



Published in final edited form as:

Pac Symp Biocomput. 2018 ; 23: 472–483.

Addressing vital sign alarm fatigue using personalized alarm thresholds

Sarah Poole¹ and Nigam Shah, MBBS, PhD¹

¹Center for Biomedical Informatics Research, Stanford University, Stanford, CA, United States

Abstract

Alarm fatigue, a condition in which clinical staff become desensitized to alarms due to the high frequency of unnecessary alarms, is a major patient safety concern. Alarm fatigue is particularly prevalent in the pediatric setting, due to the high level of variation in vital signs with patient age. Existing studies have shown that the current default pediatric vital sign alarm thresholds are inappropriate, and lead to a larger than necessary alarm load. This study leverages a large database containing over 190 patient-years of heart rate data to accurately identify the 1st and 99th percentiles of an individual's heart rate on their first day of vital sign monitoring. These percentiles are then used as personalized vital sign thresholds, which are evaluated by comparing to non-default alarm thresholds used in practice, and by using the presence of major clinical events to infer alarm labels. Using the proposed personalized thresholds would decrease low and high heart rate alarms by up to 50% and 44% respectively, while maintaining sensitivity of 62% and increasing specificity to 49%. The proposed personalized vital sign alarm thresholds will reduce alarm fatigue, thus contributing to improved patient outcomes, shorter hospital stays, and reduced hospital costs.

Keywords

Alarm Fatigue; Vital Signs; Personalized Medicine; Random Forest

1. Introduction

Vital sign monitors are an important component of inpatient care, as they provide timely alerts to clinical staff in response to extreme vital sign values^{1,2}. These vital sign alarms are intended to be a safety net in the provision of patient care, but their management in the inpatient setting is a significant patient safety issue^{3,4}. Efforts to characterize vital sign alarms have shown that 64–99% of the alarms that sound are not clinically actionable⁵. The high proportion of unnecessary alarms has led to provider desensitization, also known as alarm fatigue^{6,7}. This has been shown to increase nurse response time to subsequent alarms in both the short and the long term^{8,9}, increasing the risk to patients and contributing to adverse patient events and, in some cases, patient mortality^{7,10}. In 2013, the Joint Commissions issued Sentinel Event Alert #50 to draw attention to widespread alarm fatigue in hospital settings¹⁰, and the subsequent 2014, 2015 and 2016 National Patient Safety Goals urged hospitals to prioritize alarm system safety and ensure that alarms on medical equipment are heard and responded to on time^{11–13}.

Multiple approaches have been taken to the problem of alarm fatigue, including implementing standards for checking and changing electrocardiography lead wires and electrodes^{14–17}, escalating alarms to pages sent directly to clinical staff¹⁸, adding delays to alerts to avoid alarming for very short periods of extreme values^{14,18,19}, implementing standard time series filters²⁰, and combining alarms to obtain a more general measure of patient deterioration²¹. However, more work needs to be done to address the risk to patient safety from alarm fatigue⁶.

Alarm fatigue is particularly prevalent in the pediatric setting, due to the high level of variation in vital signs with patient age^{5,6}. Two recent studies have compared default age-based pediatric vital sign alarm thresholds from hospitals with the observed vital signs in each age group. Both studies found that default heart rate alarm thresholds fall near the 50th percentile of patient data, leading to an unnecessarily large alarm load^{22,23}. These studies conclude that patient data can successfully be used to choose more appropriate thresholds for vital sign alarms, and initial efforts in this direction have been promising²³. While the thresholds produced by these studies partially account for the expected change in vital signs with age, the performance of such default thresholds is limited, since vital signs are known to change smoothly and continuously with age²⁴, rather than displaying the ‘step’ changes that result from using discrete age groups.

Existing work addressing alarm fatigue includes only limited evaluation of the safety and efficacy Existing work addressing alarm fatigue includes only limited evaluation of the safety and efficacy of the proposed measures. This is due to the lack of large sets of gold-standard labels that indicate when alarms are crucial for patient safety and optimal outcomes, and when alarms are unnecessary and should be suppressed^{5,22,25}. As a result, evaluation has typically focused on maximizing the number of alarms suppressed, with no consideration given to the appropriateness of this suppression.

This study aims to produce improved default vital sign alarm thresholds by extending the previous work in two important ways. Firstly, models are trained to find optimal default vital sign alarm thresholds, given data available at admission, on a patient-by-patient basis, rather than using patient groups defined by discrete age categories as is currently standard. Secondly, evaluation of the resulting patient-specific alarm thresholds found is conducted, by using non-default alarm thresholds as silver-standard personalized thresholds, and by using the presence of clinical events to indicate clinical concern. Heart rate alarms are used as a proof of concept in this manuscript, as heart rate threshold alarms are very common and have been shown to have a low specificity^{19,26,27}.

