



6-2016

Clinician-in-the-Loop Annotation of ICU Bedside Alarm Data

Alexander Roederer

University of Pennsylvania, roederer@seas.upenn.edu

Joseph Dimartino

University of Pennsylvania Health System, joseph.dimartino@uphs.upenn.edu

Jacob Gutsche

University of Pennsylvania Health System, jacob.gutsche@uphs.upenn.edu

Margaret Mullen-Fortino

Thomas Jefferson University Hospital

Sachin Shah

University of Pennsylvania Health System

See next page for additional authors

Follow this and additional works at: http://repository.upenn.edu/cis_papers



Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

Recommended Citation

Alexander Roederer, Joseph Dimartino, Jacob Gutsche, Margaret Mullen-Fortino, Sachin Shah, C. William Hanson III, and Insup Lee, "Clinician-in-the-Loop Annotation of ICU Bedside Alarm Data", *First IEEE Conference on Connected Health Applications, Systems and Engineering Technologies (IEEE CHASE 2016)*, 229-237. June 2016. <http://dx.doi.org/10.1109/CHASE.2016.10>

The First IEEE Conference on Connected Health Applications, Systems and Engineering Technologies ([IEEE CHASE 2016](#)), Washington, DC, June 27-29, 2016

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/cis_papers/820

For more information, please contact repository@pobox.upenn.edu.

Clinician-in-the-Loop Annotation of ICU Bedside Alarm Data

Abstract

In this work, we describe the state of clinical monitoring in the intensive care unit and operating room, where patients are at their most fragile and thus monitoring is most heightened. We describe how large amounts of data generated by monitoring patients' physiologic signals, along with the ubiquitous aspecific threshold alarms in use today, cause dangerous alarm fatigue for medical caregivers. In order to build more specific, more useful alarms, we gathered a novel data set that would allow us to assess the number, types, and utility of alarms currently in use in the intensive care unit. To do this, we developed a system to collect physiologic monitor data, alarms, and annotations of those alarms provided electronically by clinicians. We describe the collection process for this novel data set and provide a preliminary description of the data.

Keywords

biomedical monitoring, monitoring, sensors, hospitals, fatigue, alarm systems, safety

Disciplines

Computer Engineering | Computer Sciences

Comments

The First IEEE Conference on Connected Health Applications, Systems and Engineering Technologies ([IEEE CHASE 2016](#)), Washington, DC, June 27-29, 2016

Author(s)

Alexander Roederer, Joseph Dimartino, Jacob Gutsche, Margaret Mullen-Fortino, Sachin Shah, C. William Hanson III, and Insup Lee

Clinician-in-the-Loop Annotation of ICU Bedside Alarm Data

Alexander Roederer*, Joseph Dimartino[†], Jacob Gutsche^{†‡}, Margaret Mullen-Fortino,[§] Sachin Shah[¶],
C. William Hanson[‡] and Insup Lee*

*Department of Computer and Information Science
University of Pennsylvania, Philadelphia, Pennsylvania, USA
Email: roederer@seas.upenn.edu

[†]Penn Presbyterian Medical Center, University of Pennsylvania Health System

[‡]Perelman School of Medicine, University of Pennsylvania Health System

[§]Thomas Jefferson University Hospital

[¶]Penn Medicine Information Services, University of Pennsylvania Health System

Abstract—In this work, we describe the state of clinical monitoring in the intensive care unit and operating room, where patients are at their most fragile and thus monitoring is most heightened. We describe how large amounts of data generated by monitoring patients’ physiologic signals, along with the ubiquitous aspecific threshold alarms in use today, cause dangerous alarm fatigue for medical caregivers. In order to build more specific, more useful alarms, we gathered a novel data set that would allow us to assess the number, types, and utility of alarms currently in use in the intensive care unit. To do this, we developed a system to collect physiologic monitor data, alarms, and annotations of those alarms provided electronically by clinicians. We describe the collection process for this novel data set and provide a preliminary description of the data.

I. INTRODUCTION

Hospitals have always used a wide array of medical devices to monitor patients. Many of these devices detect the intensity of various physical and chemical signals (broadly described as *physiologic signals*) in the body. Clinicians rely on the information from these signals to understand their patients’ current state of well-being, and how it changes over time.

The proliferation of inexpensive digital medical devices has allowed intensive care units (ICUs), operating rooms, emergency departments, neonatal units, and increasingly general care areas of the hospital to become “high technology” environments. In these areas of the hospital, many physiologic signals are now continuously monitored as part of routine clinical care. Clinicians (doctors, nurses, and other clinical caretakers) rely on the information from these signals to understand both a patients’ current state of well-being, and how it changes over time. Continuous digital monitoring is intended to allow clinicians to track changes in patient state more closely than would be possible with more sporadic analog measurements. The hope has been that this would allow for more accurate diagnosis, earlier anticipation of deterioration, and a clearer understanding of the impact of administered treatments, improving quality of care and lowering costs [1]. As “time to treatment” can be a key factor in morbidity and mortality rates for many dangerous conditions, continuous monitoring holds the promise to save lives.

Unfortunately, because digital medical devices can sample so frequently, they produce an overwhelming amount of data—far too much to be reviewed manually, as traditional data sources are. This data deluge is well documented ([2], [3]), and costs clinicians valuable time (see [4] for a study of how clinicians utilize their time, and how much is involved in managing health records/information retrieval). In an attempt to cope, clinicians usually either downsample the data, leading to loss of information that eliminates much of the promised benefit of continuous monitoring, or they are forced to rely on threshold alarms.

Threshold alarms are an attempt to allow clinicians to passively monitor a patient’s state continuously, with the device drawing their attention whenever the patient transitions from a normal, healthy state to a dangerous, unhealthy state. Threshold alarms are popular because they are simple, easy to implement, and easy for humans to understand. While threshold alarms can be vital in the timely detection of emergency states [5], [6], extensive research has shown them to be extremely limited, as they are unscientific in nature [7] and have a high rate of false alarms [5], [8]. The large number of erroneous alarms produced by medical devices in the ICU have been shown to cause clinicians to suffer from *alarm fatigue*, a desensitization to the presence of these alarms that causes clinicians to lose trust in the systems and begin ignoring them [9], [10].

