are tabulated as contingency tables. Specifically, we report data from the MAP-RN values, as well as four illustrative points from the MAP-AUTO curve, corresponding to $SQI = 100$, $SQI \geq 92$, $SQI \geq 18$, and $SQI \geq 0$, which are identified in Fig.3.

The MAP-AUTO and MAP-RN were statistically compared at two points and the p -values from McNemar's test are presented in Table 5 for two contingency tables: one resulting from the different sensitivities when the specificities are matched and one resulting from the different specificities when the sensitivities are matched (the points on Fig.3 where $SQI \geq 18$ and $SQI \geq 92$, respectively). The improvements in sensitivity and specificity obtained by using the MAP-AUTO versus the MAP-RN were statistically significant ($p < 0.0001$).

[Figure 3 about here]

[Table 4 about here]

[Table 5 about here]

Sensitivity Analysis

For our sensitivity analysis, we evaluated a different definition of hypotension, different methods of algorithmically processing the data, and different definitions of hemodynamic stability. Table 6 lists the permutations that we explored (the top row recapitulates the findings from the primary analysis). We found our results insensitive to these permutations, and in all cases the MAP-AUTO/SBP-AUTO was significantly more sensitive (at a matched level of specificity) and more specific (at a matched level of sensitivity) than the MAP-RN/SBP-RN. When the window length was increased from 6 minutes to 30 minutes, the difference between the two signals decreased, resulting in the highest p-value of 0.00042. All other p-values were consistently < 0.0001 .

[Table 6 about here]

For illustrative purposes, we show one ROC curve from the sensitivity analysis. In this permutation, we compared the association between archived mean arterial pressure data (MAP-RN or MAP-AUTO) versus hemodynamic instability, defined by *either* consensus hypotension *or* the doubling of the infusion rate of any vasopressor drugs. Fig. 4 shows the corresponding ROC curve and PPV for varying SQI thresholds. Comparing these curves to Fig. 3, the MAP-AUTO performance degrades slightly, while the MAP-RN performance point is similar. All the same, as in Table 5, the MAP-AUTO values are statistically superior to their matched MAP-RN counterparts.

[Figure 4 about here]

Discussion

Significance

In this analysis, RN documentation of blood pressure data in stable ICU patients does not improve the clinical validity of the ICU medical record, as compared with an automated archiving methodology. We found a small but highly significant advantage to the automated methodology, a finding that persisted throughout a set of sensitivity analyses, suggesting that this is not idiosyncratic to one method of analysis, but is probably generalizable to a spectrum of different definitions of clinical validity and different methods of automated archiving. This has

notable implications for present-day hospital operations, as well as for technologic capabilities that might develop in the future.

In today's hospitals, substantial time and effort are spent in clinical documentation [17]. If complete human attentiveness to clinical parameters was possible, it would presumably be impossible to beat the clinical team in terms of selecting representative data to aid a clinical evaluation. The findings here suggest that, for MAP and SBP at least, such clinician vital signs documentation offers no archival value, perhaps because it is impossible to maintain perfect focus given diverse work duties and the repetitiveness of some tasks. It is possible then that some of this time and effort of documentation are not strictly necessary, compared to an automated alternative (N.B., documentation may have other benefits, such as creating awareness in clinicians, which is addressed in the Limitations section). Moreover, it is standard practice for clinicians on rounds to review documented vital signs – to assess the course of a disease process, the efficacy of a therapy, the development of a new pathology, etc. – and our findings suggest that there may be a more valid alternative to reviewing an RN-documented archive of blood pressure data, and perhaps automatic archiving agents may prove valid for other vital signs, such as respiratory rate [18], urine output, etc., though this is a matter of speculation. Using MAP-AUTO (or SBP-AUTO) offers one additional advantage versus MAP-RN (SBP-RN). Specifically, the SQI can be adjusted to alter operating characteristics (sensitivity and specificity) to best suit clinical needs. MAP-AUTO has a PPV that is similar to the MAP-RN when the SQI threshold is set extremely low (e.g., $SQI > 5$). At the same time, we found that the PPV of a highly reliable (e.g., $SQI > 90$) MAP-AUTO measurement approaches 80%. Operationally it would be valuable to communicate to the caregivers this extra information, that in certain circumstances there is a *possibility* of future hypotension (e.g., moderate SQI hypotension), whereas in other circumstances there is a significant *probability* of experiencing future hypotension (e.g., high SQI hypotension). The motivation, of course, would be to communicate information so that caregivers can respond appropriately, either mitigating or preventing the subsequent hypotension. The MAP-RN, by contrast, is a fixed value, with a PPV of less than 50%. There is no easy way to modify the sensitivity or specificity of MAP-RN, or to extract information beyond what was documented.