An important distinction of this study is the use of heart rate data for training the model, rather than using a set of labeled alarms. Although a large set of alarms are available, they are lacking gold-standard labels to indicate which were unnecessary and which were crucial for patient safety and optimal outcomes. As a result, using historical alarm data to learn optimal thresholds for each patient is not possible. Instead, we take a step back and aim to develop an alarm system from first principles. Since the goal of these vital sign alarms is to indicate when concerning vital sign measurements are seen, we aim to learn the 1st and 99th percentiles of HR seen during the first day of each patient’s stay. The use of the 1st and 99th

percentiles for these thresholds was chosen carefully. Other studies have used 5th and 95th percentiles for alarm thresholds²⁸ where default thresholds are chosen for large groups of patients. Due to inter-patient variability in vital signs, the conservative 5th and 95th percentiles are chosen to ensure that very extreme values in patients who have abnormally high or low vital signs are not missed. Since this study produces thresholds at a patient-specific level, this inter-patient variability does not need to be considered, and wider percentiles can be used to improve alarm specificity. Non-default alarm thresholds are used to evaluate the choice of 1st and 99th percentiles for use as personalized alarm thresholds.

2. Methods

2.1 Data

Two main sources of data were used for this study. The Philips Research Data Export (RDE) system at Stanford's Lucille Packard Children's Hospital (LPCH) has been recording vital sign waveforms for every patient that has had their vital signs monitored, both in intensive care units and on floor units, for the past several years. An extract from this system, containing 3.5 years worth of data (5 December 2012 – 20 April 2016) has been made available for research purposes. This extract contains once per minute average heart rate and respiratory rate, as well as records of any vital sign alarms that were triggered. These data have been combined with data from the electronic health record, obtained through the Stanford Translational Research Integrated Database Environment (STRIDE)²⁹. STRIDE contains patient demographics and clinical data including ICD9 codes and medication records. These datasets were linked using patient medical record numbers, or using data showing which patient was in a specific bed location at the time data is available. Figure 1 shows this initial mapping process, and Table 1 shows the characteristics of the final cohort.

For each patient represented in the RDE dataset, the first 24 hours of their stay was isolated and processed for use in this model. All data within this 24-hour period was considered, regardless of whether it was continuously recorded or included periods of missing vital sign data. Patients with less than one hour of data in the first 24-hours of monitoring were excluded from the analysis. Of the remaining patients, 83% of patients had data spanning at least one day, and the remaining patients had data over a mean of 15.3 hours. Four values were extracted for each patient: the mean, standard deviation, 1st percentile, and 99th percentile of the heart rate data available in this 24-hour period.

2.2 Outcome

There are two outcomes of interest for each patient, corresponding to the high and low alarm thresholds. The proposed ideal values for these are the 1st and 99th percentiles of the patient's observed heart rate over the first day of hospitalization. To allow for future extensions of this work, each patient's heart rate over the first day of hospitalization is modeled as a lognormal distribution parameterized by the mean and standard deviation of the heart rate, and the 1st and 99th percentiles are obtained from this lognormal model. Figure 3 shows that the 1st and 99th percentiles of the patient's heart rate are able to be accurately recovered using the mean and standard deviation of heart rate in a lognormal distribution. Two models are built, with outcomes of mean heart rate and standard deviation

of heart rate. The outputs of these models are then used as the parameters of patient-specific lognormal distributions, from which the expected 1st and 99th percentiles of the patient's heart rate are found. Evaluations are performed on these resulting 1st and 99th percentiles, as these are the proposed alarm thresholds.

2.3 Imputing weight

Including patient weight in the model was a prime consideration, as weight is known to impact heart rate. Weight data was not available for all patients, and for some patients weight at the time of the vital sign recording was not available. An imputation process was developed for weight data, using standard pediatric growth charts. Growth charts were used to find which percentile the patient fell into for their age at the time that weight was recorded. The growth charts were then used to determine the weight that the patient would be at the time of vital sign recording, assuming that they remained in the same percentile. If patients had multiple weights recorded, the mean percentile was used. 603 patients had no weight data recorded, so were assumed to be at the 50th percentile of weight for their age. The percentile found for each patient was also used as an input to the model.

2.4 Diagnosis Information

The STRIDE dataset includes diagnosis related groups (DRGs)³⁰, which are designed to group patients according to the medical services they receive, but can also be used to provide a rough grouping by clinical complaint. 45 DRGs were present as admit diagnoses in the cohort of interest. DRGs that contained less than 10 observations were combined into an 'OTHER' group, leaving a total of 22 distinct DRGs. This categorical variable was converted to 22 variables with Boolean values, with the constraint that for a given sample only one of the values can be set to 1.

The floor departments at LPCH are arranged such that patients with particular care needs are grouped together. For example, one floor unit typically houses patients with cardiac issues, while patients with pulmonary-related problems are cared for in another unit. The department in which the patient is located was used as a feature in our model, as it provides a rough grouping according to diagnosis.