Often, in an effort to reduce the number of alarms and seek relief from alarm fatigue, clinicians attempt to re-adjust settings on the monitor, or turn off certain alarms altogether [11]. Alarm adjustments may impede the ability to detect early signs of deterioration resulting in decreased quality of care [5], [12], [6]. Thus, the huge number of false alarms produced by bedside monitors is one of the most pressing issues in clinical care today ([13]). The Association for the Advancement of Medical Instrumentation, the US Food and Drug Administration, the Joint Commission, the American College of Clinical Engineering, and the ECRI Institute have all highlighted a link between alarm fatigue and patient deaths in their joint report from the 2011 Medical Device Alarms Summit [14]. In 2014,

reducing alarm fatigue was considered so vital that the Joint Commission declared alarm management a national patient safety goal [15]:

Clinical alarm systems are intended to alert caregivers of potential patient problems, but if they are not properly managed, they can compromise patient safety.... There is general agreement that this is an important safety issue.

A reduction in the number of false alarms has been shown to improve patient safety [11], and various efforts have been made to reduce alarm fatigue by reducing false alarms. These efforts typically focus on improving *workflow*, establishing appropriate patient-customized thresholds, or identifying situations where alarms are not clinically relevant and shutting them off [14], [16], [17].

While efforts to improve workflow or disable irrelevant alarms have had some impact on reducing alarms, the fundamental problem remains: isolated threshold alarms cannot capture sufficient nuance in patient state to eliminate false alarms, and provide little assistance to clinicians seeking to make tangible decisions about patient care. Clinicians most often use multiple vital signs in concert to understand a patient’s state. For example, a low heart rate (bradycardia) can be normal and healthy. However, if a low heart rate occurs in conjunction with an abnormal blood pressure or a low blood oxygen level, this can be cause for concern. Thus, there has been a call for the development of *smart alarm* systems, which would consider multiple vital signs before triggering an alarm. The hope is that this would reduce false alarms, thus reducing alarm fatigue and leading to improved care. Such a smart alarm system can be viewed as a simple clinical decision support system [18]; it would combine multiple sources of patient information with preexisting health knowledge to produce more accurate alarms that could be used more reliably by clinicians to make decisions about patient care.

Unfortunately, efforts to create smarter alarm systems have been hampered by a number of challenges. Perhaps the most fundamental challenge often faced is a lack of understanding of the number and types of alarms going off in the hospital environment, and which, if any, of these alarms are considered useful by clinicians. A lack of medical device interoperability hampers efforts to collect physiologic signals and alarm records; even within a single hospital, medical devices are designed and manufactured by numerous different vendors, who lack an incentive to expose interfaces for accessing their data in a systematic way. Even when this data is available, retrospective analysis of how alarms may have impacted care is difficult, if not impossible.

To that end, we set out to develop a novel data set including full records of all physiologic signals captured and bedside alarms which sounded in a surgical ICU. In order to understand which alarms were deemed useful by clinicians, we developed a novel, touchscreen-based method to collect clinician annotations on alarm usefulness in real time. In this work, we describe the data collection methodology we used, and give a broad overview of the dataset collected. We leverage

the contents of the dataset to provide an in-depth look at the alarm environment of the surgical ICU, with an eye to our planned future work of building smarter alarm systems.

II. BACKGROUND

Before digital medical devices became widespread, clinicians used analog devices to measure aspects of patient physiology, which was often laborious and time-consuming. Because of the work involved, measurements were usually taken at a low frequency, such as once per hour or once per day. Clinicians would record this information by hand (often in paper records) and review it by visual inspection [19]. Though many analog sensors are still in routine use, the advance of digital technology has enabled the development of a huge number of new digital sensors that have quickly become part of routine clinical care. Many digital sensors can collect data at high frequencies (ranging from 0.01 Hz for devices like continuous glucometers [20] to 100 – 300 Hz for sensors measuring electrical signals generated by the body, like electrocardiograms and electroencephalograms [21]).

Digital medical devices have also enabled new technologies for improved care. For example, as digital signals can also be transmitted easily across long distances, remote monitoring of at-risk patients (often dubbed “telemedicine”) is now possible, expanding the opportunity for rapid response and thus better patient care [22], [23], [24]. In practice, these new technologies mean that clinicians today are not only responsible for physical care of their patients, but are also responsible for coordinating and comprehending patient data over time and among multiple providers and settings, which requires processing and managing vast amounts of information [25].

A. Clinician Data Access

While the amount of available patient data increased dramatically over the past several decades, the way in which clinicians interact with this data has failed to change significantly. Two major ways that clinicians access physiologic data are through electronic health records, and by data visualized on patient bedside monitors.

1) *Electronic Health Records*: The deluge of medical data in the modern hospital quickly made the limitations of paper-based information management clear [26], and this coupled with mandatory government requirements lead to a rapid shift toward storing patient data digitally in *electronic medical records* (EMRs) [1].¹ While EMRs have become ubiquitous within the United States, they serve primarily to *store* patient data; it is usually difficult or impossible to access, visualize, or analyze data from within the EMR system. Without additional technology, EMRs are “essentially just copies of paper-based records stored in electronic form” [27], but containing an unwieldy amount of data. Additionally, due to lack of interoperability between medical device sensors and EMRs, as well as storage size constraints, it is rare for continuous physiologic data to be stored in EMRs. Most commonly, vital

¹Electronic medical records are also called electronic health records (EHRs). The terms are used interchangeably in the literature.



Fig. 1. A bedside monitor at Penn Presbyterian Medical Center in an (empty) patient room. An annotation tablet, with our application running, is mounted below the lower right side of the screen (see Section IV-A).

signs are read from a medical device display (such as a bedside monitor) and hand-entered into a patient’s EMR, resulting in severely downsampled data.

2) *Bedside Monitors*: Typically, within a patient’s room, a number of different vital sign sensors are connected to a single bedside monitor, which displays sensed data in an attempt to allow clinicians tending to the patient to easily view all of a patient’s current vital signs on one screen. Bedside monitors most commonly display the patient’s electrocardiogram, (if the patient has electrodes placed), respiration rate, heart rate, blood oxygen saturation (via a pulse oximeter), and arterial blood pressure (if a patient has an arterial line inserted). A photo of a bedside monitor can be seen in Figure 1.

Many devices in the ICU, including ventilators, dialysis machines, feeding machines, and even patient beds, produce alarms. However, owing to their ability to connect to a number of different sensors and their presence in every patient room, bedside monitors are often the source of the majority of alarms in the ICU. In particular, many of the alarms produced by bedside monitors are threshold alarms.

B. Threshold Alarms

Many medical devices currently in use, including bedside monitors, are configured to generate *threshold alarms*, alarms that activate when a specified physiologic signal being measured crosses a predefined threshold [28], [29]. An illustration of a threshold alarm is shown in Figure 2. Threshold alarms allow clinicians to set a certain value (the *threshold*) for a specific physiologic signal being monitored. If the value of the signal crosses the chosen threshold (either by exceeding it, in the case of a high threshold, or dropping below it, in the case of a low threshold) an alarm is produced to alert a clinician. Threshold alarms may be *latched*; latched alarms will continue

to sound even if the signal re-crosses the threshold to return to the “normal” zone.