Our findings also inform technologic capabilities that may be developed in the future. First of all, there is substantial interest in early warning scores (EWS), in which continual vital signs and other data are monitored and, when abnormal conditions are detected, a clinician response team is mobilized to respond to the incipient deterioration of a patient [19, 20]. It is possible that such EWS functionality could be automated, and our findings offer preliminary evidence that human oversight, e.g., ensuring that spurious blood pressure data are excluded, may not be necessary, or even desirable. Automatically archived blood pressure data, using an adjustable SQI, may be the best source of input data for such decision-support algorithms in the ICU. It is possible that these results also pertain to non-ICU hospital wards, where the benefits of rapid response teams (RRTs, which are typically activated on the basis of abnormal vital signs data) have been quite inconsistent in published reports [20]. We found that human documented blood pressure data are inferior, and it is likely that some RNs are better than others in terms of charting clinically valid blood pressure data. We speculate that one reason RRT programs have had varied success is because of inconsistent vital signs collection practices by different nursing staffs. An automated archive may offer a more valid, continual, and consistent method of data collection for EWS applications.

Finally, these findings suggest another interesting hypothesis, to be developed and tested in future work: the “secretarial” aspects of documentation, i.e., recording a tedious list of parameters and findings for future review, may be distracting from the real-time benefits of documentation, namely, obliging caregivers to re-examine their patients on a regular basis. It would seem ideal if, in the future, clinical processes emphasized the continual re-examination of patients by the clinical staff (rather than the secretarial tasks) employing novel Clinical Information Systems to reduce the effort of data archiving, and to automatically highlight interesting patterns/changes in the clinical parameters ensuring that such patterns were not accidentally overlooked. In terms of data display, our findings suggest it might be reasonable for clinicians on rounds to review automatically archived BP records, with the associated BP reliability measures indicated, so the clinicians can assess themselves which data are most meaningful. This current report is important because it suggests that there is real room for improvement of today’s ICU documentation, with future work justified in optimizing computer-clinician interactions to yield the best patient care.

Overall, our results provide very preliminary evidence that clinician oversight is not strictly necessary for valid collection of physiologic data. Therefore, extremely large records of physiologic data may not be limited by the requirement for clinician oversight, i.e., archiving continuous physiologic data for days at a time during an inpatient admission (though whether such a practice would offer any clinical benefit, and justify the substantial data storage requirement, remain completely open questions).

Limitations

There are several important limitations to consider. Our findings relate only to blood pressure data for initially stable ICU patients within a single institution. The findings may not apply to other vital signs, to consistently unstable ICU patients, to non-ICU patients, or to ICU patients in other institutions. However, there is gathering evidence that automated methods for excluding unreliable vital signs measurements and maximizing the clinical validity of the data, may prove as good as or better than clinicians [18, 21, 22].

Our findings do not apply to patients with ongoing hemodynamic instability; we only examined records with periods of antecedent blood pressure stability. For actively unstable patients, there may be reasons why clinician documented data would be more valid, e.g., because the RN attention is more focused and reliable, or because the pathophysiologic condition is too complex for a simple computer algorithm, etc. However, the consequence of this reasoning is that our results may be even more applicable outside the ICU where there is a lower staff to patient ratio and hence a lower probability of identifying infrequent events such as hypotension.