2.5 Training

The combined RDE and STRIDE data set was randomly split into training and testing cohorts using a 75%/25% split at the patient level, resulting in cohorts of 6,383 and 2,124 patients.

As previously described, two models were trained: one to identify the mean heart rate, and the second to identify the standard deviation of the heart rate. The output of these models is used in a lognormal distribution to calculate the 1st and 99th percentiles of heart rate, which are proposed for use as the alarm thresholds. Two sets of these two models were trained. Figure 2 describes the training process. First, loess models³¹ were used to capture nonlinear variation in the mean and variance of heart rate with age. The thresholds calculated from the output of these models are referred to as 'personalized thresholds: age only' thresholds. The output from these models was used as inputs to two random forest models (one each for

mean heart rate and standard deviation), along with additional demographic (age, weight, gender, ethnicity and race) and diagnostic features (DRG and hospital department). A random forest model was chosen to avoid bias that would be introduced by a linear model.

2.6 Evaluations

The results can be evaluated directly by comparing the modeled 1st and 99th percentiles of the vital signs to the actual values. We include comparisons with the original LPCH vital sign thresholds, and age-based thresholds previously described in^{32,33}. A record of the alarms that sounded is available, so the number of alarms that would have been suppressed if the predicted 1st and 99th percentiles were used as thresholds can be found. However, evaluation of the clinical meaningfulness of these results is difficult, as gold-standard labels indicating whether alarms were meaningful are not available.

To estimate the appropriateness of the proposed alarm thresholds, we use the dataset of alarms that sounded in LPCH to find alarm thresholds that are not the default values, indicating that clinical staff manually chose this threshold for the patient. The non-default alarms from patients in both the training and the test set were able to be compared to the proposed thresholds without biasing the result of the evaluation, since the actually used thresholds that were used in practice were not input to the models. A total of 727 and 2,242 alarms with non-default settings were found for patients in the test set and the training set respectively.

As a second estimate of the appropriateness of the proposed alarm thresholds, we looked for significant clinical events in the 4 hours following an alarm. The label of ‘clinically meaningful alarms’ implies that to meet this criterion, some clinical action should have been taken in response to the alarm. Two lists of clinical events were formulated through consultation with clinicians and clinical experts, and are shown in Table 2. The presence of a clinical event from list A is considered to imply that the alarm was clinically meaningful, while an event from list B implies that the alarm was unnecessary. If events from both list A and list B occur in the 4-hour period following the alarm, this is considered ambiguous and no label is assigned to the alarm. A discharge event where the patient dies within 30 days is treated as a discharge to end of life care. 6.9% of all alarms recorded were assigned a label using this process (8.3% of low alarms and 6.1% of high alarms).

A record of the alarms that sounded is available, so the number of alarms that would have been suppressed if the predicted 1st and 99th percentiles were used as thresholds can be found. The status of the alarms if the new thresholds were used is compared to the labels created using the clinical events in Table 2 to obtain estimates of the sensitivity and specificity of the alarms. These values are not available to evaluate any other vital sign alarming method, so comparisons with previous methods are not possible. We are also unable to evaluate the performance of the original or age-based LPCH thresholds, as these were used to trigger the alarms.

3. Results

As shown in Figure 3, using the mean and standard deviation of a patient's heart rate over a 24-hour period as parameters in a lognormal distribution gives an accurate estimate of the 1st and 99th percentiles of the heart rate over this period. This shows that a lognormal model is well suited to the distribution of an individual patient's heart rate over a 24-hour period.

Figure 4 shows that a model with a single variable of continuous age is able to recover the 1st and 99th percentiles of heart rate more closely than the age-based thresholds previously developed at LPCH. Adding additional demographic and diagnostic features slightly decreases the variance in the error. Figure 4 also compares the vital sign thresholds to the thresholds of non-default alarms in the data set. The low error suggests that the use of 1st and 99th percentiles as threshold values is an appropriate one. The continuous age only model has a similar error to the age-based thresholds, while adding demographic and diagnostic features decreases this error.

Table 3 shows that over 50% of low heart rate alarms would be suppressed using the proposed thresholds, as well as upwards of 35% of high heart rate alarms, depending on the threshold scheme.

Using clinical events to infer labels for the alarms allowed us to estimate sensitivity and specificity of the proposed thresholds, shown in Table 4. Previous studies have shown that the specificity of heart rate alarms ranges from 1% to 36%⁵, suggesting that our proposed alarm thresholds would improve the specificity of heart rate alarms. Since it is not possible to measure false negative alarms, no studies have been conducted to determine the sensitivity of existing vital sign alarms, however this is generally considered to be extremely high, close to 100%.

4. Discussion

This study has shown that the 1st and 99th percentiles of observed heart rate over the first day of an inpatient stay are able to be predicted using a random forest with demographic and diagnostic features.