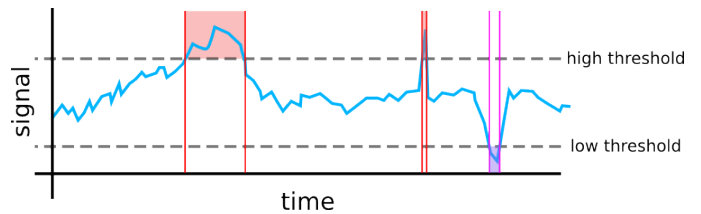


Fig. 2. Illustration of two high threshold alarms (shaded in red) and one low threshold alarm (shaded in purple).

As previously described, threshold alarms are particularly susceptible to false alarms [5], [8]. These false alarms can be caused by non-significant, random fluctuations in the patient’s physiologic signals, or by noise due to external stimuli (the most common example is patient movement, which can cause sensors to move, get compressed, or detach from the patient entirely). Threshold alarms are particularly sensitive to these fluctuations for a number of reasons. First, though clinicians can usually manually adjust thresholds to tailor them to their patients, many medical devices have interfaces which make this time-consuming and difficult. Thus, thresholds are often left at default values, which are not appropriate for many patients. A threshold alarm set close to a patient’s normal baseline will alarm for even small or brief deviations in that baseline. Second, by design, threshold alarms cannot take advantage of trend information or waveform shape information that would make them less sensitive to noise or artifact. Third, each threshold alarm is applied to a physiologic data source individually, and so alarms are generated using no contextual information from other physiologic signals. As a result, an alarm triggered from a single source threshold violation may not be cause for concern. For these reasons, threshold alarms are on the whole very poor at separating true changes in patient state from random noise fluctuations. Worse still, even when they do capture a true change in patient state, the presence of an alarm itself provides no contextual information beyond the fact that a single physiologic signal had a value that was outside the threshold at some point in the past [5], [30]. It becomes the task of the clinician to piece together retrospective clues as to why an alarm occurred, which can be an arduous process for even one alarm; the sheer number of alarms going off typically makes this impossible.

III. RELATED WORK

Few studies of physiologic signals have focused on incorporating alarms from existing alarm systems. [31] combined lab tests with monitor alarms to build “SuperAlarm” patterns, which provided fair accuracy at detecting code blue events while maintaining low rates of false alarm. There have been more significant efforts to filter out alarms from existing systems. [17] describes a system that filters out alarms by integrating multiple vital signs using “fuzzy logic.” The system dramatically reduced false alarms and increased sensitivity,

meaning clinicians would be presented with more useful alarms. [32] describes online signal filtering of threshold alarms. However, on the whole, analysis of which alarms are useful is often difficult, and learning to categorize alarms based on utility is rare.

There have been numerous attempts to increase the accuracy of alarm systems by using multiple vital signs in concert. [33], [34], [35] created and evaluated a smarter alarm system by hand-curating sets of rules from clinicians that combined vital signs and produced alerts only when nurses reported that these combinations were indicative of concern. The system was tested with hand-curated data from patients and manual, individual observations, reporting a dramatic decrease in false alarm rate with no true missed events. Alarms during anesthesia have also been the target of decision support development efforts. [36] developed a diagnostic alarm system prototype for monitoring pathological events during anesthesia. [37] created a clinical decision support system to help anesthesiologists identify critical events in patients under spinal analgesia with sedation. These systems reduced time to detect and treat, and the latter reduced the number of episodes of hypoxemia.

Despite these efforts, there remain numerous roadblocks to developing improved alarm systems: access to physiologic signals from patients is often restricted by device manufacturers, and most hospitals do not regularly quantify the number and types of alarms which go off in the ICU, nor do they assess which of these alarms were useful. Without baseline alarm information, it can be difficult to definitively demonstrate that a new clinical decision support system improves over threshold alarms. Few studies have attempted to address these issues. As a followup to their previous work, the most recent alarm study by King *et al.* collected alarm responses through prospective observations, which was labor intensive and limited the amount of data which could be collected [35]. A landmark study by Bonafide *et al.* describes efforts to use video to capture physiologic monitor alarms and clinician responses to those alarms [38]. Annotations were added retrospectively through manual review of video footage by clinicians, which dramatically increased the amount of data which could be collected, but was also labor intensive. In contrast, we believe our study is the first to describe a mechanism to capture annotations of alarms by clinicians in a minimally invasive way with no additional retrospective observational labor.

IV. METHODOLOGY

We sought to collect a novel data set that would allow us to gain a better understanding of how many alarms go off in the ICU and which of them are useful. We provide a summary of the data set collected, then present our analysis scheme along with preliminary results.

A. Alarm Data Set Collection Methods

We developed this framework for collecting data at Penn Presbyterian Medical Center. Penn Presbyterian is an urban tertiary care teaching hospital with 331 beds. In each surgical ICU room, a Philips IntelliVue MP70 physiologic monitor is

mounted at the bedside. (Each room houses a single patient.) Our objective was to collect all physiologic signals which passed through these bedside monitors and all generated alarms, as well as a set of annotations by clinicians which would indicate which alarms were useful or not, from sixteen beds in the surgical ICU for a period of fifteen months.

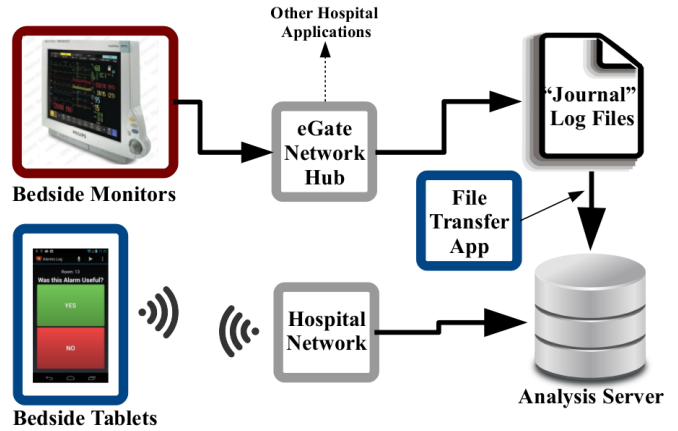


Fig. 3. Architecture of the alarm data set collection scheme.

