Safety Analysis of Proposed Data-Driven Physiologic Alarm Parameters for Hospitalized Children

Veena V. Goel, MD1,2*, Sarah F. Poole, BS3, Christopher A. Longhurst, MD, MS4,5, Terry S. Platchek, MD1,6, Natalie M. Pageler, MD, MEd2,7, Paul J. Sharek, MD, MPH1,8, Jonathan P. Palma, MD, MS2,9

1Division of Pediatric Hospital Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California; 2Department of Clinical Informatics, Stanford Children’s Health, Stanford, California; 3Biomedical Informatics Training Program, Stanford University, Stanford, California; 4Department of Pediatrics, University of California San Diego School of Medicine, San Diego, California; 5Department of Biomedical Informatics, University of California San Diego School of Medicine, San Diego, California; 6Performance Improvement Department, Stanford Children’s Health, Stanford, California; 7Division of Pediatric Critical Care, Department of Pediatrics, Stanford University School of Medicine, Stanford, California; 8Center for Quality and Clinical Effectiveness, Stanford Children’s Health, Stanford, California; 9Division of Neonatal and Developmental Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California.

INTRODUCTION: Modification of alarm limits is one approach to mitigating alarm fatigue. We aimed to create and validate heart rate (HR) and respiratory rate (RR) percentiles for hospitalized children, and analyze the safety of replacing current vital sign reference ranges with proposed data-driven, age-stratified 5th and 95th percentile values.

METHODS: In this retrospective cross-sectional study, nurse-charted HR and RR data from a training set of 7202 hospitalized children were used to develop percentile tables. We compared 5th and 95th percentile values with currently accepted reference ranges in a validation set of 2287 patients. We analyzed 148 rapid response team (RRT) and cardiorespiratory arrest (CRA) events over a 12-month period, using HR and RR values in the 12 hours prior to the event, to determine the proportion of patients with out-of-range vitals based upon reference versus data-driven limits.

RESULTS: There were 24,045 (55.6%) fewer out-of-range measurements using data-driven vital sign limits. Overall, 144/148 RRT and CRA patients had out-of-range HR or RR values preceding the event using current limits, and 138/148 were abnormal using data-driven limits. Chart review of RRT and CRA patients with abnormal HR and RR per current limits considered normal by data-driven limits revealed that clinical status change was identified by other vital sign abnormalities or clinical context.

CONCLUSIONS: A large proportion of vital signs in hospitalized children are outside presently used norms. Safety evaluation of data-driven limits suggests they are as safe as those currently used. Implementation of these parameters in physiologic monitors may mitigate alarm fatigue. Journal of Hospital Medicine 2016;11:817–823. © 2016 Society of Hospital Medicine

The management of alarms in the hospital setting is a significant patient safety issue. In 2013, the Joint Commission issued Sentinel Event Alert #50 to draw attention to the fact that tens of thousands of alarms occur daily throughout individual hospitals, and 85% to 99% are false or not clinically actionable. These alarms, designed to be a safety net in patient care, have the unintended consequence of causing provider desensitization, also known as “alarm fatigue,” which contributes to adverse events as severe as patient mortality. For this reason, a 2014 Joint Commission National Patient Safety Goal urged hospitals to prioritize alarm system safety and to develop policies and procedures to manage alarms and alarm fatigue.

Multiple efforts have been made to address alarm fatigue in hospitalized adults. Studies have quantified the frequency and types of medical device alarms, and some proposed solutions to decrease excess alarms. One such solution is to change alarm limit settings, an intervention shown to be efficacious in the literature. Although no adverse patient outcomes are reported in these studies, none of them included a formal safety evaluation to evaluate whether alarm rate reduction occurred at the expense of clinically significant alarms.