The comparison of the predicted 1st and 99th percentiles to the non-default alarm settings that were used in practice gives insight into the appropriateness of using the output from these models as alarm thresholds. Despite not being trained with the non-default alarm settings as inputs, the models recover these values well.

As shown in Table 4, the specificity of the proposed thresholds is higher than that of current heart rate alarms, but the sensitivity of the proposed thresholds is likely to be lower than the sensitivity of current alarms. In theory, this increases the chance that truly concerning heart rates will fail to sound an alarm, which could lead to negative outcomes for the patient. However, an alarm sounding is of no help to the patient in distress if it is not responded to, as may happen in a situation where clinical staff are suffering from alarm fatigue³⁴. While studies have shown that various forms of alarm fatigue can increase nurse response time^{8,9}, no studies have quantified the effective sensitivity of alarms given the presences of alarm

fatigue. We propose that the reduced number of alarms that will sound if these personalized thresholds are adopted (see Table 3) will reduce the problem of alarm fatigue, and that this reduction in the desensitization of health care providers will reduce the instance of negative patient outcomes related to missed vital sign events, despite the lower expected alarm sensitivity.

Limitations of this study include the lack of gold standard alarm labels to evaluate our proposed alarm thresholds. The evaluation methods used in lieu of gold standard labels (comparing to patient 1st and 99th percentiles, comparing to non-default alarm limits, and using clinical events to infer alarm labels) improve upon previous studies that have lacked any evaluation, but are still limited. For example only 7% of alarms could be labeled using clinical events. This also limits the accuracy of the performance metrics reported for the proposed alarm thresholds.

In conclusion, this study presents a model to accurately identify the 1st and 99th percentiles of an individual's heart rate during their first day of vital sign monitoring, using demographic and diagnosis features as input to a random forest. This is a proof of concept that personalized alarm thresholds can be learned, and demonstrates promising results for use of such personalized thresholds to reduce false alarms and address alarm fatigue. Patient-specific alarm thresholds represent a first step towards personalized medicine, and the resulting reduction in alarm fatigue will improve patient outcomes while also contributing to lower healthcare costs.

Acknowledgements

The authors would like to thank Eric Helfenbein, Principal Scientist at Philips Healthcare, for his help in obtaining the RDE dataset. We also thank Stephen Pfohl and Dr. Tavpritesh Sethi, both members of Stanford's Biomedical Informatics Research group, for their helpful feedback on the manuscript.

This work was supported by the Stanford Clinical and Translational Science Award (CTSA) to Spectrum (UL1 TR001085). The CTSA program is led by the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Force C Impact of clinical alarms on patient safety: a report from the American College of Clinical Engineering Healthcare Technology Foundation. *J Clin Eng* 2007.
2. Chopra V, McMahon LF Jr., Redesigning hospital alarms for patient safety: alarmed and potentially dangerous. *JAMA* 2014; 311(12): 1199–200. [PubMed: 24590296]
3. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. *AACN Adv Crit Care* 2013; 24(4): 378–86; quiz 87–8. [PubMed: 24153215]
4. Bridi AC, Louro TQ, da Silva RC. Clinical Alarms in intensive care: implications of alarm fatigue for the safety of patients. *Rev Lat Am Enfermagem* 2014; 22(6): 1034–40. [PubMed: 25591100]
5. Paine CW, Goel VV, Ely E, et al. Systematic Review of Physiologic Monitor Alarm Characteristics and Pragmatic Interventions to Reduce Alarm Frequency. *J Hosp Med* 2016; 11(2): 136–44. [PubMed: 26663904]
6. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med* 1994; 22(6): 981–5. [PubMed: 8205831]