Our initial attempts to collect physiologic signal data directly from bedside monitors was unsuccessful, as data output by monitors was encoded in a proprietary fashion, and existing hospital infrastructure was not equipped to support this access point. However, bedside monitors were already configured to send data across the hospital’s network to downstream telemedicine applications via their “eGate” Network Hub. As such, we partnered with Penn Medicine’s Information Services department, who logged these HL7 data packets sampled from the hospital network once per minute as “Journal” files. While the data was at a lower frequency than ideal, using the existing data logs made data acquisition straightforward.

In order to capture alarm data, we requested that the manufacturer of the physiologic monitors in use at the hospital (Philips) enable a software switch that allowed bedside monitors to output alarm messages to the hospital network. In this way, alarm data would be logged with the aforementioned network data as HL7 data packets at a rate of once per minute. Hospital information technology requested an initial testing period to ensure that the hospital network could handle these additional messages and that they did not impact downstream applications. In order to access the logged vital sign and alarm messages, we then developed an File Transfer Application which would first fully de-identify the journal files within the hospital’s firewall, then automatically and securely transfer them from the Information Services department to a secure server for analysis. The architecture of the overall process was IRB approved, and is shown in Figure 3.

In order to collect annotations on alarm utility from clinicians with minimal disruption to clinician workflow, we developed a simple mobile annotation collection application for the Android operating system, and deployed it on Google Nexus 7 tablets. Nexus 7 tablets were selected for their low price and

their large, easily visible displays. A tablet was placed in each patient room of the target surgical ICU, as well as at the two central nursing stations (where each nursing station monitor displays vital signs and alarms alerts for half of the rooms in the ICU). The tablets were affixed using adhesive velcro in an unobtrusive location below the physiologic monitor. (A photo of a tablet affixed to a bedside monitor can be seen in Figure 1.) The tablets were placed near the “alarm silence” button on the bedside monitors (located on the lower right of the monitor screen), for ease of use.

For the tablets mounted at the bedside, the application displays a question (“Was this Alarm Useful?”) and two large buttons (a green “Yes” button and an orange-red “No” button). Both buttons were labeled in addition to color-coding, in case of colorblind users. In addition to the in-room monitor, there exists a central monitoring station located within clinicians’ core workstation. Information from each patient housed in a certain section of the unit is displayed on a central monitor to allow for easy viewing by clinicians who may not be in the room at the time an alarm is triggered. For the tablets mounted at these central monitoring stations, the application displays the same question and two buttons for each of the rooms that that nursing station covered. A screenshot of the application interfaces can be seen in Figure 4. Tablet time was synchronized to NTP time via the tablet’s Wi-Fi connection. Clinicians were invited to provide annotations by pushing the button which appropriately described their view of the utility of current alarms, when it would not cause them an undue burden in the course of patient care. Records of buttons pushed (which did not contain any sensitive information) were transmitted via the hospital’s Wi-Fi connection to a secure server for analysis.

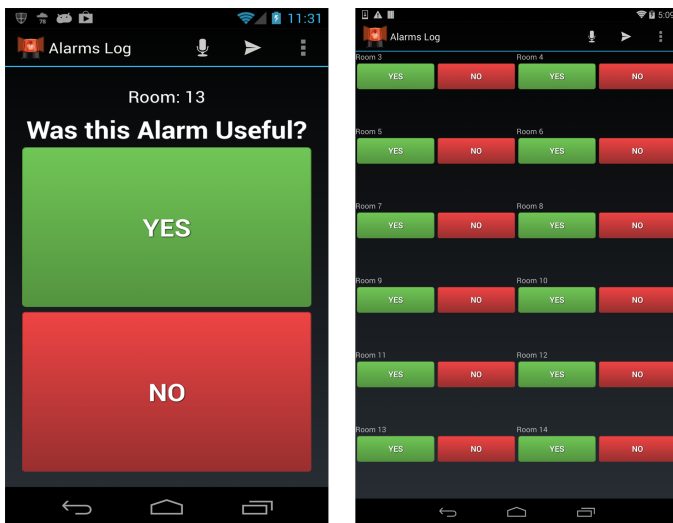


Fig. 4. Screenshot of the annotation applications allowing clinicians to provide alarm annotations. On the left, the display presented at the bedside in each room. On the right, the display presented at the nursing stations, which allows annotation of multiple rooms.

V. RESULTS

We collected the aforementioned data from the surgical ICU for a period of fifteen months (from March 2014 to June 2015). The unique data set we collected has allowed us to quantify the amount of physiologic data that passes through the bedside monitor, the number and type of bedside monitor alarms which sound in the surgical ICU, and to describe which of these alarms were annotated as useful/not useful by clinicians.

A. Data Set Description

In this section, we provide an overview of the physiologic signals, alarms, and clinician annotations collected.

1) *Physiologic Signals*: For each ICU room we collected an average of 828.8 hours of physiologic signal data per month, where physiologic data was recorded at a frequency of 0.016 Hz (one sample per minute). When physiologic signals were unavailable for a room, it was often because the ICU room in question was empty (in anticipation of a patient, after a patient was discharged, or during cleaning). Table I displays number of samples and percent time available for physiological signals which were present for more than one percent of the time when data was available. Note that the majority of recorded signals

Signal Name	Count	Percent Present
Heart Rate	5517	99.8
Premature Ventricular Contraction	5510	99.7
Respiratory Rate	5333	96.5
Pulse Rate	5077	91.9
Pulse Oxygenation	4723	85.4
Mean Arterial Blood Pressure	3339	60.4
Diastolic Arterial Blood Pressure	3313	60.0
Systolic Arterial Blood Pressure	3313	60.0
Mean Central Venous Pressure	2289	41.4
Mean Pulmonary Arterial Pressure	1159	21.0
Diastolic Mean Pulmonary Arterial Pressure	1139	20.6
Systolic Mean Pulmonary Arterial Pressure	1139	20.6
Signal Quality Index	160.7	2.9
Electromyograph	158.0	2.9
Bispectral Index	142.6	2.6
Aortic Mean Pressure	136.2	2.4
Aortic Diastolic Pressure	135.7	2.4
Aortic Systolic Pressure	135.7	2.4
Suppression Ratio	124.7	2.2
Mean Noninvasive Blood Pressure	86.78	1.6
Diastolic Noninvasive Blood Pressure	84.08	1.6
Systolic Noninvasive Blood Pressure	84.08	1.6
Core Temp	57.69	1.0

TABLE I
AVERAGE COUNT OF PHYSIOLOGIC SIGNALS RECORDED PER ROOM IN ONE MONTH.

were respiration rate (RR), heart rate (HR), pulse oxygenation (SpO_2), and, less commonly (as it requires a patient had an arterial line) arterial blood pressure (ABP). Though the data set occasionally contains other signals, these four “vital signs” are sensed for the majority of patients. No other signals routed through the bedside monitor were captured nearly as often. For this reason, further analysis will focus on these four signals (with blood pressure occasionally expressed in its systolic and diastolic components).