Specific to pediatrics, frameworks to address alarm fatigue have been proposed, and the relationship between nurse response time and frequency of exposure to nonactionable alarms has been reported. However, efforts to address alarm fatigue in the pediatric setting are less well studied overall, and there is little guidance regarding optimization of pediatric alarm parameters. Although multiple established reference ranges exist for pediatric vital signs, a systematic review in 2011 found that only 2 of 5 published heart rate (HR) and 6 respiratory rate (RR) guidelines cited any references, and even these had weak underpinning evidence. Consequently, ranges defining normal pediatric vital signs are derived either from small sample observational data in healthy outpatient children or consensus opinion. In a 2013 study by Bonafide et al., charted vital sign data from hospitalized children were used to develop percentile...
curves for HR and RR, and from these it was estimated that 54% of vital sign measurements in hospitalized children are out of range using currently accepted normal vital sign parameters.24 Although these calculated vital sign parameters were not implemented clinically, they called into question reference ranges that are currently widely accepted and used as parameters for electronic health record (EHR) alerts, early warning scoring systems, and physiologic monitor alarms.

With the goal of safely decreasing the number of out-of-range vital sign measurements that result from current, often non–evidence-based pediatric vital sign reference ranges, we used data from non–critically ill pediatric inpatients to derive HR and RR percentile charts for hospitalized children. In anticipation of local implementation of these data-driven vital sign ranges as physiologic monitor parameters, we performed a retrospective safety analysis by evaluating the effect of data-driven alarm limit modification on identification of cardiorespiratory arrests (CRA) and rapid response team (RRT) activations.

METHODS
We performed a cross-sectional study of children less than 18 years of age hospitalized on general medical and surgical units at Lucile Packard Children’s Hospital Stanford, a 311-bed quaternary-care academic hospital with a full complement of pediatric medical and surgical subspecialties and transplant programs. During the study period, the hospital used the Cerner EHR (Millennium; Cerner, Kansas City, MO) and Philips IntelliVue bedside monitors (Koninklijke Philips N.V., Amsterdam, the Netherlands). The Stanford University Institutional Review Board approved this study.

Establishing Data-Driven HR and RR Parameters
Vital sign documentation in the EHR at our institution is performed primarily by nurses and facilitated by bedside monitor biomedical device integration. We extracted vital signs data from the institution’s EHR for all general medical and surgical patients discharged between January 1, 2013 and May 3, 2014. To be most conservative in the definition of normal vital sign ranges for pediatric inpatients, we excluded critically ill children (those who spent any part of their hospitalization in an intensive care unit [ICU]). Physiologically implausible vital sign values were excluded as per the methods of Bonafide et al.24 The data were separated into 2 different sets: a training set (patients discharged between January 1, 2013 and December 31, 2013) and a test set for validation (patients discharged between January 1, 2014 and May 3, 2014). To avoid oversampling from both particular time periods and individual patients in the training set, we randomly selected 1 HR and RR pair from each 4-hour interval during a hospitalization, and then randomly sampled a maximum of 10 HR and RR pairs per patient. Using these vital sign measurements, we calculated age-stratified 1st, 5th, 10th, 50th, 90th, 95th, and 99th percentiles for both HR and RR.

Based on a combination of expert opinion and local consensus from our Medical Executive and Patient Safety Committees, we selected the 5th and 95th percentile values as proposed data-driven parameter limits and compared them to the 5th and 95th percentile values generated in the 2013 study24 and to the 2004 National Institutes of Health (NIH)–adapted vital sign reference ranges currently used at our hospital.25 Using 1 randomly selected HR and RR pair from every 4-hour interval in the validation set, we compared the proportion of out-of-range HR and RR observations with the proposed 5th and 95th percentile data-driven parameters versus the current NIH reference ranges. We also calculated average differences between our data-driven 5th and 95th percentile values and the calculated HR and RR values in the 2013 study.24

Safety Analysis
To assess the safety of the newly created 5th and 95th percentile HR and RR parameters prior to clinical adoption, we retrospectively reviewed data associated with all RRT and CRA events on the hospital’s medical/surgical units from March 4, 2013 until March 3, 2014. The RRT/CRA event data were obtained from logs kept by the hospital’s code committee. We excluded events that lacked a documented patient identifier, occurred in locations other than the acute medical/surgical units, or occurred in patients >18 years old. The resulting charts were manually reviewed to determine the date and time of RRT or CRA event activation. Because evidence exists that hospitalized pediatric patients with CRA show signs of vital sign decompensation as early as 12 hours prior to the event,26–29 we extracted all EHR-charted HR and RR data in the 12 hours preceding RRT and CRA events from the institution’s clinical data warehouse for analysis, excluding patients without charted vital sign data in this time period. The sets of patients with any out-of-range HR or RR measurements in the 12-hours prior to an event were compared according to the current NIH reference ranges25 versus data-driven parameters. Additionally, manual chart review was performed to assess the reason for code or RRT activation, and to determine the role that out-of-range vital signs played in alerting clinical staff of patient decompensation.