It could be possible that our methodology is unfair to the nurse: after documenting hypotension, the nurse might therapeutically intervene, and so avoid future hypotension. Then, even though the documentation of hypotension was valid, our methodology would treat this scenario as a false-positive for MAP-RN because there was no future hypotension. As a result, we may be underestimating test characteristics, such as PPV, for MAP-RN. However, these occurrences are unlikely to alter our major findings:

- Overall, there were substantial differences between MAP-RN and MAP-AUTO (e.g., PPV as plotted in Fig. 3).

- In the sensitivity analysis, when we modified the outcome definition to include therapeutic interventions to hypotension – consensus hypotension *or* increase in vasopressor infusion rate – there was only marginal improvement in the MAP-RN PPV (see Fig. 4), while MAP-AUTO remained significantly superior in terms of sensitivity and specificity, without any change in the p-value < 0.001 (see Table 6).

Furthermore, note that such occurrences would never reduce the PPV of MAP-AUTO, nor would they alter the finding that adjusting the SQI cut-off yields a wide range of PPVs for MAP-AUTO.

As a final study limitation, we note that vital signs documentation is not merely for archival purposes. The process of reviewing and documenting vital signs may alert the clinician to a troublesome condition, or to malfunctioning equipment. While it may be that clinician's do not need to formally document blood pressure data, it is likely that an alternative, perhaps more time-efficient, mechanism, such as an interactive graphical user interface, would be necessary to ensure that the clinician is well aware of the current blood pressure of the patient. It is possible that some day, fully automated care of ICU patients may be possible (perhaps using some of the automated techniques employed in this study), but our findings are limited to only the *archival* value of MAP-AUTO versus MAP-RN data. Real clinicians do not have the luxury of looking back many minutes in the past for the last reliable blood pressure measurement when making real-time management decisions in critically-ill patients.

CONCLUSIONS

In an initially stable ICU patient population, clinician-documented blood pressure values were inferior to an automated archiving agent with signal quality filtering, as early indicators of hemodynamic instability. These findings suggest that human oversight may not be necessary for creating a valid archive of vital signs data within an electronic medical record. Moreover, if clinician documentation is an unreliable early indicator of hemodynamic instability, then an automated archive may be a preferable source of data for early warning systems that identify patients at-risk of decompensation.

Acknowledgements

This work was supported in part by the National Library of Medicine (NLM) Medical Informatics Traineeship (LM 07092), the U.S. National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the National Institutes of Health (NIH) under Grant Number R01 EB001659, Philips Healthcare and the Information and Communication University (ICU), Korea. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NLM, the NIBIB, the NIH, Philips Healthcare, or ICU Korea.

	MAP-RN: Documentation of Stable Blood Pressure	MAP-RN: Documentation of Hypotension
MAP-AUTO: Documentation of Stable Blood Pressure	<p>Valid stable blood pressure [AUTO valid; RN valid]</p> <p><i>or</i></p> <p>Valid (transient) hypotension incorrectly rejected by AUTO because waveform appears noisy/unreliable AND RN deems valid hypotension insignificant or fails to notice transient episode [AUTO invalid; RN invalid]</p>	<p>Valid (transient) hypotension incorrectly rejected by AUTO because waveform appears noisy/unreliable AND Hypotension properly identified by RN [AUTO invalid; RN valid]</p> <p><i>or</i></p> <p>Erroneous or insignificant hypotension properly rejected by AUTO because waveform appears noisy/unreliable AND Erroneous or insignificant hypotension inappropriately accepted by RN [AUTO valid; RN invalid]</p>
MAP-AUTO Documentation of Hypotension	<p>Hypotension properly identified by AUTO because waveform appears clean AND RN deems valid hypotension insignificant or fails to notice transient episode [AUTO valid; RN invalid]</p> <p><i>or</i></p> <p>Erroneous or insignificant hypotension inappropriately accepted by AUTO because waveform appears clean AND Erroneous or insignificant hypotension properly rejected by RN [AUTO invalid; RN valid]</p>	<p>Valid hypotension [AUTO valid; RN valid]</p> <p><i>or</i></p> <p>Erroneous or insignificant hypotension inappropriately accepted by AUTO because waveform appears clean AND Erroneous or insignificant hypotension inappropriately accepted by RN [AUTO invalid; RN invalid]</p>