Signals displayed by the bedside monitor do not encompass all physiologic signals used by clinicians to make decisions about the care of their patients. ² We chose to focus on these signals as they are monitored for all patients and tend to be the primary signals used by clinicians as they monitor a patient’s overall status. While the frequency of the physiologic data we acquired for the dataset was not as high as we had originally hoped, we expect that the granularity of the data is high enough to capture adverse the majority of signal changes that would truly indicate the need for an alarm.

2) *Alarms:* Philips IntelliVue MP70 bedside monitors include three clinical alarm levels (“Green,” “Yellow,” and “Red” alarms, ranging from least urgent to most urgent, respectively). The vast majority of alarms issued in the ICU are Green (mostly technical alarms) and Yellow (which include vital sign threshold alarms). Yellow and Red alarms produce auditory alerts and visual on-screen alerts; Green alarms only produce visual alerts. All alerts can be silenced for a short period of time through a button on the bedside monitor or nursing station.

First, we calculated the average number of hours out of a twenty-four hour day that some alarm is going off for each clinical alarm level (see Figure 5). For each bed a yellow or red alarm is sounding for almost 11 out of the 24 hours of an average day. As ICUs at large hospitals often have between 16 and 20 beds, yellow alarms are sounding near-constantly.

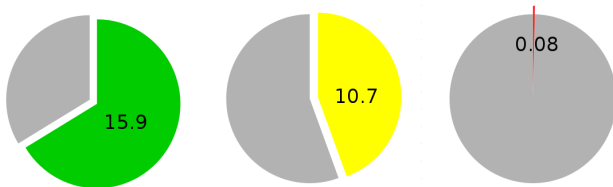


Fig. 5. For each bed, the number of hours of each day (out of 24) where some alarm is going off for, from left to right, “Green”, “Yellow”, and “Red” alarm levels.

Table II lists the most common bedside monitor alarms, along with their types. The “Hours Per Bed Per Month” column displays the number of hours said alarm went off for a single bed over the course of the average month of monitoring. (For space reasons, alarms which went off for less than one hour per month not included in the table.) Among the most common alarms are numerous threshold alarms, particularly those for SpO_2 (which is a particularly noisy vital sign) as well as for ABP (which often has particularly tight threshold cutoffs). Also high on this list are alarms related to premature ventricular contractions (PVC), which trigger when a patient experiences a particular type of irregular heartbeat. In Figure 6, we plot each alarm type’s alarming frequency (the average hours per bed per month that the alarm is triggered) ordered from most frequent to least. This figure illuminates the types

²For example, ventilator settings and corresponding detailed respiratory information that are vital for detecting many forms of acute respiratory distress are often not routed through the bedside monitor.

Alarm Name	Type	Hours Per Bed Per Month
**SpO <	YELLOW	9.65
ABP Reduce Size	GREEN	8.70
**ABPm >	YELLOW	8.69
Some ECG AlarmsOff	GREEN	8.26
SpO Sensor Off	GREEN	7.33
**ABPm <	YELLOW	6.83
SpO Low Perf	GREEN	3.82
**RR <	YELLOW	3.40
CVP Reduce Size	GREEN	3.28
PAP Reduce Size	GREEN	3.26
EcgOut Equip Malf	GREEN	3.01
** Pair PVCs	YELLOW	2.97
Resp Leads Off	GREEN	2.77
NBP Interrupted	GREEN	2.69
**RR >	YELLOW	2.22
**ABPs <	YELLOW	2.20
** Multiform PVCs	YELLOW	1.65
**CVpm >	YELLOW	1.64
**PAPd >	YELLOW	1.63
SpO No Pulse	GREEN	1.60
** Irregular HR	YELLOW	1.59
LA Lead Off	GREEN	1.29
**ABPs >	YELLOW	1.19
**HR >	YELLOW	1.18
RL Lead Off	GREEN	1.14

TABLE II
AVERAGE MONTHLY ALARM DURATION FOR THE MOST COMMON ALARMS IN A SURGICAL ICU. ONLY ALARMS WHICH SOUNDED

Alarm Counts

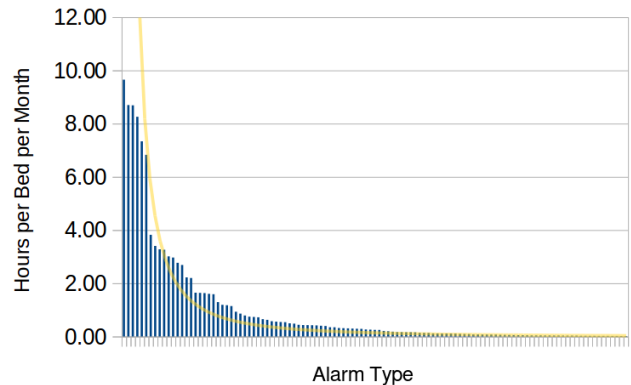


Fig. 6. Frequency of alarms plotted in order from most frequent to least frequent. A power curve fit to the data is superimposed in yellow. A long tail of alarms which occurred no more than once per month was truncated to improve graph legibility.

of alarms that contribute to alarm density in the ICU. A few types of alarms make up a bulk of the alarms which go off; while the distribution has a long tail (there are many different alarm types), most of these are very rare. Thus a reduction in the most common alarms could have a significant impact in overall alarm burden. This coupled with even simple counts of which alarms are going off the most can have a dramatic impact on practice of care. Six months into the data collection period, armed with the information that Pair PVCs were the most common yellow alarms (sounding for hundreds of hours

each month), the surgical ICU in question decided to begin automatically suppressing Pair PVC alarms, as the alarms were already considered non-actionable and thus only served as a nuisance, which eliminates nearly three hours of nuisance alarm sounds per bed per month.

3) *Annotations*: Annotations, while the most novel part of the dataset, are also the most challenging to analyze. Over the course of the study, we collected over 3500 unique annotations. We suspect some of these are erroneous, as some occur when no vital signs are being recorded on the bedside monitor.³ However, most of these erroneous annotations can be removed with little difficulty; we have attempted to remove as many as possible, and have identified many specific cases of definitively labeled alarms in the annotations that remain.

Figure 7 displays the cumulative number of each type of annotation, as well as total annotations, acquired during the data collection period. On the whole, clinicians provided significantly more annotations indicating alarms were not useful than annotations indicating alarms were useful, lending further credence to initial discussions of alarms as mostly false. The figure also illustrates that as the data collection period progressed, clinicians provided fewer annotations, an expected result of clinicians being desensitized to the presence of the annotation tablets. In anticipation of this, we scheduled four separate visits to the ICU over the course of the data collection period to remind clinicians of the goals of the ongoing study and encouraging their participation. Arrows indicate times when educational inservices were held, and we note increases in provided annotations after each of these inservices. The flat area on the graph, marked with an asterisk, demarcates a period during which the ICU in question relocated to a different wing of the hospital and thus we were forced to suspend annotation collection.

In Table III, we break down the numbers of annotations made per month by the two tablet locations, the bedside tablets and the nursing station tablets. On the whole, more annotations were made via the nursing station tablets than via the bedside tablets, which mirrors feedback provided by nurses, who indicated they often silence alarms from the nursing stations as it is more efficient (they can review multiple rooms at once) and does not require disturbing the patient. Second, we note that non-useful alarms were more often logged from the tablets mounted at the central monitoring stations, while useful alarms were more often logged from the bedside tablets. As clinicians were instructed to provide annotations in the course of their usual care, we hypothesize that a useful alarm would be more likely to prompt a nurse to visit a patient at the bedside, where they would enter their annotation, while obviously false alarms could safely be dismissed from the nursing station, outside the patient’s room.

The design of our annotation collection procedure was intended to minimize the impact of providing annotations on clinicians. As such, annotations were provided voluntarily, not

³Some of these erroneous annotations likely occurred when the tablets were being wiped with cleaning solution while the room was empty.

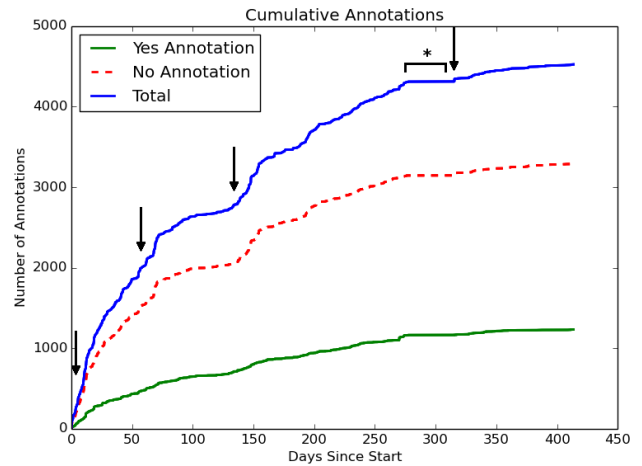


Fig. 7. Plot of the number of annotations collected over time. The red dashed line indicates “No” annotations (the clinician indicated the current alarms were not useful), the green (lower) line indicates “Yes” annotations (the clinician indicated the current alarms were useful), and blue line is the cumulative total. Arrows indicate points at which educational inservices were held.

	Bedside	Central Station
No	92.67	204.8
Yes	57.56	44.22
Total	150.2	249

TABLE III

AVERAGE NUMBERS OF ALARM ANNOTATIONS PROVIDED BY CLINICIANS VIA TABLETS, PER MONTH, WITH BREAKDOWNS BY LOCATION AND TYPE. “No” ANNOTATIONS INDICATED THE CURRENT ALARM WAS NOT USEFUL. “Yes” ANNOTATIONS INDICATED THE CURRENT ALARM WAS USEFUL.

at random, and the annotations may be biased. For example, clinicians may have provided annotations only for those alarms they felt most strongly about, or those alarms which were easiest to categorize. However, as the goal of this data set is to enable the creation of smarter alarm systems, and any such system must be designed with clinician satisfaction in mind, such bias may be a useful indication of clinicians’ priorities when addressing alarms. Additionally, as we have previously described, the alarm environment is so saturated with alarms that any reduction in false alarms would lead to improvement.

The proliferation of alarms in the ICU means that the number of annotations provided is far smaller than the total number of alarms, meaning many alarms in the data set do not have annotations. Again, we believe that even a small subset of annotated alarms could help guide the development process for a smarter alarm system. Previous efforts to develop such systems relied on far fewer annotations (owing to the labor involved in curating them), but were still able to achieve reductions in false alarm rates. We believe our work only builds on these efforts. Information about clinician opinion of alarm utility as provided by our annotation set has not previously been available at scale in alarm reduction research. We believe a large enough set of these sort of annotations, logged in real-time by clinicians with minimal impact to their

workflow, provides a major missing component of current research efforts, which we will capitalize on in future work.

B. Multivariate Analysis

While we reserve more in-depth analysis for future work, in an effort to illustrate the relationship between the four major vital signs we captured, we performed a preliminary multivariate analysis. Figure 8 plots several thousand pairs of these four major vital signs, drawn from a random subset of patients. To highlight major clusters of typical vital sign values, we used a one-class SVM [39] with a radial basis function kernel ($\nu = 0.5$, $\gamma = 0.1$) and set our outliers fraction to $\%0.5$. This simple plotting alone reveals several relationships that could not be captured by simple threshold alarms. For example, it is clear that there is a non-linear relationship between respiration rate (RR) and SpO_2 ; Any single threshold on RR would thus produce a large number of alarms. While 8 does not immediately reveal a solution for creating smarter alarms, it does suggest relationships between vital signs might be useful in further investigation.

VI. CONCLUSION

This work outlines a new data set intended to enable the study of physiologic monitor alarms. We outlined the collection process for the data contained in the data set, including a novel clinician annotation collection mechanism using tablets affixed to bedside monitors in patient rooms. We described the data collected, attempting to provide some insight into the type and frequency of the large number of alarms present in the surgical ICU, and described the annotations provided by clinicians in response to those alarms.

Preliminary analysis of the described data set confirms that threshold alarms occur with high frequency in the surgical ICU, a large portion of these alarms are deemed not useful by clinicians, and that threshold alarms are inadequate to identify meaningful changes in physiologic signals. Our description of the data hints at numerous interesting relationships between physiologic signals, and planned future work will build on this foundation to provide further insights into strategies for false alarm reduction.

ACKNOWLEDGMENT

This research was supported in part by NSF CNS 1035715 and the DGIST Research and Development Program of the Ministry of Science, ICT and Future Planning of Korea (CPS Global Center).