7. Varpio L, Kuziemy C, MacDonald C, King WJ. The helpful or hindering effects of in-hospital patient monitor alarms on nurses: a qualitative analysis. *Comput Inform Nurs* 2012; 30(4): 210–7. [PubMed: 22156767]
8. Bonafide CP, Lin R, Zander M, et al. Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. *J Hosp Med* 2015; 10(6): 345–51. [PubMed: 25873486]
9. Bonafide CP, Zander M, Graham CS, et al. Video methods for evaluating physiologic monitor alarms and alarm responses. *Biomed Instrum Technol* 2014; 48(3): 220–30. [PubMed: 24847936]
10. Sentinel Event Alert Issue 50: Medical device alarm safety in hospitals.
11. Commission TJ. The Joint Commission Announces 2014 National Patient Safety Goal. 2014.
12. Commission TJ. 2015 Hospital National Patient Safety Goals. 2015.
13. Commission TJ. 2016 Hospital National Patient Safety Goals. 2016.
14. Dandoy CE, Davies SM, Flesch L, et al. A team-based approach to reducing cardiac monitor alarms. *Pediatrics* 2014; 134(6): e1686–94. [PubMed: 25384493]
15. Albert NM, Murray T, Bena JF, et al. Differences in alarm events between disposable and reusable electrocardiography lead wires. *Am J Crit Care* 2015; 24(1): 67–73; quiz 4. [PubMed: 25554556]
16. Cvach MM, Biggs M, Rothwell KJ, Charles-Hudson C. Daily electrode change and effect on cardiac monitor alarms: an evidence-based practice approach. *J Nurs Care Qual* 2013; 28(3): 265–71. [PubMed: 23187092]
17. Sendelbach S, Wahl S, Anthony A, Shotts P. Stop the Noise: A Quality Improvement Project to Decrease Electrocardiographic Nuisance Alarms. *Crit Care Nurse* 2015; 35(4): 15–22; quiz 1p following [PubMed: 26232798]
18. Cvach MM, Frank RJ, Doyle P, Stevens ZK. Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. *J Nurs Care Qual* 2014; 29(1): 9–18. [PubMed: 23963169]
19. Gorges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesth Analg* 2009; 108(5): 1546–52. [PubMed: 19372334]
20. Tsien C Reducing False Alarms in the Intensive Care Unit: A Systematic Comparison of Four Algorithms. 1997.
21. Hu X, Sapo M, Nenov V, et al. Predictive combinations of monitor alarms preceding in-hospital code blue events. *J Biomed Inform* 2012; 45(5): 913–21. [PubMed: 22465785]
22. Goel VV, Poole SF, Longhurst CA, et al. Safety analysis of proposed data-driven physiologic alarm parameters for hospitalized children. *J Hosp Med* 2016; 11(12): 817–23. [PubMed: 27411896]
23. Jacques SJ, Fauss EK, Sanders JA, et al. Patient-centered design of alarm limits in a complex pediatric population. *Health and Technology* 2016.
24. Fleming S, Thompson M, Stevens R, et al. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *Lancet* 2011; 377(9770): 1011–8. [PubMed: 21411136]
25. de Waele S, Nielsen L, Frassica J. Estimation of the patient monitor alarm rate for a quantitative analysis of new alarm settings. *Conf Proc IEEE Eng Med Biol Soc* 2014; 2014: 5727–30. [PubMed: 25571296]
26. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med* 1997; 25(4): 6149.
27. Schoenberg R, Sands DZ, Safran C. Making ICU alarms meaningful: a comparison of traditional vs. trend-based algorithms. *Proc AMIA Symp* 1999: 379–83. [PubMed: 10566385]
28. Karnik A, Bonafide CP. A framework for reducing alarm fatigue on pediatric inpatient units. *Hosp Pediatr* 2015; 5(3): 160–3. [PubMed: 25732990]
29. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE--An integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009; 2009: 391–5. [PubMed: 20351886]
30. Fetter RB, Shin Y, Freeman JL, Averill RF, Thompson JD. Case mix definition by diagnosis-related groups. *Med Care* 1980; 18(2 Suppl): iii, 1–53.

31. Cleveland WS. LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician* 1981; 35(1): 54.
32. Goel V, Poole S, Kipps A, et al. Implementation of Data Drive Heart Rate and Respiratory Rate parameters on a Pediatric Acute Care Unit. *Stud Health Technol Inform* 2015; 216: 918. [PubMed: 26262220]
33. Kipps AK, Poole SF, Slaney C, et al. Inpatient-Derived Vital Sign Parameters Implementation: An Initiative to Decrease Alarm Burden. *Pediatrics* 2017.
34. Bridi AC, da Silva RC, de Farias CC, Franco AS, dos Santos Vde L. [Reaction time of a health care team to monitoring alarms in the intensive care unit: implications for the safety of seriously ill patients]. *Rev Bras Ter Intensiva* 2014; 26(1): 28–35. [PubMed: 24770686]