Table 1: Possible clinical interpretations for paired measurements of MAP-AUTO and MAP-RN. Distinguishing between valid and invalid measurements on a case-by-case basis is problematic. Our key assumption was that, on average, the more valid measurements would have significantly higher associations with future hemodynamic states. See Methods section for details.

AUTO Filter Settings

Parameter	Baseline	Alternative(s)
Pressure Type	MAP	SBP
Averaging Filter	Median	Min
Filter Window	6 Minutes	60 Minutes
Outcome	Consensus Hypotension	Consensus Hypotension <i>or</i> Pressor Increase
Consensus SQI thresholds	70	{0, 90}

Table 2: Parameter permutations used for sensitivity analyses.

Subject Characteristics

Characteristic	Value (\pm stddev)
ICU Visits	2,320
Total Data Segments	35,659
Age	65 \pm 16 years
Admit Weight	82 \pm 26 kg
Sex	56.9% M, 43.1% F
Hospital Mortality	24.3%
Incidence of Pressor Infusions	44.9%
Incidence of Mech. Ventilation	58.9%
CSRU Service	22.5%
MICU Service	32.0%
CCU Service	29.4%
NSICU Service	11.9%
MSICU Service	3.4%
CSICU Service	7.3%

Table 3: Characteristics for study subjects

	Documented MAP	Episodes of Future Hypotension (≤ 4 hrs)	Episodes of Future Stability (≤ 4 hrs)	Predictive Values
MAP-RN	Hypotension	802	1263	0.39 PPV
	Stable blood pressure	3040	30554	0.91 NPV
MAP-AUTO (SQI = 100)	Hypotension	489	171	0.74 PPV
	Stable blood pressure	3353	31646	0.90 NPV
MAP-AUTO (SQI ≥ 92)	Hypotension	790	283	0.74 PPV
	Stable blood pressure	3052	31534	0.91 NPV
MAP-AUTO (SQI ≥ 18)	Hypotension	1059	1245	0.46 PPV
	Stable blood pressure	2783	30572	0.92 NPV
MAP-AUTO (SQI ≥ 0)	Hypotension	1106	9298	0.11 PPV
	Stable blood pressure	2736	22519	0.89 NPV

Table 4: Contingency tables for documented MAP data from stable patients versus future hemodynamics. We report summary data for the RN documented data, as well as from three illustrative points from the MAP-AUTO curve that are labeled in Figure3. PPV=Positive Predictive Value; NPV=Negative Predictive Value.

Sensitivity Comparison (Specificity=0.960)				Specificity Comparison (Sensitivity=0.209)			
MAP-AUTO (w/ SQI \geq 18)				MAP-AUTO (w/ SQI \geq 92)			
	TP	FN		TN	FP		
MAP-RN	TP	708	94	MAP-RN	TN	30271	283
	FN	351	2689		FP	1263	0
$\chi^2 = 147, p < 0.0001$ MAP-AUTO Sensitivity: 0.276 MAP-RN Sensitivity: 0.209				$\chi^2 = 620, p < 0.0001$ MAP-AUTO Specificity: 0.991 MAP-RN Specificity: 0.960			

Table 5: The MAP-AUTO and MAP-RN were statistically compared with McNemar’s test at two points, one resulting from the different sensitivities when the specificities are matched (i.e., MAP-AUTO using SQI \geq 18) and one resulting from the different specificities when the sensitivities are matched (i.e., MAP-AUTO using SQI \geq 92); TN = True Negative; FN = False Negative.

Signal	win [min]	Filter Type	Consensus SQI	Hypotension or pressors	Matched Specificity			Matched Sensitivity		
					AUTO Sensitivity	RN Sensitivity	<i>p</i>	AUTO (1-Spec)	RN (1-Spec)	<i>p</i>
MAP	6	Median	≥ 70	No	0.276	0.209	<0.001 [†]	0.009	0.040	<0.001 [†]
SBP	6	Median	≥ 70	No	0.220	0.180	<0.001 [†]	0.005	0.018	<0.001 [†]
MAP	30	Median	≥ 70	No	0.250	0.229	<0.001 [†]	0.010	0.042	<0.001 [†]
MAP	6	Min	≥ 70	No	0.319	0.168	<0.001 [†]	0.012	0.038	<0.001 [†]
MAP	6	Median	≥ 0	No	0.264	0.201	<0.001 [†]	0.013	0.020	<0.001 [†]
MAP	6	Median	≥ 90	No	0.277	0.219	<0.001 [†]	0.008	0.045	<0.001 [†]
MAP	6	Median	≥ 70	Yes	0.222	0.184	<0.001 [†]	0.012	0.037	<0.001 [†]

Table 6: Sensitivity analyses obtained by altering the values for one parameter (in bold) of the baseline (first row) at a time. The *p*-values marked with † indicate significance at the 0.05 level.

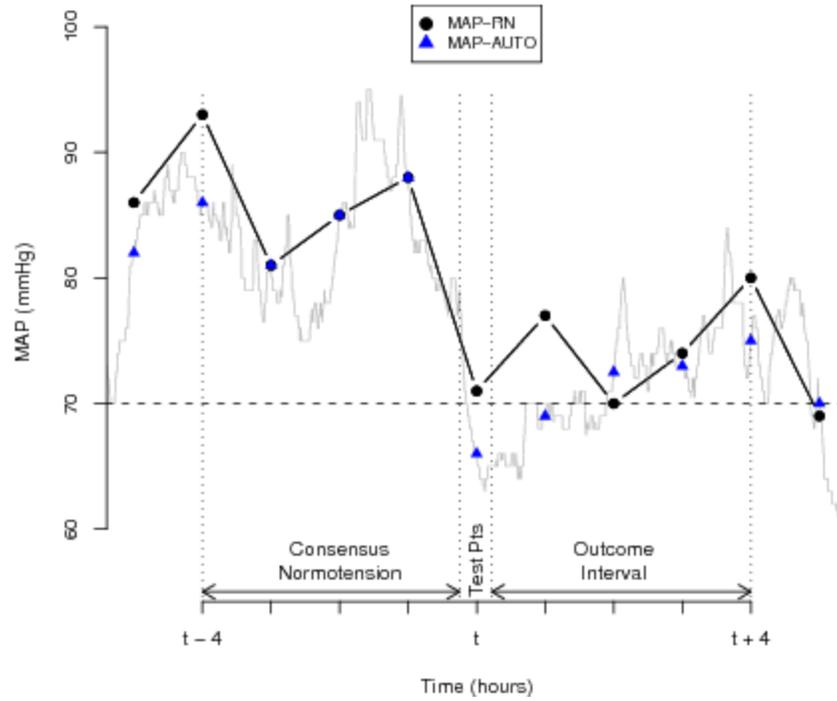


Figure 1: Example of a false positive (for the MAP-AUTO) and a true negative (for the MAP-RN). At time t , the MAP-AUTO value drops below the hypotensive threshold of 70 mmHg, but the two signals fail to reach consensus hypotension in the subsequent 4 hours (the consensus hypotension at $t+5$ is outside of our 4 hour window). Note that the antecedent 4 hours of consensus stability prior to time t , in which both MAP-RN and MAP-AUTO are >70 mmHg, satisfies the inclusion criteria.