REFERENCES

- [1] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, p. 3, 2014.
- [2] Frost & Sullivan, "Drowning in Big Data? Reducing Information Technology Complexities and Costs For Healthcare Organizations," Tech. Rep., 2012.
- [3] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbes, C. Roxburgh, and A. H. Byers, "Big data : The next frontier for innovation , competition , and productivity," Tech. Rep. June, 2011.
- [4] Z. Tang, L. Weavind, J. Mazabob, E. J. Thomas, M. Y. L. Chu-Weininger, and T. R. Johnson, "Workflow in intensive care unit remote monitoring: A time-and-motion study*," *Critical care medicine*, vol. 35, no. 9, pp. 2057–2063, 2007.
- [5] Clinical Alarms Task Force, "Impact of clinical alarms on patient safety," *Journal of Clinical Engineering*, vol. 32, no. 1, pp. 22–33, 2007.
- [6] M. Imhoff and S. Kuhls, "Alarm algorithms in critical care monitoring," *Anesthesia and Analgesia*, vol. 102, no. 5, pp. 1525–1536, 2006.
- [7] L. A. Lynn and J. P. Curry, "Patterns of unexpected in-hospital deaths: a root cause analysis," *Patient Safety in Surgery*, vol. 5, 2011.
- [8] S. Siebig, L. Juhls, M. Imhoff, U. Gather, U. Scholmerich, and C. Wrede, "Plug-and-play for medical devices: Experiences from a case study," *Crit Care Med*, vol. 38, no. 2, pp. 451–455, 2010.
- [9] S. Siebig, S. Kuhls, M. Imhoff, U. Gather, J. Schölmerich, and C. E. Wrede, "Intensive care unit alarms—how many do we need?," *Critical care medicine*, vol. 38, no. 2, pp. 451–6, Feb. 2010.
- [10] B. J. Drew, P. Harris, J. K. Zègre-Hemsey, T. Mammone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, and X. Hu, "Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients," *PLoS one*, vol. 9, no. 10, p. e110274, Jan. 2014.
- [11] J. Edworthy and E. Hellier, "Fewer but better auditory alarms will improve patient safety," *Quality & safety in health care*, vol. 14, no. 3, pp. 212–215, Jun. 2005.
- [12] Y. Donchin and F. J. Seagull, "The hostile environment of the intensive care unit," *Current opinion in critical care*, vol. 8, no. 4, pp. 316–320, 2002.
- [13] T. Clark, Y. David, M. Baretich, and T. Bauld, "Impact of clinical alarms on patient safety," ACCE Healthcare Technology Foundation, Tech. Rep., 2006.
- [14] "Clinical Alarms - Priority Issues from the Medical Device Alarms Summit," Association for the Advancement of Medical Instrumentation, Tech. Rep., 2011.
- [15] T. J. Commission, "National patient safety goal on alarm management," *Joint Commission Perspectives*, vol. 33, Jul. 2013.
- [16] G. Clifford, W. Long, G. Moody, and P. Szolovits, "Robust parameter extraction for decision support using multimodal intensive care data," *Philosophical Transactions of the Royal Society*, vol. 367, no. 1887, pp. 411–429, Jan. 2009.
- [17] C. Oberli, J. Urzua, C. Saez, M. Guarini, A. Cipriano, B. Garayar, G. Lema, R. Canessa, C. Sacco, and M. Irarrazaval, "An expert system for monitor alarm integration," *Journal of Clinical Monitoring and Computing*, vol. 15, no. 1, pp. 29–35, 1999.
- [18] A. X. Garg, N. K. J. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. J. Devereaux, J. Beyene, J. Sam, and R. B. Haynes, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review," *Jama*, vol. 293, no. 10, pp. 1223–38, Mar. 2005.
- [19] M. Elliott and A. Coventry, "Critical care: the eight vital signs of patient monitoring," *British Journal of Nursing*, vol. 21, no. 10, pp. 621–625, 2012.
- [20] S. Finfer, J. Wernerman, J.-C. Preiser, T. Cass, T. Desai, R. Hovorka, J. I. Joseph, M. Kosiborod, J. Krinsley, I. Mackenzie, D. Mesotten, M. J. Schultz, M. G. Scott, R. Slingerland, G. Van den Berghe, and T. Van Herpe, "Clinical review: Consensus recommendations on measurement of blood glucose and reporting glycemic control in critically ill adults," *Critical care (London, England)*, vol. 17, no. 3, p. 229, Jan. 2013.
- [21] National Instruments, "Biomedical Engineering Education Portal: Typical Measurements in Bioinstrumentation Classes," National Instruments, Tech. Rep. October 2005, 2013.
- [22] B. A. Kohl, M. Fortino-Mullen, A. Praestgaard, C. W. Hanson, J. DiMartino, and E. A. Ochroch, "The effect of icu telemedicine on mortality and length of stay," *Journal of telemedicine and telecare*, vol. 18, no. 5, pp. 282–286, 2012.
- [23] M. Mullen-Fortino, F. D. Sites, M. Soisson, and J. Galen, "Innovative use of tele-icu in long-term acute care hospitals," *AACN advanced critical care*, vol. 23, no. 3, pp. 330–336, 2012.
- [24] M. Mullen-Fortino and J. Dimartino, "Evaluation of critical care telemedicine nursing interventions on patient outcomes, professional practice, and organizational processes," *Computers Informatics Nursing*, vol. 27, no. 5, pp. 335–336, 2009.
- [25] M. Peleg and S. Tu, "Decision Support , Knowledge Representation and Management in Medicine," *IMIA Yearbook of Medical Informatics*, vol. 45, pp. 72–80, 2006.