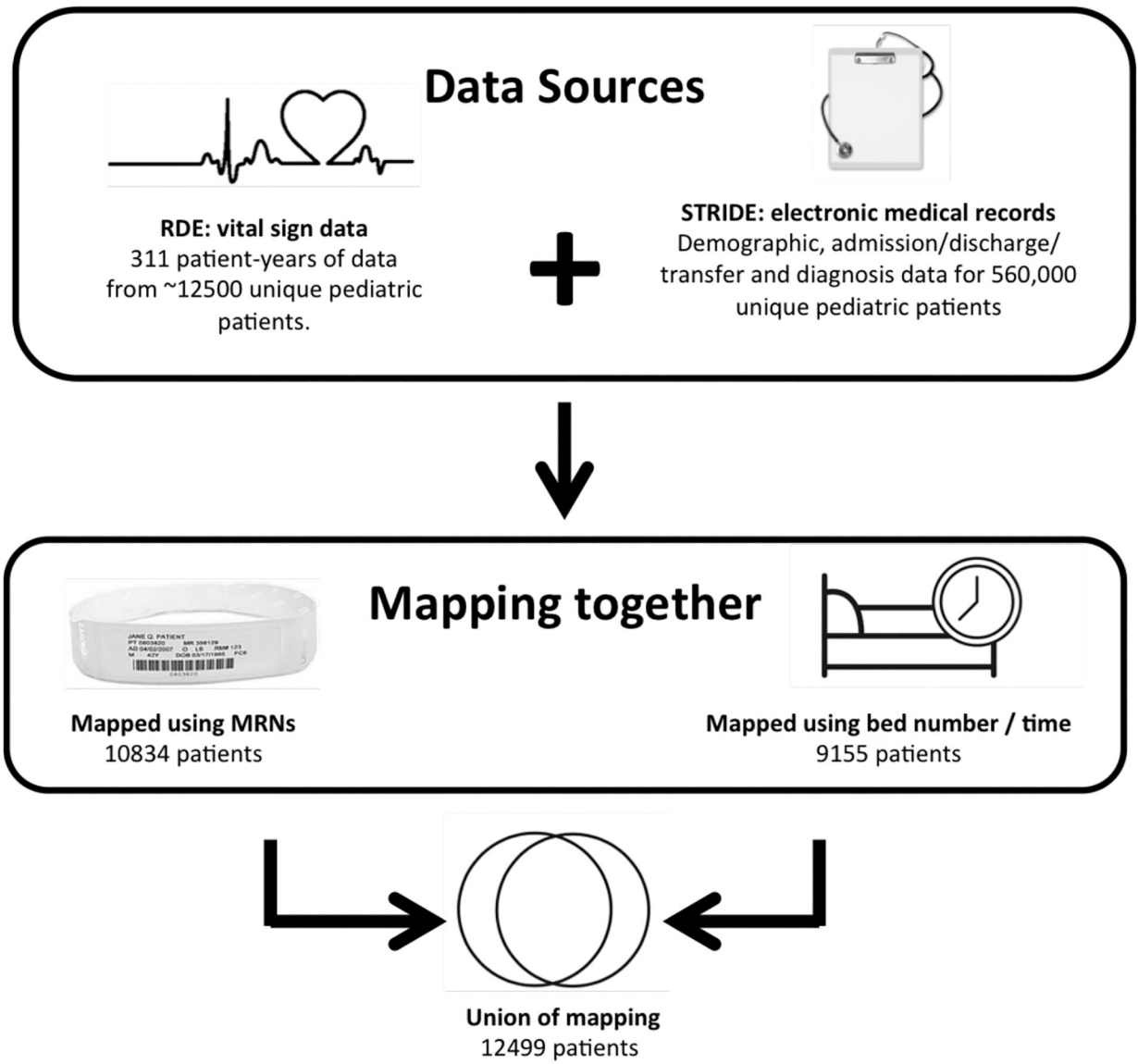


Figure1: Merging of RDE and STRIDE data using patient Medical Record Numbers (MRNs) where available, and using bed number and time where a unique patient was recorded as occupying the bed at the time of interest. Instances where the two mapping schemes gave different patients were removed.