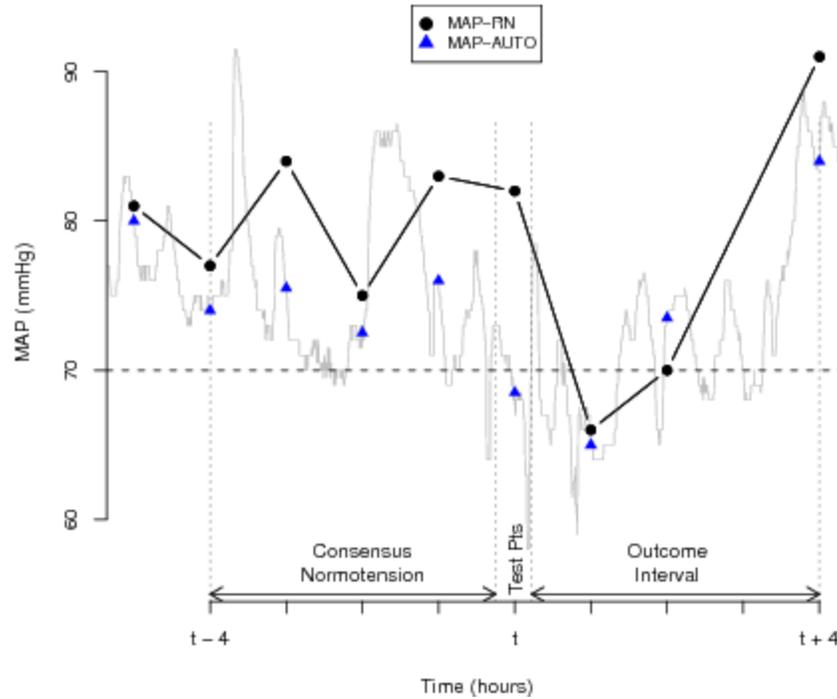


Figure 2: Example of a true positive (for the MAP-AUTO) and a false negative (for the MAP-RN). At time t , the MAP-AUTO value drops below the hypotensive threshold of 70 mmHg, and the two signals reach consensus hypotension in the subsequent 4 hours. Note that the antecedent 4 hours of consensus stability prior to time t , in which both MAP-RN and MAP-AUTO are >70 mmHg, satisfies the inclusion criterion.

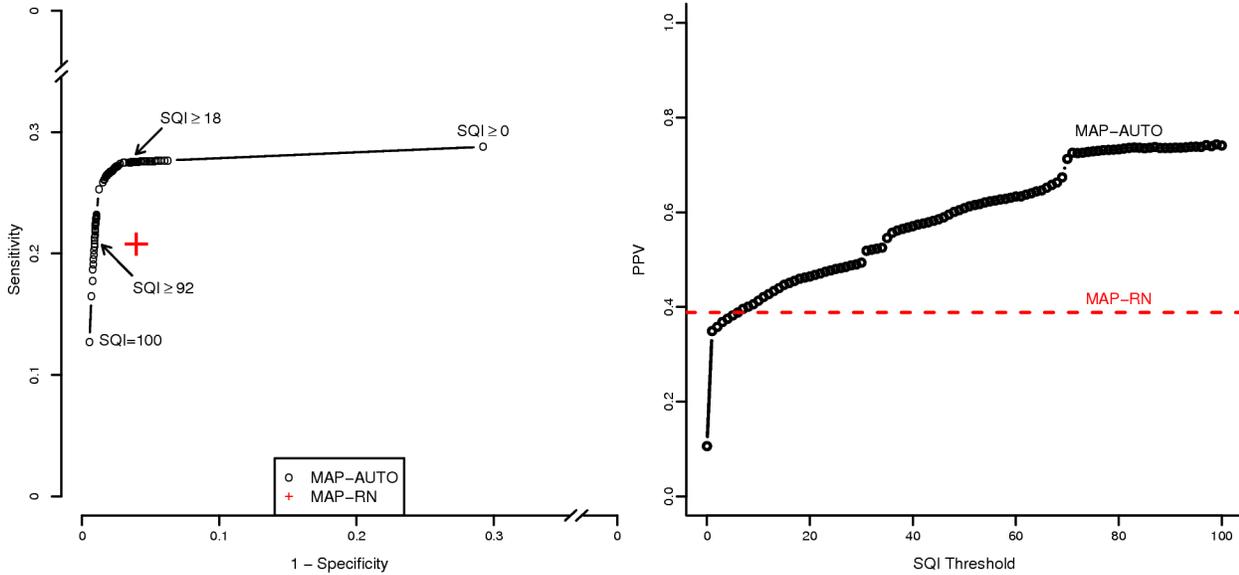


Figure 3: Receiver operating characteristic curve (left) and corresponding positive predictive value (PPV) curve (right) for the association between any archived blood pressure value and subsequent hypotension (for clarity there is a break in the y-axis of the ROC curve). Shown: MAP-AUTO (black circles/solid line) through a range of signal quality indices 0-100 (SQIs) versus MAP-RN (red cross/dash). With $SQI \geq 18$, MAP-AUTO has an equal specificity as MAP-RN. With $SQI \geq 92$, MAP-AUTO has an equal sensitivity as MAP-RN.

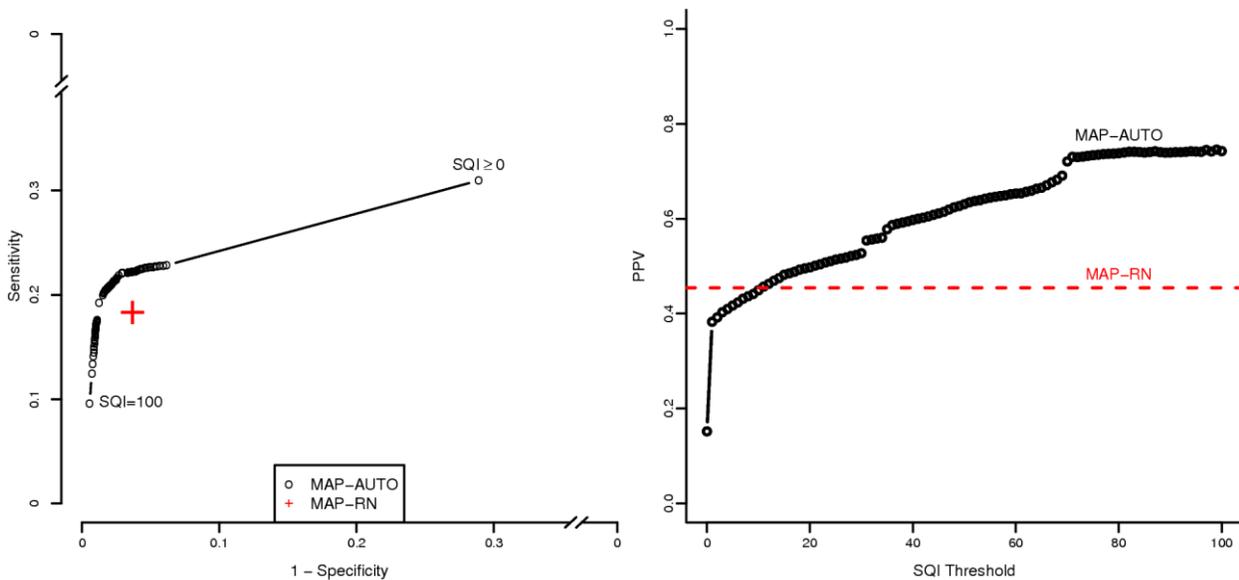


Figure 4: An example from the sensitivity analyses: Receiver operating characteristic curve (left) and corresponding positive predictive value curve (right) for the association between any archived MAP value and one alternative outcome definition, which was the development of hypotension or the increase in infusion rate of vasopressors.

REFERENCES

1. Kaiser W, Findeis M: **Artifact processing during exercise testing.** *J Electrocardiol* 1999, **32 Suppl**:212-219.
2. Tsien CL, Fackler JC: **Poor prognosis for existing monitors in the intensive care unit.** *Crit Care Med* 1997, **25**(4):614-619.
3. Edmonds ZV, Mower WR, Lovato LM, Lomeli R: **The reliability of vital sign measurements.** *Ann Emerg Med* 2002, **39**(3):233-237.
4. Lovett PB, Buchwald JM, Sturmman K, Bijur P: **The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage.** *Ann Emerg Med* 2005, **45**(1):68-76.
5. Jones DW, Appel LJ, Sheps SG, Roccella EJ, Lenfant C: **Measuring blood pressure accurately: new and persistent challenges.** *JAMA* 2003, **289**(8):1027-1030.
6. Friesdorf W, Konichezky S, Gross-Alltag F, Fattroth A, Schwilk B: **Data quality of bedside monitoring in an intensive care unit.** *Int J Clin Monit Comput* 1994, **11**(2):123-128.
7. Kacmarek RM: **Alarms.** In: *Principles and Practice of Intensive Care Monitoring.* Edited by Tobin MJ. New York: McGraw-Hill, Inc.; 1998: 133-140.
8. Goldman JM, Schrenker RA, Jackson JL, Whitehead SF: **Plug-and-play in the operating room of the future.** *Biomed Instrum Technol* 2005, **39**(3):194-199.
9. Amoores JN: **A simulation study of the consistency of oscillometric blood pressure measurements with and without artefacts.** *Blood Press Monit* 2000, **5**(2):69-79.
10. Portet F, Hernandez AI, Carrault G: **Evaluation of real-time QRS detection algorithms in variable contexts.** *Med Biol Eng Comput* 2005, **43**(3):379-385.
11. Li Q, Mark RG, Clifford GD: **Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter.** *Physiol Meas* 2008, **29**(1):15-32.
12. Li Q, Mark RG, Clifford GD: **Artificial arterial blood pressure artifact models and an evaluation of a robust blood pressure and heart rate estimator.** *Biomed Eng Online* 2009, **8**:13.
13. Sun J, Reisner AT, Mark RG: **A signal abnormality index for arterial blood pressure waveforms.** *Computers in Cardiology* 2006, **33**:13-16.
14. Zong W, Moody GB, Mark RG: **Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure.** *Med Biol Eng Comput* 2004, **42**(5):698-706.
15. Hug C, Clifford GD: **An analysis of the errors in recorded heart rate and blood pressure in the ICU using a complex set of signal quality metrics.** *Computers in Cardiology* 2007, **34**:641-644.
16. Clifford GD, Scott DJ, Villarreal M: **User guide and documentation for the MIMIC-II database. MIMIC-II database version 2, release 1.** In. Cambridge: Massachusetts Institute of Technology; 2009.
17. Poissant L, Pereira J, Tamblyn R, Kawasumi Y: **The impact of electronic health records on time efficiency of physicians and nurses: a systematic review.** *J Am Med Inform Assoc* 2005, **12**(5):505-516.
18. Chen L, Reisner AT, Gribok A, McKenna TM, Reifman J: **Can we improve the clinical utility of respiratory rate as a monitored vital sign?** *Shock* 2009, **31**(6):574-580.

19. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M: **Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards.** *Cochrane Database Syst Rev* 2007(3):CD005529.
20. Winters BD, Pham JC, Hunt EA, Guallar E, Berenholtz S, Pronovost PJ: **Rapid response systems: a systematic review.** *Crit Care Med* 2007, **35**(5):1238-1243.
21. Aboukhalil A, Nielsen L, Saeed M, Mark R, Clifford G: **Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform.** *Journal of biomedical informatics* 2008.
22. Reisner AT, Chen L, McKenna TM, Reifman J: **Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation.** *J Trauma* 2008, **65**(4):915-923.