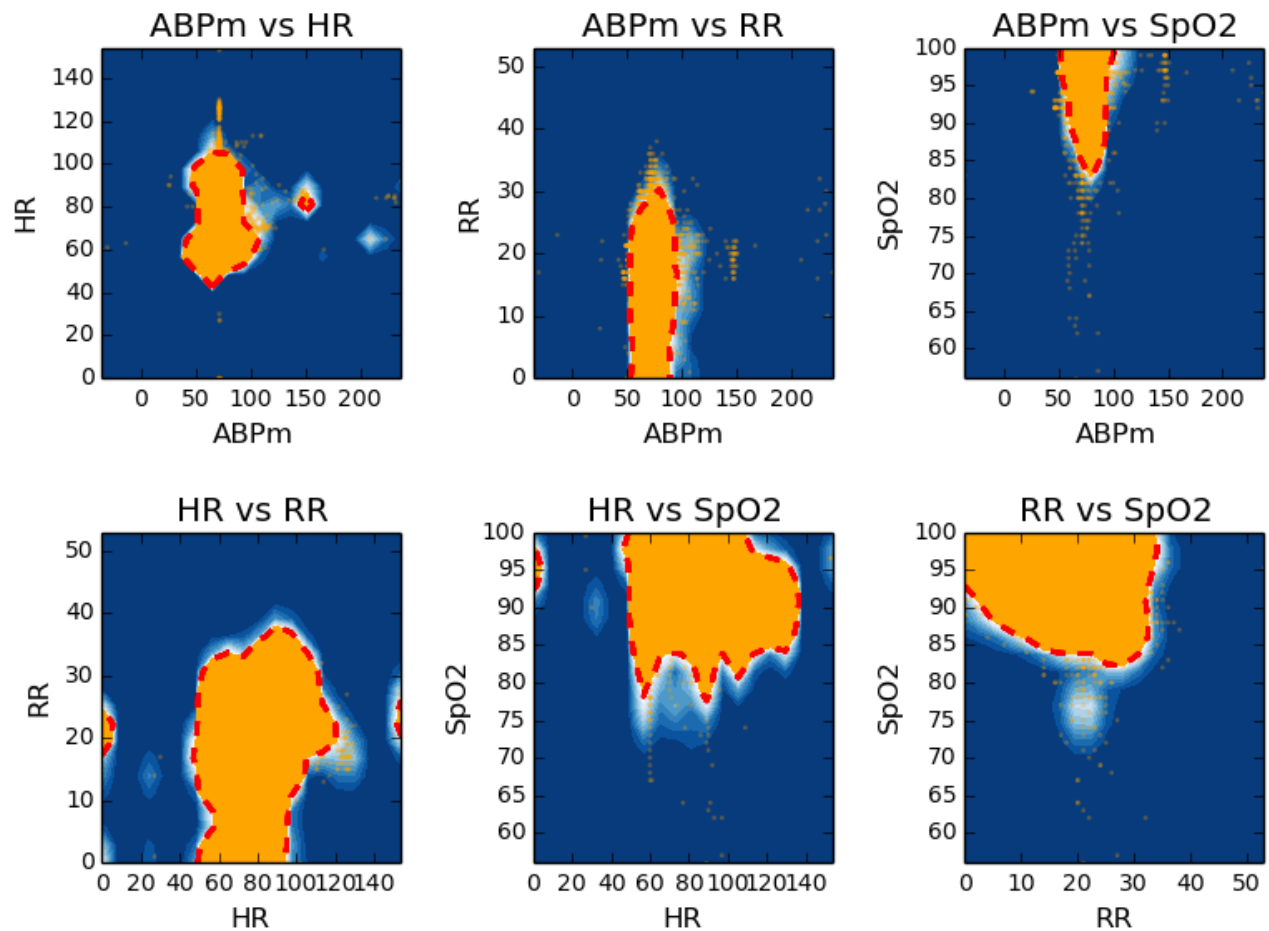


Fig. 8. Plots of each pair of four most commonly collected physiologic signals. Clusters are outlined with a red dashed line. Density of points represented by blue-white-orange (low to high) color grade.

[26] B. Chaudhry, J. Wang, and S. Wu, "Systematic review: impact of health information technology on quality, efficiency, and costs of medical care," *Annals of internal medicine*, vol. 144, pp. 742–752, 2006.

[27] C. Bennett and T. Doub, "Data mining and electronic health records: Selecting optimal clinical treatments in practice," in *Proceedings of the 6th International Conference on Data Mining*, 2011, pp. 313–318.

[28] A. Otero, P. Félix, S. Barro, and F. Palacios, "Addressing the flaws of current critical alarms: a fuzzy constraint satisfaction approach," *Artificial Intelligence in Medicine*, vol. 47, no. 3, pp. 219–238, Nov. 2009.

[29] F. Mora, G. Passariello, G. Carrault, and J.-P. Le Pichon, "Intelligent patient monitoring and management systems: a review," *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, no. 4, pp. 23–33, Dec. 1993.

[30] S. Lawless, "Crying wolf: False alarms in a pediatric intensive care unit," *Critical Care Medicine*, vol. 22, pp. 981–985, 1994.

[31] Y. Bai, D. H. Do, P. R. E. Harris, D. Schindler, N. G. Boyle, B. J. Drew, and X. Hu, "Integrating monitor alarms with laboratory test results to enhance patient deterioration prediction." *Journal of biomedical informatics*, Sep. 2014.

[32] M. Borowski, S. Siebig, C. Wrede, and M. Imhoff, "Reducing false alarms of intensive care online-monitoring systems: an evaluation of two signal extraction algorithms," *Computational and Mathematical Models in Medicine*, vol. 2011, Jan. 2011.

[33] A. King, A. Roederer, D. Arney, S. Chen, A. Giannareas, M. Mullenfortino, C. W. I. Hanson, V. Kern, N. Stevens, J. Tannen, A. V. Trevino, J. Wanderer, S. Park, O. S. P. D., I. L. P. D., and A. Vi, "GSA: a framework for rapid prototyping of smart alarm systems," in *Proceedings of the 1st ACM International Health Informatics Symposium*, 2010.

[34] A. King, A. Roederer, S. Chen, N. Stevens, P. Asare, O. Sokolsky, I. Lee, M. Mullen-Fortino, and S. Park, "Demo of the Generic Smart Alarm: a framework for the design, analysis, and implementation of smart alarms and other clinical decision support systems," in *ACM Wireless Health 2010*, 2010, pp. 1–2.

[35] A. King, K. Fortino, N. Stevens, S. Shah, M. Fortino-Mullen, and I. Lee, "Evaluation of a smart alarm for intensive care using clinical data." in *Conference proceedings : 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, vol. 2012, Jan. 2012, pp. 166–169.

[36] B. Gohil, "Diagnostic Alarms in Anaesthesia," Ph.D. dissertation, 2007.

[37] C. Zaouter, M. Wehbe, S. Cyr, J. Morse, R. Taddei, P. a. Mathieu, and T. M. Hemmerling, "Use of a decision support system improves the management of hemodynamic and respiratory events in orthopedic patients under propofol sedation and spinal analgesia: a randomized trial." *Journal of clinical monitoring and computing*, vol. 28, no. 1, pp. 41–7, Feb. 2014.

[38] C. P. Bonafide, M. Zander, C. S. Graham, C. M. W. Paine, W. Rock, A. Rich, K. E. Roberts, M. Fortino, V. M. Nadkarni, R. Lin, R. Keren, and P. Christopher, "Video Methods for Evaluating Physiologic Monitor Alarms And Alarm Responses," no. June, pp. 220–231, 2014.

[39] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.