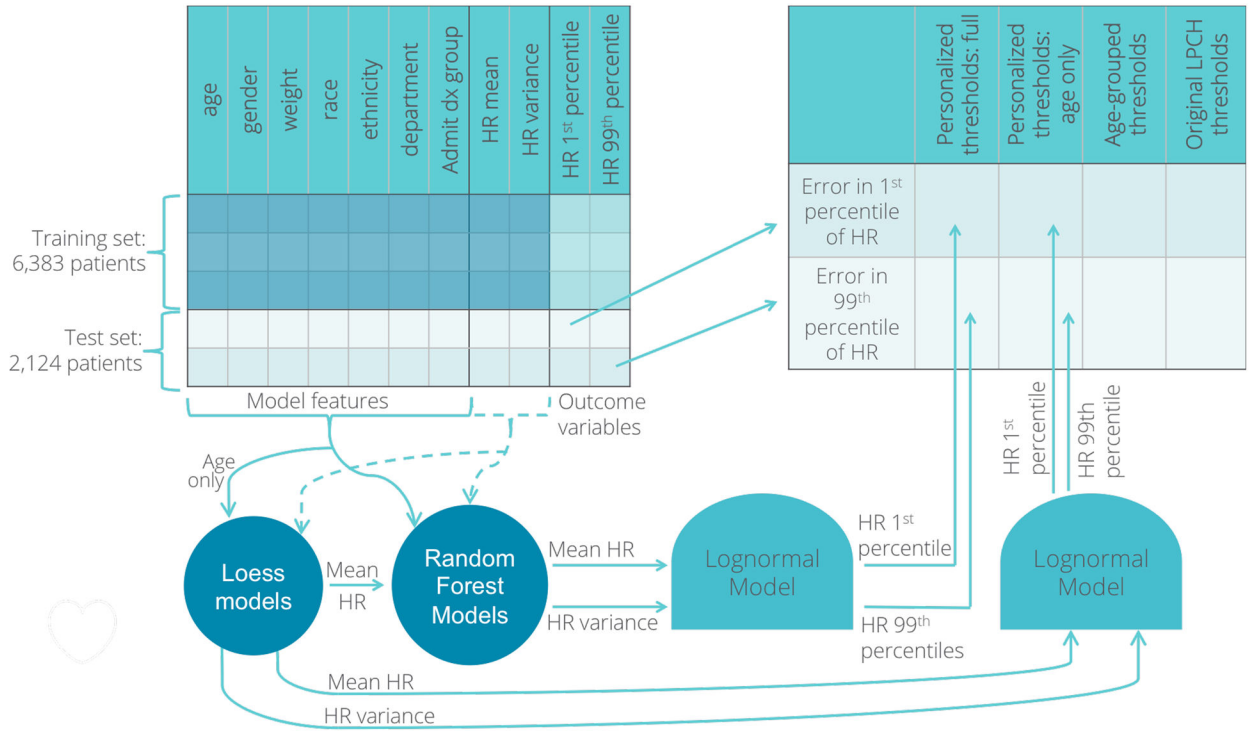


Figure 2:

Schematic of the model training process. First the training set is used to fit loess models to the outcomes of mean heart rate and heart rate variance, using age as the only feature. The outputs of these models are used as the parameters of a lognormal model to estimate the 1st and 99th percentiles of heart rate. These resulting estimates are proposed as personalized thresholds: age only. The output of the loess model fitting to mean heart rate is also used as a feature in a pair of random forest models, one fit to the outcome of mean heart rate, and the other fit to variance of heart rate. These random forest models also have gender, weight, race, ethnicity, hospital department, and admit diagnosis group (DRG) as additional features. The mean and variance of heart rate for each patient, as predicted by the random forest models, are used as the parameters of a lognormal model, allowing the 1st and 99th percentiles of heart rate to be estimated. The trained models are used to estimate the 1st and 99th percentiles of heart rate for patients in the test set, which can then be compared to the actual values observed over the first day of monitoring. The previously used original LPCH thresholds and age-grouped thresholds can also be compared to the observed 1st and 99th percentiles of heart rate.

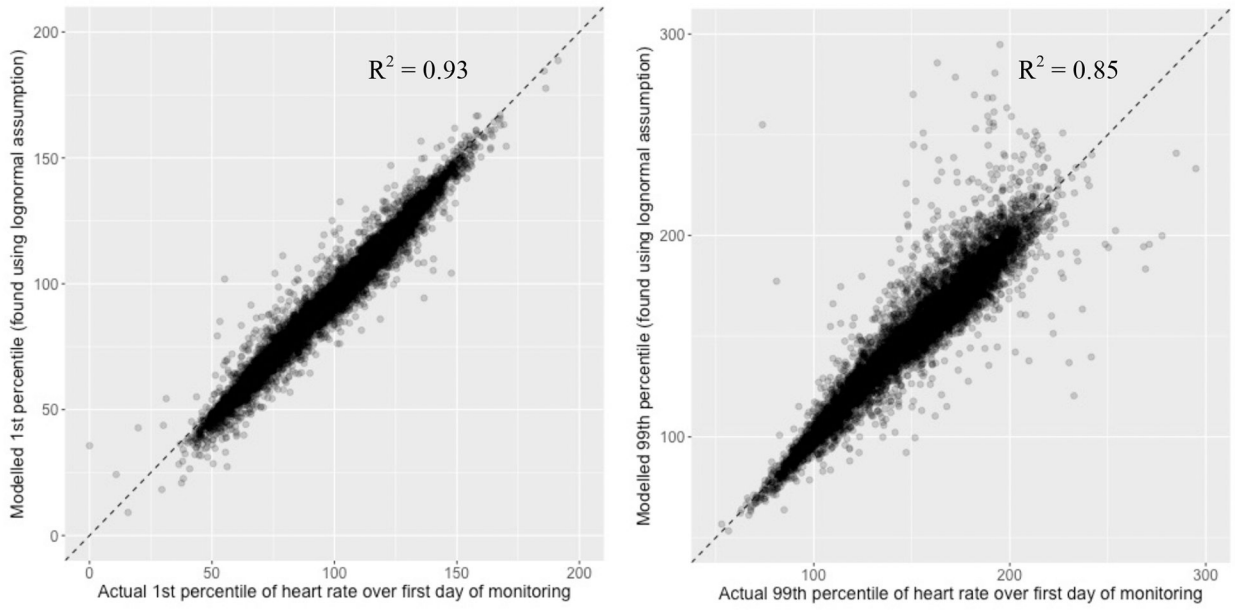


Figure 3:
 Error using mean and standard deviation of heart rate with lognormal assumption to find 1st (left) and 99th (right) percentile of heart rate.