Statistical Analysis
All analysis was performed using R statistical package software (version 0.98.1062 for Mac OS X 10_9_3; The R Foundation for Statistical Computing, Vienna, Austria) with an SQL database (MySQL 2015; Oracle Corp., Redwood City, CA).

RESULTS

Data-Driven HR and RR Parameters

We established a training set of 62,508 vital sign measurements for 7202 unique patients to calculate 1st, 5th, 10th, 50th, 90th, 95th, and 99th percentiles for HR and RR among the 14 age groups (see Supporting Information, Appendix 1, in the online version of this article). Figures 1 and 2 compare the proposed

FIG. 1. Comparison of HR ranges. Data-driven HR 5th/95th percentile ranges compared with Bonafide et al.’s data-driven HR 5th/95th percentiles24 and with the currently adopted NIH 2004 reference ranges.25 Abbreviations: bpm, beats per minute; HR, heart rate; NIH, National Institutes of Health.

FIG. 2. Comparison of RR Ranges. Data-driven RR 5th/95th percentile ranges compared with Bonafide et al.’s data-driven RR 5th/95th percentiles24 and with the currently adopted NIH 2004 reference ranges.25 Abbreviations: bpm, breaths per minute; NIH, National Institutes of Health; RR, respiratory rate.
data-driven vital sign ranges with (1) our current HR and RR reference ranges and (2) the 5th and 95th percentile values created in the similar 2013 study. The greatest difference between our study and the 2013 study was across data-driven 95th percentile RR parameters, which were an average of 4.8 points lower in our study.

Our validation set consisted of 82,993 vital sign measurements for 2287 unique patients. Application of data-driven HR and RR 5th and 95th percentile limits resulted in 24,045 (55.6%) fewer out-of-range measurements compared to current NIH reference ranges (19,240 vs 43,285). Forty-five percent fewer HR values and 61% fewer RR values were considered out of range using the proposed data-driven parameters (see Supporting Information, Appendix 2, in the online version of this article).

Safety
Of the 218 unique out-of-ICU RRT and CRA events logged from March 4, 2013 to March 3, 2014, 63 patients were excluded from analysis: 10 lacked identifying information, 33 occurred outside of medical/surgical units, and 20 occurred in patients >18 years of age. The remaining 155 patient charts were reviewed. Seven patients were subsequently excluded because they lacked EHR-documented vital signs data in the 12 hours prior to RRT or CRA team activation, yielding a cohort of 148 patients (128 RRT events, 20 CRA events).

Table 1 describes the analysis of vital signs in the 12 hours leading up to the 148 RRT and CRA events.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No. Patients With HR Out of Range</th>
<th>No. Patients With RR Out of Range</th>
<th>No. Patients With HR or RR Out of Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH ranges</td>
<td>121</td>
<td>137</td>
<td>144</td>
</tr>
<tr>
<td>Data-driven ranges</td>
<td>129</td>
<td>104</td>
<td>138</td>
</tr>
<tr>
<td>Difference (causal threshold)</td>
<td>+8 (low HR)</td>
<td>-20 (high RR), -5 (low RR)</td>
<td>+2 (low HR), -8 (high RR)</td>
</tr>
</tbody>
</table>

**NOTE:** Abbreviations: CRA, cardiorespiratory arrest; HR, heart rate; NIH, National Institutes of Health; RR, respiratory rate; RRT, rapid response team. *Vital signs in the 12 hours preceding RRT or CRA event.

All 121 patients with out-of-range HR values using NIH reference ranges also had out-of-range HR values with the proposed data-driven parameters; an additional 8 patients had low HR values using the data-driven parameters. Of the 137 patients with an out-of-range RR value using NIH reference ranges, 33 (24.1%) were not considered out of range by the data-driven parameters. Of these, 28 had high RR and 5 had low RR according to NIH reference ranges.

After evaluating out-of-range HR and RR individually, the 148 RRT and CRA events were analyzed for either out-of-range HR values or RR values. In doing so, 144 (97.3%) patients had either HR or RR measurements that were considered out of range using our current NIH reference ranges. One hundred thirty-eight (93.2%) had either HR or RR measurements that were considered out of range with the proposed parameters. One hundred thirty-six (94.4%) of the 144 patients with out-of-range HR or RR measurements according to NIH reference ranges were also considered out of range using proposed parameters. The data-driven parameters identified 2 additional patients with low HR who did not have out-of-range HR or RR values using the current NIH reference ranges. Manual chart review of the RRT/CRA events in the 8 patients who had normal HR or RR using the data-driven parameters revealed that RRT or CRA team interventions occurred for clinical indications that did not rely upon HR or RR measurement (eg, laboratory testing abnormalities, desaturation events) (Table 2).

**DISCUSSION**
This is the first published study to analyze the safety of implementing data-driven HR and RR parameters in hospitalized children. Based on retrospective analysis of a 12-month cohort of patients requiring RRT or CRA team activation, our data-driven HR and RR parameters were at least as safe as the NIH-published reference ranges employed at our children’s hospital. In addition to maintaining sensitivity to RRT and CRA events, the data-driven parameters resulted in an estimated 55.6% fewer out-of-range measurements among medical/surgical pediatric inpatients.

Improper alarm settings are 1 of 4 major contributing factors to reported alarm-related events, and data-driven HR and RR parameters provide a means by which to address the Joint Commission Sentinel...
Event Alert and National Patient Safety Goal regarding alarm management safety for hospitalized pediatric patients. Our results suggest that this evidence-based approach may reduce the frequency of false alarms (thereby mitigating alarm fatigue), and should be studied prospectively for implementation in the clinical setting.

The selection of percentile values to define the new data-driven parameter ranges involved various considerations. In an effort to minimize alarm fatigue, we considered using the 1st and 99th percentile values. However, our Medical Executive and Patient Safety Committees determined that the 99th percentile values for HR and RR for many of the age groups exceeded those that would raise clinical concern. A more conservative approach, applying the 5th and 95th percentile values, was deemed clinically appropriate and consistent with recommendations from the only other study to calculate data-driven HR and RR parameters for hospitalized children.

When taken in total, Bonafide et al.’s 2013 study demonstrated that up to 54% of vital sign values were abnormal according to textbook reference ranges. Similarly, we estimated 55.6% fewer out-of-range HR and RR measurements with our data-driven parameters. Although our 5th and 95th HR percentile and 5th percentile RR values are strikingly similar to those developed in the 2013 study, the difference in 95th percentile RR values between the studies was potentially clinically significant, with our data-driven upper RR values being 4.8 breaths per minute lower (more conservative) on average. Bonafide et al. transformed the RR values to fit a normal distribution, which might account for this difference. Ultimately, our safety analysis demonstrated that 24% fewer patients were considered out of range for high RR prior to RRT/CRA events with the data-driven parameters compared to NIH norms. Even fewer RRT/CRA patients would have been considered out of range per Bonafide’s less conservative 95% RR limits.

Importantly, all 8 patients in our safety analysis without abnormal vital sign measurements in the 12 hours preceding their clinical events according to the proposed data-driven parameters (but identified as having high RR per current reference ranges) had RRT or CRA events triggered due to other significant clinical manifestations or vital sign abnormalities (eg, hypoxia). This finding is supported by the literature, which suggests that RRTs are rarely activated due to single vital sign abnormality alone. Prior analysis of RRT activations in our pediatric hospital demonstrated that only approximately 10% of RRTs were activated primarily on the basis of HR or RR vital sign abnormalities (5.6% tachycardia, 2.8% tachypnea, 1.4% bradycardia), whereas 36% were activated due to respiratory distress. The clinical relevance of high RR in isolation is questionable given a recent pediatric study that raised all RR limits and decreased alarm frequency without adverse patient safety consequences. Our results suggest that modifying HR and RR alarm parameters using data-driven 5th and 95th percentile limits to decrease alarm frequency does not pose additional safety risk related to identification of RRT and CRA events. We encourage continued work toward development of multivariate or “smart” alarms that analyze multiple simultaneous vital sign measurements and trends to determine whether an alarm should be triggered.

The ability to demonstrate the safety of data-driven HR and RR parameters is a precursor to hospital-wide implementation. We believe it is crucial to perform a safety analysis prior to implementation due to the role vital signs play in clinical assessment and detection of patient deterioration. Though a few studies have shown that modification of alarm parameters decreases alarm frequency, to our knowledge no formal safety evaluations have ever been published. This study provides the first published safety evaluation of data-driven HR and RR parameters.

By decreasing the quantity of out-of-range vital sign values while preserving the ability to detect patient deterioration, data-driven vital sign alarm limits have the potential to decrease false monitor alarms, alarm-generated noise, and alarm fatigue. Future work includes prospectively studying the impact of adoption of data-driven vital sign parameters on monitor alarm burden and monitoring the safety of the changes. Additional safety analysis could include comparing the sensitivity and specificity of early warning score systems when data-driven vital sign ranges are substituted for traditional physiologic parameters. Further personalization of vital sign parameters will involve incorporating patient-specific characteristics (eg, demographics, diagnoses) into the data-driven approach, and could revolutionize the way hospitalized patients are monitored.

Numerous relevant issues are not yet addressed in this initial, single-institution study. First, although the biomedical device integration facilitated the direct import of monitor data into the EHR (decreasing transcription errors), our analysis was performed using EHR-charted data. As such, the effect on bedside monitor alarms was not directly evaluated in our study, including those due to technical alarms or patient artifact. Second, our overall sample size for the training set was quite large; however, in some cases the number of patients per age category was limited. Third, although we evaluated the identification of severe deterioration leading to RRT or CRA events, the sensitivity of the new limits to the need for other interventions (eg, fluid bolus for dehydration or
escalation of respiratory support for asthma exacerbation) or unplanned transfers to the ICU was not assessed. Fourth, the analysis was retrospective, and so the impact of data-driven alarm limits on length of stay and readmission could not be defined. Fifth, excluding all vital sign measurements from patients who spent any time in the ICU setting decreased the amount of data available for analysis. However, excluding sicker patients probably resulted in narrower data-driven HR and RR ranges, leading to more conservative proposed parameters that are more likely to identify patient decompensation in our safety analysis. Finally, this was a single-site study. We believe our data-driven limits are applicable to other tertiary or quaternary care facilities given the similarity to those generated in a study performed in a comparable setting,24 but generalizability to other settings may be limited if the local population is sufficiently different. Furthermore, because institutional policies (eg, indications for care escalation) differ, individual institutions should determine whether our analysis is applicable to their setting or if local safety evaluation is necessary.

CONCLUSION

A large proportion of HR and RR values for hospitalized children at our institution are out of range according to current vital sign reference ranges. Our new data-driven alarm parameters for hospitalized children provide a potentially safe means by which to modify physiologic bedside monitor alarm limits, a first step toward customization of alarm limit settings in an effort to mitigate alarm fatigue.

Acknowledgements

The authors thank Debby Huang and Joshua Glandorf in the Information Services Department at Stanford Children’s Health for assistance with data acquisition. No compensation was received for their contributions.

Disclosures: All authors gave approval of the final manuscript version submitted for publication and agreed to be accountable for all aspects of the work. Dr. Venna V. Goel conceptualized and designed the study; collected, managed, analyzed and interpreted the data; prepared and reviewed the initial manuscript; and approved the final manuscript as submitted. Ms. Sarah F. Poole contributed to the design of the study and performed the primary data analysis for the study. Ms. Poole critically revised the manuscript for important intellectual content and approved the final manuscript as submitted. Dr. Goel and Ms. Poole had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Dr. Paul J. Sharek and Dr. Jonathan P. Palma contributed to the study design and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. J Adv Nurs. 2009;65(9):1844–1852.