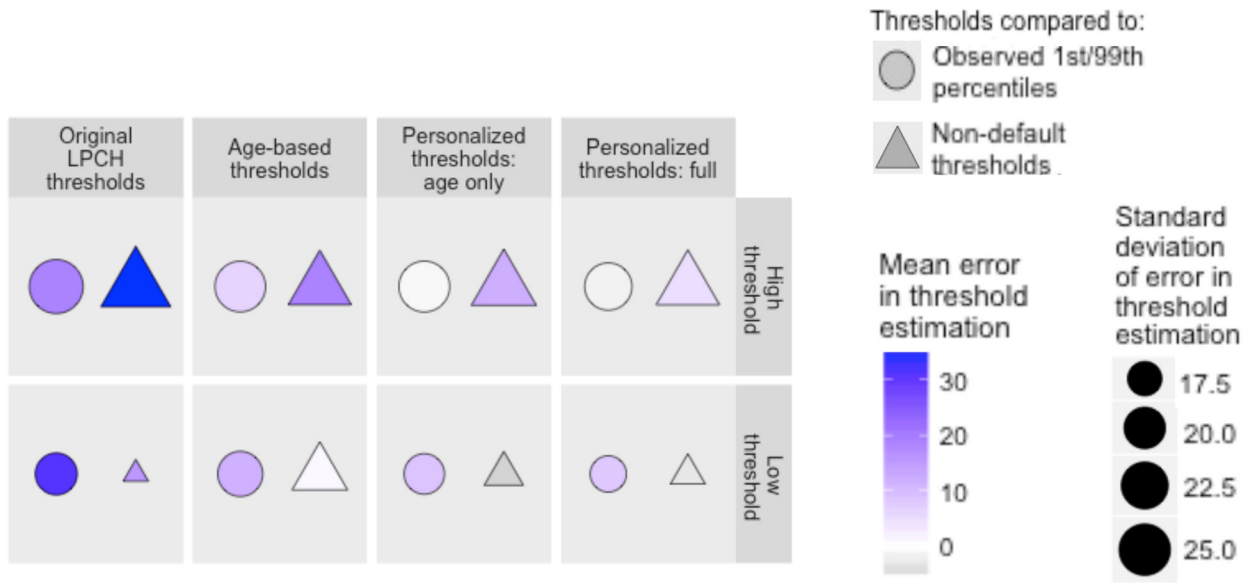


Figure 4: Comparison of alarm thresholds with the 1st (for low thresholds) and 99th (for high thresholds) percentiles of heart rate observed over the first 24 hours of monitoring (circles), and comparison of alarm thresholds with the recorded non-default thresholds (triangles).

Table 1:

Characteristics of patient cohort

	Count	Percentage
Total number of patients:	8,507	
Total HR alarms triggered:	1,930,493	
<i>Low</i>	693,516	35.69%
<i>High</i>	1,236,977	64.1%
Mean HR observations per patient:	14385 minutes (9.98 days)	
Standard deviation of HR observations per patient:	29942 minutes (20.8 days)	
Demographic breakdown:		
Gender:		
<i>Male</i>	3,921	46.1%
<i>Female</i>	4,586	53.9%
Ethnicity:		
<i>Hispanic</i>	2,121	24.9%
<i>Not Hispanic</i>	3,339	39.3%
<i>Unknown</i>	3,047	35.8%
Race:		
<i>Asian</i>	767	9.0%
<i>Black</i>	150	1.8%
<i>Native American</i>	6	0.1%
<i>Other</i>	1,937	22.8%
<i>Pacific Islander</i>	193	2.3%
<i>Unknown</i>	3,087	36.3%
<i>White</i>	2,367	27.8%
Age (years):		
<i>Min</i>	0.00	
<i>Median</i>	1.85	
<i>Mean</i>	4.96	
<i>Max</i>	17.98	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Clinical events used to indicate whether an alarm was clinically meaningful (list A) or unnecessary (list B).

List A (indicates clinically meaningful alarm)	List B (indicates unnecessary alarm)
Patient death	Patient discharged
Patient transferred to higher acuity unit	Patient transferred to lower acuity unit
Manual change of alarm thresholds to become more conservative	Manual change of alarm thresholds to become less conservative
Patient discharged to end of life care	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Percentage of alarms suppressed using proposed thresholds

	% low alarms suppressed	% high alarms suppressed
Personalized thresholds: age-only	53.1%	35.2%
Personalized thresholds: full	50.5%	44.1%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Performance metrics of proposed thresholds calculated using presence of clinical events to label alarms.

	Sensitivity	Specificity	Positive Predictive Value
Personalized thresholds: age-only	0.67	0.44	0.072
Personalized thresholds: full	0.62	0.49	0.079

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript