KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Joshi R., van Pul C., Atallah L., Feijs L., Van Huffel S., Andriessen P., ``Pattern discovery in critical alarms originating from neonates under intensive care'', *Physiological Measurement*, vol. 37, 2016, pp. 564-579.
Archived version	Final publisher's version / pdf
Published version	http://dx.doi.org/10.1088/0967-3334/37/4/564
Journal homepage	http://iopscience.iop.org/journal/0967-3334;jsessionid=8123A22CF7A9747CAFB652EDAD3DF702.c3.
Author contact	Sabine.vanhuffel@esat.kuleuven.be Klik hier als u tekst wilt invoeren.
IR	https://lirias.kuleuven.be/handle/123456789/537622

(article begins on next page)

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **37** (2016) 564–579 doi:10.1088/0967-3334/37/4/564

Pattern discovery in critical alarms originating from neonates under intensive care

Rohan Joshi[1,2], Carola van Pul[2,3], Louis Atallah[4], Loe Feijs[1], Sabine Van Huffel[5] and Peter Andriessen[6,7]

[1] Eindhoven University of Technology, Department of Industrial Design, Laplace 32, 5612 AZ Eindhoven, The Netherlands
[2] Máxima Medical Center, Clinical Physics, Veldhoven, The Netherlands
[3] Eindhoven University of Technology, School of Medical Physics and Engineering, Eindhoven, The Netherlands
[4] Patient Care & Measurements Department, Philips Research, Eindhoven, The Netherlands
[5] KU Leuven, Department of Electrical Engineering (ESAT), Division Stadius, Leuven, Belgium and iMinds Medical IT, Leuven, Belgium
[6] Máxima Medical Center, Neonatology, Veldhoven, The Netherlands
[7] Faculty of Health, Medicine and Life Science, Maastricht University, The Netherlands

E-mail: r.joshi@tue.nl

Received 3 November 2015, revised 28 January 2016
Accepted for publication 2 February 2016
Published 30 March 2016



Abstract

Patient monitoring generates a large number of alarms, the vast majority of which are false. Excessive non-actionable medical alarms lead to alarm fatigue, a well-recognized patient safety issue. While multiple approaches to reduce alarm fatigue have been explored, patterns in alarming and inter-alarm relationships, as they manifest in the clinical workspace, are largely a black-box and hamper research efforts towards reducing alarms. The aim of this study is to detect opportunities to safely reduce alarm pressure, by developing techniques to identify, capture and visualize patterns in alarms.

Nearly 500 000 critical medical alarms were acquired from a neonatal intensive care unit over a 20 month period. Heuristic techniques were developed to extract the inter-alarm relationships. These included identifying the presence of alarm clusters, patterns of transition from one alarm category to another, temporal associations amongst alarms and determination of prevalent sequences in which alarms manifest.

Desaturation, bradycardia and apnea constituted 86% of all alarms and demonstrated distinctive periodic increases in the number of alarms that were

 © 2016 Institute of Physics and Engineering in Medicine Printed in the UK

synchronized with nursing care and enteral feeding. By inhibiting further alarms of a category for a short duration of time (30 s/60 s), non-actionable physiological alarms could be reduced by 20%. The patterns of transition from one alarm category to another and the time duration between such transitions revealed the presence of close temporal associations and multiparametric derangement. Examination of the prevalent alarm sequences reveals that while many sequences comprised of multiple alarms, nearly 65% of the sequences were isolated instances of alarms and are potentially irreducible.

Patterns in alarming, as they manifest in the clinical workspace were identified and visualized. This information can be exploited to investigate strategies for reducing alarms.

Keywords: alarm fatigue, neonates, pattern discovery, NICU, monitoring, critical alarms, alarm reduction

S Online supplementary data available from stacks.iop.org/PM/37/564/mmedia

(Some figures may appear in colour only in the online journal)

Introduction

Medical alarms originating from continuous patient monitoring are intended to draw the attention of caregivers to the status of patients at critical times. Reliable alarming enables caregivers to pursue other important tasks, safe in the knowledge that patients are stable. An overload of alarms, the vast majority of which are false or clinically irrelevant leads to alarm fatigue—a situation that compromises clinical workflow and patient safety (Borowski *et al* 2011a, Cvach 2012, Sendelbach 2012).

Alarm fatigue is a serious patient-safety concern across diverse medical settings such as intensive care units, operating rooms, emergency rooms, progressive care units, nursing homes and even dental offices (Chambrin 2001, Atzema *et al* 2006, Siebig *et al* 2010, Schmid *et al* 2011, Cvach *et al* 2013, Weaver 2013). Published concerns about excessive and inappropriate sounding of alarms dates back to at least three decades (O'Carroll 1986). With the increasing use of monitoring and the growing number of physiological parameters being monitored, this problem is ever-increasing. In recent years, multiple organizations such as The Joint Commission, the ECRI Institute, the Healthcare Technology Foundation, the Association for the Advancement of Medical Instrumentation and the American Association of Critical-Care Nurses have recognized the hazards originating from alarm fatigue (Graham and Cvach 2010, Keller 2012, Mitka 2013, Drew *et al* 2014, Funk *et al* 2014).

Multiple factors influence medical alarms and therefore various approaches with the potential to minimize nuisance alarms have been investigated. On the one hand, these include organizational and institutional changes, while on the other, technical approaches have been explored.

Some of the organizational and institutional reforms that have helped reduce alarm fatigue include customization of alarm thresholds to patient characteristics, daily electrode change for electrocardiogram (ECG), employing disposable ECG electrodes, judging the appropriateness of continuous physiological monitoring for different patient populations, selecting devices with low false-alarm frequencies and rapid communication of alarms to healthcare workers by employing pagers, wireless phones etc (Ahlborn *et al* 2000, Atzema *et al* 2006, Graham and Cvach 2010, Bonzheim *et al* 2011, Cvach *et al* 2013, Albert *et al* 2015, van Pul *et al* 2015).

From a technical point of view, both improved signal processing and more sophisticated classification techniques (using both single and multiple sources of physiological signals) have shown promise in reducing nuisance alarms (Schoenberg *et al* 1999, Imhoff *et al* 2006, Imhoff *et al* 2009, Borowski *et al* 2011b, Baumgartner *et al* 2012). Approaching this problem from a different perspective, a study by Görges *et al* that used approximately 1200 alarms from an intensive care unit, showed that delaying the onset of alarms can eliminate many ineffective ones. This indicates that many alarm states resolve quickly and without intervention (Görges *et al* 2009). Up till now the focus has been on minimizing false and non-actionable alarms which are, in fact, detrimental to the interests of the patients (Bliss *et al* 2000, Bustamante *et al* 2007, Blum and Tremper 2010, Konkani *et al* 2012).

Given the fact that alarms are merely treated as acute alerts, the longer-term patterns in alarming remain largely a black-box, especially given the absence of quantitative techniques to analyze such data and to visualize and synthesize this information. This fact is further highlighted by a recent editorial in *Acta Paediatrica* by Walsh *et al*, which states that often, multiple parameters derange together in neonates and that enhanced monitoring systems that can recognize such time-linked events could correctly identify them as a single one—and thus minimize alarm fatigue (Walsh *et al* 2015). In addition, we recognize the need for quantifying such relationships amongst alarms, identifying the prevalent phenotypic alarm data (alarm sequences), developing diagnostic alarms and predictive warnings (Blum and Tremper 2010, Borowski *et al* 2011a, McCartney 2012). Exploring alarm patterns as well as the temporal relations between them can pave the path to providing diagnostic alarms and reducing alarm fatigue by recognizing multiple alarms that are closely clustered in time. This can then lead to a reduced number of alarms by combining a cluster of alarms into a single super-alarm.

While alarm fatigue is a problem across diverse medical settings, the neonatal patient population and in particular premature neonates are especially vulnerable to the excessive sounding of false alarms. Loud transient noise such as those from alarm sounds have negative short-term effects on the cardiovascular and respiratory systems of preterm infants, cause disruption of sleep and are also believed to impair neurodevelopment (Wachman and Lahav 2011, McMahon *et al* 2012, Altimier and Phillips 2013). Therefore, the need to reduce alarms in the neonatal intensive care unit (NICU) until only actionable alarms remain is especially important.

The aim of this work was to retrospectively explore and quantify patterns in critical alarms that originated from neonates under intensive care. We hypothesize, based on insights from our earlier work, that interrelationships between alarms and patterns in alarming exist and can be exploited to safely reduce the number of alarms (Pul *et al* 2015). In order to examine this possibility, we developed exploratory techniques to identify and visualize such relationships for large volumes of alarm data. Specifically, we (i) quantitatively analyzed the temporal relationship of alarms of the same category, (ii) identified the transition patterns and the times to transition between select alarm categories and, (iii) identified and analyzed the prevalent alarm sequences, for example an ABD event, which is an apnea (A) followed by a bradycardia (B) and a desaturation (D). We then propose a set of strategies for using these relationships for reducing redundant alarms.

Materials and methods

The NICU environment

Máxima Medical Centre has an 18-bed, level III, tertiary NICU with an annual admission rate of approximately 380 babies. This NICU has nine single rooms, five double rooms and a triple room, where a maximum of 18 beds can be simultaneously used. Due to the private

nature of the patient rooms an alarm distribution system was implemented, after risk analysis, details of which are published elsewhere (Pul *et al* 2015). A brief summary of this system is provided here.

The primary alarm chain comprised of the patient monitor (Philips IntelliVue MX 800, Germany), the inter-bed communication system and the Philips central monitors at the nursing stations. On average, nurses were responsible for no more than two babies. The patient-monitor alarms that were sent to the central nursing stations and to the other patient monitors ('inter-bed communication') could also be sent to handheld devices (i62, Ascom, Sweden) carried by the nurses. This was achieved by sending the monitor alarms to an Emergin server (Philips Medical Systems, USA), following which they were routed to a distributed alarming system (Ascom, Sweden) from which they were wirelessly transferred to the handheld devices.

Patient monitoring was used to continuously acquire the ECG, respiration rate (using impedance pneumography), oxygen saturation (SpO_2), invasive arterial blood pressure (BP) (when available) and temperature of all babies. When clinically necessary, additional devices such as those measuring transcutaneous oxygen and carbon dioxide were interfaced to the patient monitors via the Philips IntelliBridge EC10. This not only allowed the signals recorded from these devices to be displayed on the patient monitors but also for the alarms originating from these devices to be transferred to the alarm chain.

Based on urgency, patient monitor alarms were categorized into 'red' and 'yellow' alarms. The physiological thresholds for generating these alarms, taking into account the gestational ages of the babies, were decided by a multidisciplinary team and are tabulated in our earlier work (Pul *et al* 2015). The red alarms corresponded to potentially life-threatening situations, whereas yellow alarms served as warnings or alerts. Both yellow and red alarms were sent to the central stations and the inter-bed communication system (patient monitors in other rooms). Red alarms, however, were also transferred to the handheld devices carried by the nurses. The handheld devices beeped thrice for every red alarm and then stayed silent for 45 s for alarms of the same category. If the alarm state persisted, a repeat alarm was generated. For alarms of a different category, the handheld devices would beep again. For red alarms, if the nurse did not silence the alarm using the patient monitor or if the alarming situation did not resolve by itself within 45 s, a repeat alarm was sent to both, the primary and the buddy nurse (an additional nurse who cares for the baby if the primary nurse is preoccupied).

Alarms originating from devices interfaced to the EC10 generated yellow alarms only. It is worth noting that ventilators were not part of the alarm chain. Infusion pumps were connected to the alarm chain via the device call system (DCS) while the parents/guardians of the babies could call the nurse by means of a patient call system (PCS). For red alarms generated by the patient monitors, the alarm message displayed on the handheld devices of nurses included the type of alarm (e.g. bradycardia) along with the vital parameters of the patients, while for DCS and PCS only the message 'alarm' was displayed.

Patient demographics

A total of 8522 patient days, with at least one physiological alarm, were present in the period from 1 May 2012 to 31 December 2013. Patient demographics of the 593 subjects that were analyzed are presented in table 1. There were no exclusion criteria. As is standard practice, this retrospective analysis, aimed at quality improvement, did not require ethical approval. We analyzed this large amount of data on *anonymized alarms* without applying constraints on patient metadata like GA, PMA, etc—i.e. without stratifying them on the basis of patient metadata.

Table 1. Characteristics of the patient population.

Feature	Median	25th percentile	75th percentile
Gestational age (week)	31.3	28.9	34.6
Birthweight (g)	1540	1121	2174
Length of stay (day)	7	4	16
PMA at discharge (week)	33.4	32.0	37

Alarm demographics

The Emergin server stored all red alarm logs from the patient monitors. These logs contained the timestamp, bed number, alarm category, nature of alarm (first or repeat alarm) and the threshold and breach values for heart rate (HR), SpO_2 and BP. For our analysis, the alarm logs of the 20 month study period were processed and analyzed using Matlab (MathWorks, USA).

All unique categories of red alarms were identified. These alarm categories corresponded to either physiological monitor alarms or to technical ones. There was a small subset of alarms, from periods of testing which were excluded from all analysis. The total number of alarms, the number of repeat alarms, the different categories of alarms and the alarm rate for each category of alarm were obtained. Statistics of alarm occurrence per patient per day were extracted in order to explore the variance in daily alarm pressure generated by different patients.

If alarms occur at random, analysis of a large number of alarms would yield a constant alarm pressure with respect to time. If however, alarming is not an entirely stochastic process and is influenced by dynamic environmental or clinical variables, this would be reflected in a non-uniform alarm pressure with respect to time. In order to identify the presence of patterns, the average number of alarms generated by 100 patients for each consecutive 10 min interval of time was plotted through the course of the day for physiological, technical and all alarms combined. Differences in patterns of alarming between the weekdays and weekends were also investigated.

Technical analysis

In order to investigate the relationships amongst alarms, three heuristic techniques were developed for analysis and are shown in figure 1. Analysis was restricted to the three most prevalent alarm categories of desaturation, bradycardia and apnea; the small number of repeat alarms were excluded for simplification of analysis.

Cluster analysis–temporal relationship between alarms of the same category. In order to identify whether alarms from each of the three categories (considered independently) occurred in clusters and to determine the temporal relationship between them, we use a heuristic approach, described here with an example.

For every alarm of a particular category (e.g. desaturation, bradycardia and apnea, A_0), we counted all alarms of that category in a time window of ($\pm T$) as shown in figure 1(a). To gain insight into the temporal relationship between the occurrences of alarms, we plot a normalized histogram of the times at which alarms of that category occurred (with respect to A_0 which occurs at $t = 0$). In a normalized histogram, the y-axis represents the probability density function (pdf), and is a likelihood estimate ranging from zero to one for every value found in the x-axis.

In fact, the shape of the histogram, as developed here, is bound to be symmetric about $t = 0$ and gives insight into the time-constants of the physiological processes that generate alarms.

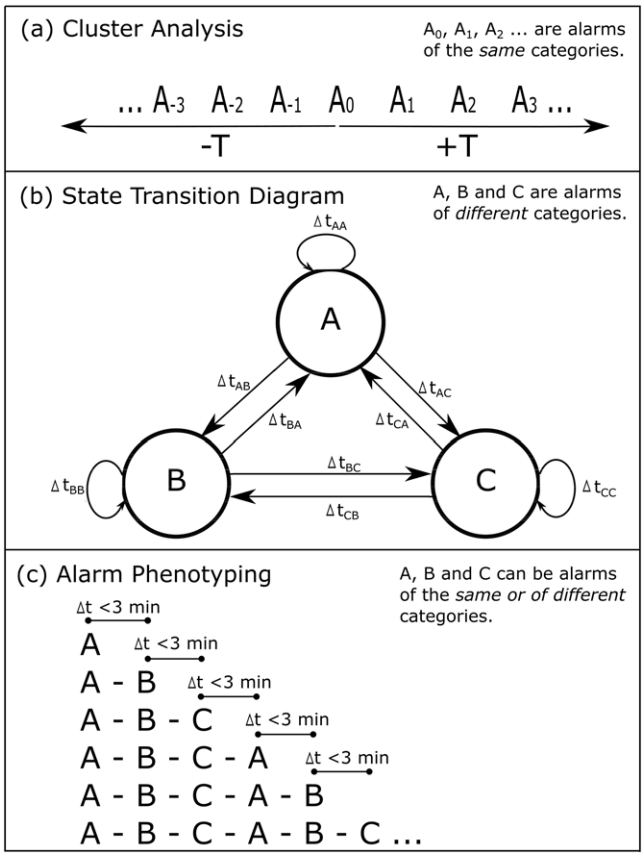

Figure 1. This schematic illustrates the three techniques developed to discover patterns in alarms. (a) Cluster analysis extracts the temporal relationship between alarms of the same category. (b) The state transition diagram captures the frequencies and the time to transition from one alarm category to the next. (c) Alarm phenotyping captures the prevalent sequences of alarm manifestation.

We empirically choose $T = 30$ min in order to identify alarm patterns in a clinically relevant time window and carry out this analysis for all desaturation, bradycardia and apnea alarms.

If patients rapidly oscillate between stable and alarm states, multiple redundant alarms of the same category can occur in short intervals of time giving rise to alarm clusters. For example, a patient may become bradycardic, recover quickly and then deteriorate again, leading to a cluster of bradycardia alarms. The expected reduction in alarm pressure that would have been generated by inhibiting further alarms of the cluster, after the alarm at $t = 0$ was explored. Inhibition was implemented for the time range of 0–5 min in increments of half a minute.

State transition diagram. If two or more alarms, of similar or different categories, occur due to the same precipitating factor, they will be closely connected in time. In order to explore the time-linkages between alarms of different categories that might arise due to the same physiological deterioration, we explored the percentage of times (i.e. frequency) with which alarms transit from one alarm state to another along with the times to transition, for desaturation, bradycardia and apnea alarms. All other alarms were excluded. For instance, if a desaturation

Table 2. Prevalence rates of all physiological alarms, in descending order.

Type of alarm	% of all physiological alarms
Desaturation	66.5
Bradycardia	21.8
Apnea	4.5
Asystole	2
High ABP	1.8
Ventricular fibrillation/tachycardia	1.5
Low ABP	1.4
Tachycardia	0.4

ABP = arterial blood pressure.

was followed by a low ABP and a bradycardia alarm, the transition would be registered as a desaturation to bradycardia. This assumption is reasonable because alarms of other categories constitute only 7% of the total alarms. These transition times are represented by normalized histograms censored at five minutes. In addition, a summary measure in terms of median and interquartile range (IQR) is also provided. These transition frequencies and transition times are diagrammatically represented in what we call the 'state transition diagram' (figure 1(b)). This analysis was carried out per patient per day, i.e. a transition is valid as long as it occurs on the same day. Therefore the first and last alarms of the day are ignored for analysis since they have no backward and forward reference.

Alarm phenotyping—sequencing. A string of alarms, such as multiple desaturation alarms, may arise as a result of specific physiological instabilities. All alarm sequences with inter-alarm durations, between consecutive alarms, of 3 min or less were identified and extracted in order to explore the prevalent sequences that arise in clinical practice (figure 1(c)). This choice of 3 min was largely empirical and was motivated by the inhibition time of the monitor to exclude effects of silencing. The total number of unique sequences, their frequencies of occurrence and the values of HR and SpO_2 corresponding to each sequence—obtained by averaging the HR and SpO_2 values associated with every alarm of the sequence—were extracted.

Results

Alarm demographics

A total of 489 780 alarms were logged through the course of the study. 93% of all alarms, as diagnosed by the patient monitor, were physiological and 6% were technical (e.g. sensor malfunction). The remaining originated from test periods. Five of the ten technical alarm-categories (5.9% of all alarms) resulted from disconnections or defects in the ECG, SpO_2 or BP sensors and the remaining five from EC10-based issues. A total of 2% of the alarms were repeat alarms and were removed from further analysis.

Table 2 gives the prevalence rates of the physiological alarms. Since desaturation (66.5%), bradycardia (21.8%) and apnea (4.5%) accounted for 92.8% of all physiological alarms, further quantitative analysis was carried out on these categories. The number of physiological alarms generated per patient per day showed considerable variability—the median and interquartile values, obtained from the analysis of 8522 patient days were 22 and 8–53, respectively.

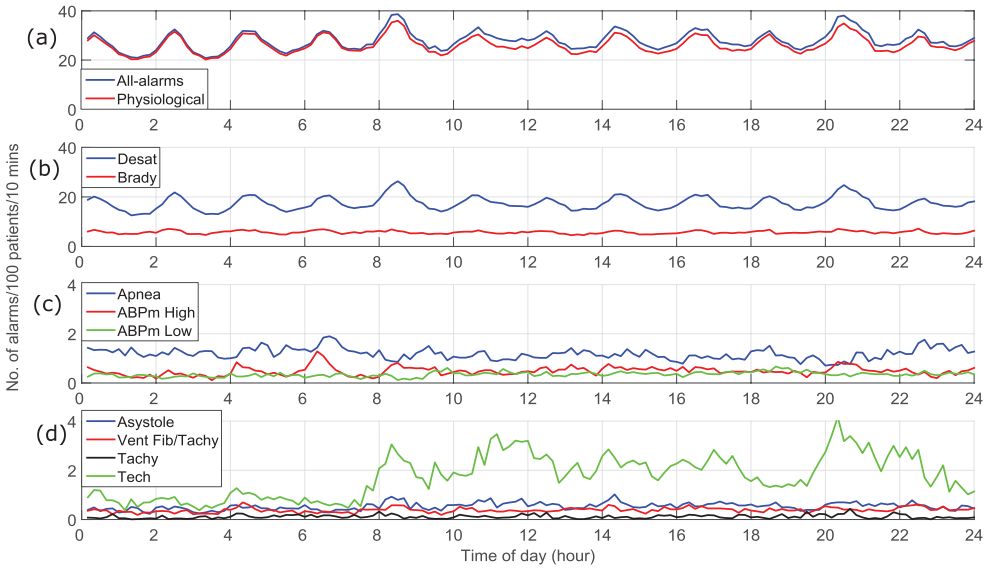

Figure 2. (a) The average values for all-alarms (blue) and physiological alarms (red) plotted through the course of the day. (b) The average values for desaturation (blue) and bradycardia (red) alarms. (c) The average values for apnea (blue), high ABP (red) and low ABP (green) alarms. (d) The average values for the three arrhythmia alarms—asystole (blue), ventricular fibrillation (red) and tachycardia (black) along with all technical alarms (green). The resolution of the x-axis is 10 min. The y-axis represents the number of alarms—the top two panels have a scale of 0–40 and are a factor of ten higher than the bottom two panels. The alarm rate is shown for 100 patients per 10 min.

Patterns in alarming through the course of the day

Figure 2(a) shows the average number of all alarms (physiological + technical + repeat) as well as the average number of physiological alarms through the course of the day. We observed a distinct two-hourly periodicity, with peaks at approximately half past all the even hours of the day (figure 2(a)). Since physiological alarms constitute the majority of the alarm load, analysis of the contributions of individual physiological alarm categories is shown in the next panels (figures 2(b)–(d)).

The two-hourly periodicity in desaturation and bradycardia alarms is clearly visible. A similar trend is present for apnea, although it is less prominent. Alarms of high ABP have two prominent peaks at 04:15 and 06:30 while the remaining physiological alarms are unremarkable.

In figure 2(d), technical alarms are shown alongside arrhythmia alarms—as diagnosed by the patient monitor). Arrhythmia alarms are addressed, because it is well known that the majority of arrhythmia alarms in the NICU are false positives. Sources of erroneous arrhythmia alerts include algorithmic errors, low signal-to-noise ratio, etc. Technical alarms are noticeably higher between 08:00–23:00 h, with an increasing number at 08:00, 11:00 and 20:00 h.

Repeat alarms, while not displayed here, occurred at a rate of 0.64 alarms per patient per day and demonstrated no periodic relationship with the time of day. Analysis of the patterns in alarming between the weekdays and the weekend showed no significant differences.

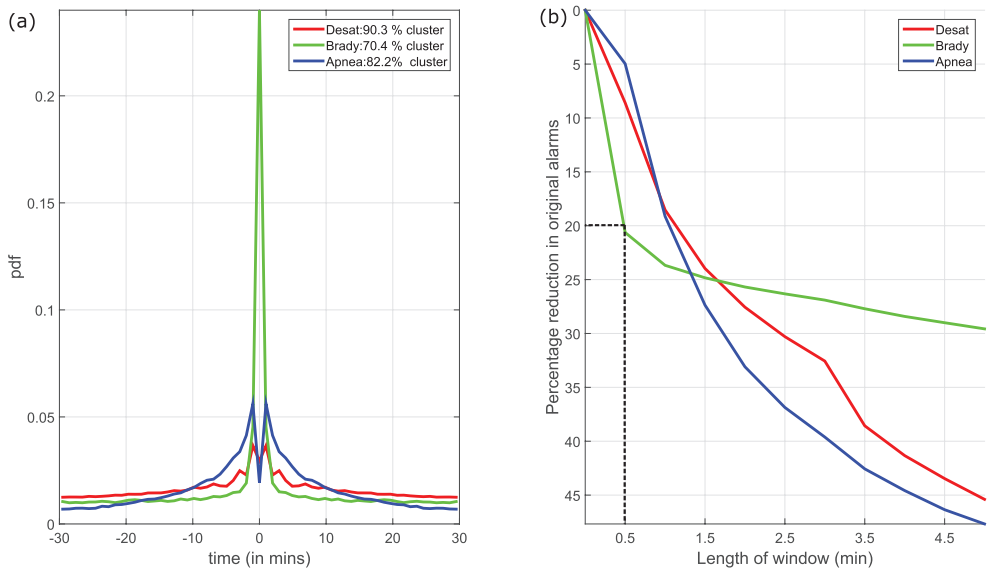

Figure 3. (a) The normalized histograms of the times of occurrence of desaturation (red), bradycardia (green) and apnea (blue) alarms centered about $t = 0$. As can be seen from the height of the pdf and the steepness of the slope at $t = 0$, the vast majority of bradycardia clusters are small while desaturation clusters occur over longer durations of time. The legend indicates the percentage of alarms from each category that cluster (i.e. another alarm of the same category was present within ± 30 min). The y-axis represents the pdf (normalized histogram). (b) The percentage reduction in original alarms that would have occurred if further alarms of the cluster, following the first alarm, were inhibited for various durations of time. For example, inhibition of bradycardia alarms for 30 s would reduce the number of bradycardia alarms by 20% (dashed line).

Cluster analysis–temporal relationship between alarms of the same type

The results from cluster analysis are shown in figure 3(a) and it displays the normalized histograms of the times of occurrence for desaturation, bradycardia and apnea alarms. 90, 70 and 82 percent of these alarm-categories cluster—i.e. another alarm of the same category was present within ± 30 min. Notably, as seen from the height of the pdf and the steepness of the slope at $t = 0$, the vast majority of bradycardia clusters are small. On the other hand, subjects generate desaturation alarms over longer periods of time, resulting in larger cluster-sizes for desaturation. The behavior of apnea lies midway between the two. Notably, the histograms of desaturation and apnea exhibit a notch at $t = 0$.

The percentage reduction in original alarms by inhibiting alarms within a cluster (after the one at $t = 0$) is visually represented in figure 3(b). We can see that the steepness of decline for bradycardia alarms begins to stabilize beyond 30 s. If all bradycardia alarms were inhibited for this time period, following the first bradycardia alarm, and if similar inhibitions were implemented for desaturation and apnea for durations of 60 s, the total number of non-actionable physiological alarms would reduce by approximately 20%.

State transition diagram

The state transition diagram for desaturation, bradycardia and apnea is shown in figure 4(a). The thickness of the arrows encode the frequencies of transition (in %), with thicker arrows

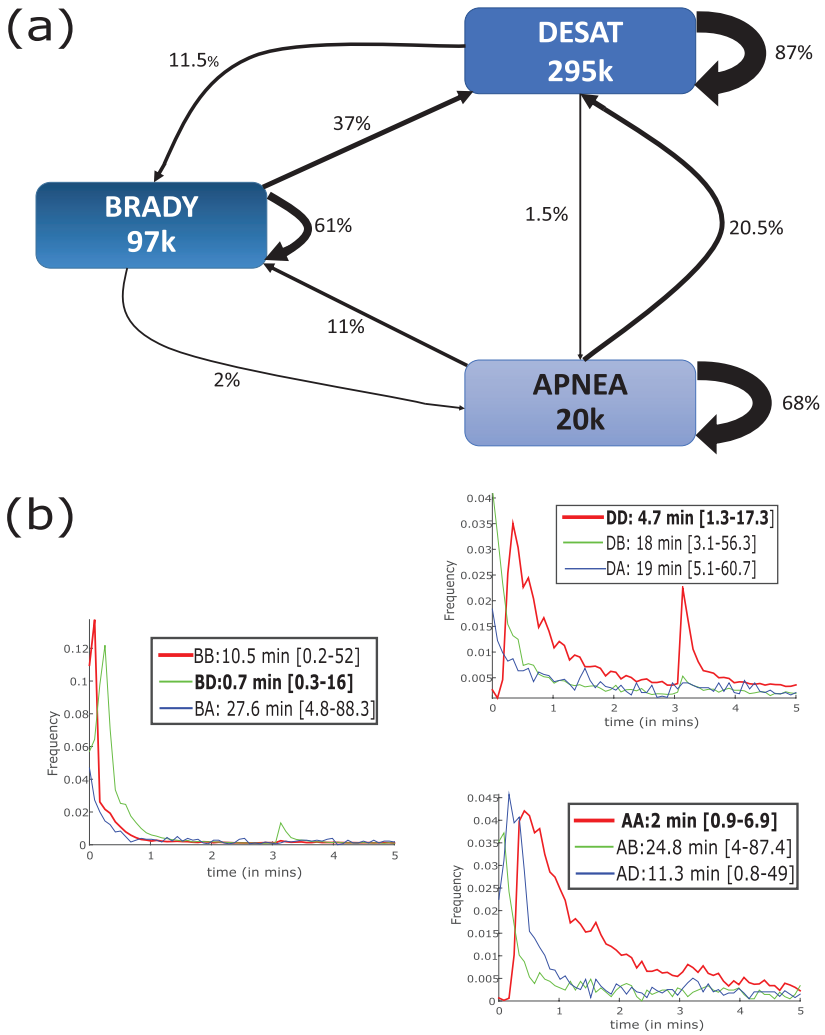

Figure 4. (a) The three most prevalent alarm states along with the total number of alarms in each of these categories. For each alarm state, the percentage of times the alarm transited to another alarm state is listed beside the arrows. The thickness of the arrows encode the frequencies of transition (in %) with thicker arrows representing higher frequencies. (b) For each alarm category, the normalized histograms of the time-to-transition are shown up to a duration of five minutes (beyond which data are sparse). Desaturation, bradycardia and apnea are abbreviated as 'D', 'B' and 'A'. The legend reads the direction of transition, e.g. 'BA' refers to transition from bradycardia to apnea along with the median (IQR) time to transition in minutes. Alarm transitions with short median times to transition are highlighted in bold. Note that the *y*-axes represent the pdf (normalized histogram) and that the *y*-axis for transitions from bradycardia is thrice the scale of the other two.

representing higher frequencies. Figure 4(b) shows the transition times of alarms originating from each category. Notably, the median times for transition from desaturation to desaturation again and from apnea to apnea again, along with the transition times from bradycardia to desaturation are short and are highlighted in bold. The first quartile transition times of

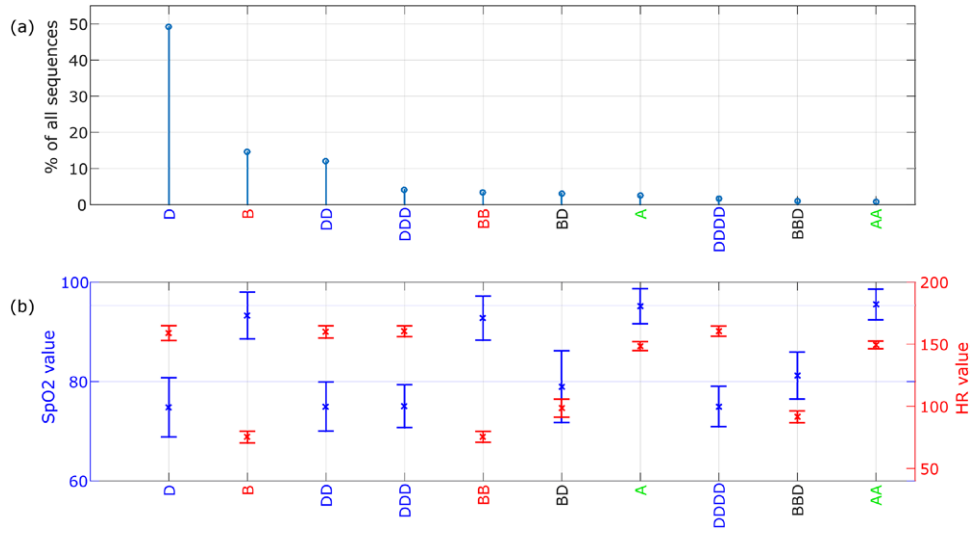

Figure 5. The ten most prevalent sequences, from the 1036 unique alarm sequences that were identified, are shown along the *x*-axis. (a) The percentage contribution of each sequence to the total number of sequences (255 223). (b) The mean and standard deviation of the SpO_2 (blue) and HR (red) values for the top ten alarm sequences—the HR and SpO_2 of each sequence is obtained by averaging the HR and SpO_2 values associated with each alarm of the sequence, as indicated in the alarm message.

bradycardia to bradycardia again and from apnea to desaturation are remarkably short too. Notably, 37% of all bradycardia alarms are followed by a desaturation with a median transition time of just 42 s (0.7 min).

The histograms of the three transitions that originate from each of the alarm categories of desaturation, bradycardia and apnea are shown in figure 4(b). The spatial arrangement of these histograms follows that of the frequency pattern in figure 4(a). Importantly, despite a long median time to transition in some cases, for instance desaturation to bradycardia (DB), a significant number of these transitions might occur in very short time-windows indicating that they are reflecting the same physiological instability. Thus, while the median and inter-quartile times give a summary of the transition times, the histograms give valuable complementary information. The effect of silencing alarms and the generation of a new alarm following the 3 min of silent period are clearly seen in the peaks of DD and BD at the 3 min interval.

Alarm phenotyping—sequencing

Sequencing, with an intra-alarm duration of 3 min or less between consecutive alarms, resulted in 1036 unique sequences and a total of 255 223 alarm sequences. The ten most prevalent sequences comprised more than 90% of all alarm sequences and figure 5(a) shows the prevalence rates of these top ten most frequently observed sequences. Figure 5(b) shows the means and standard deviations of the HR and SpO_2 values of these sequences.

It is noteworthy that homogenous alarm sequences such as those comprising solely of desaturation or bradycardia alarms are associated with drops in either SpO_2 (desaturation) or HR (bradycardia) and not both. Remarkably, we found that even sequences comprising of six desaturations showed no decrease in HR. Interestingly, sequences comprising solely apneas

show seemingly stable values of SpO_2 and HR. We checked the 25 most prevalent sequences (97% of all alarm sequences) and found that drops in both HR and SpO_2 are noted only when the sequence contains at least one instance of both bradycardia and desaturation (with the exception of AD).

Discussion

Our research develops and deploys heuristic techniques (cluster analysis, state transition diagram and phenotyping) on a very large dataset of critical patient-monitor alarms that originated from neonates under intensive care. We explore patterns in these alarms, gaining insight into the typical manifestation of patient deterioration. We also look into the redundancies of multiple, temporally clustered alarms.

Analysis of nearly 500 000 critical patient-monitor alarms shows that 93% of all alarms, as determined by the monitor, are of physiological origin and that nearly 86% of all alarms are desaturation, bradycardia or apnea—therefore our efforts are focused on these categories alone. Remarkably the two-hourly periodicity in alarms (figure 2) is found to be synchronized with nursing care, which commences at the even hours of the day and involves nursing, suctioning the bronchopulmonary tract and feeding the babies. These peaks appear to be correlated with periods of enteral feeding which were determined to take place at approximately 30 min after the beginning of nursing care. We hypothesize that enteral feeding contributes to cardio-respiratory instability and physiological distress due to the timing of the peaks in the alarm pressure of desaturation. In our unit, administration of feeding is the last aspect of nursing care and does not involve handling the baby (incubator is closed). This fact along with the observation that the peaks in alarm pressure are approximately 30 min after the start of nursing rounds suggests that enteral feeding is the major contributor to the increase in alarm pressure, although we cannot entirely exclude that nurse handling may play a role. This is consistent with our earlier work where we saw a similar pattern of two-hourly increases (Pul *et al* 2015). In our current study we use considerably more data and a finer resolution (10 min, and not hourly) allowing us to confidently correlate enteral feeding with increased alarm pressure.

Furthermore, we observe (figure 2) that peaks exist in both physiological and technical alarms at around 08:00 and 20:00 h respectively. These periods coincide with the 'large' morning round and parent visits (entailing kangaroo care, etc) respectively. Around 11:00 h, the patterns of technical alarms are noticeably different. At this time radiological rounds occur that consist of cerebral and cardiac ultrasonography and x-ray imaging.

At 06:30 and 04:30, on the weekdays and the weekend respectively, arterial blood samples are collected and the associated devices are calibrated. High ABP alarms shows corresponding peaks at these times, which suggests an association with nursing intervention and device calibration.

Cluster analysis (figure 3) reveals that alarms infrequently occur in isolation—i.e. temporally distant from other alarms of the same category. The cluster patterns of desaturation, bradycardia and apnea are distinctively dissimilar. For instance, bradycardia alarms have the highest likelihood of occurring in isolation (30%). However, when they do in fact cluster, they are closely associated in time and arise in small clusters. Desaturation and apnea alarms often occur in larger clusters and over longer durations of time. The presence of long desaturation clusters suggests that the physiological systems responsible for saturation are slower to respond to clinical/nursing intervention and that this inertia is being reflected in multiple instances of desaturation alarms arising over relatively short periods of time. The graphs for both desaturation and apnea show notches centered at $t = 0$. Likely, these result from the fact

that before attending to the neonate, nurses often silence monitors preventing the generation of further alarms. The default silencing time is 3 min, even for red alarms. For bradycardia, this notch is absent because neonates who often get bradycardic, recover rapidly only to deteriorate again (oscillatory behavior) before the responsible nurse can respond. This does not indicate that nurse response times are slow since analysis from our earlier work supports the opposite (Pul *et al* 2015). We term this oscillatory manifestation of arrhythmia as flash-bradys.

These patterns of clustering in alarms offer the opportunity to reduce non-actionable alarms by inhibiting *further alarms* for various durations of time, following the first alarm of a cluster. While this approach is similar to that of delays, suggested by Görges *et al* (2009), it has the advantage that, unlike delays, the window of opportunity for the nurse to act on the alarm is not reduced—an absolute necessity when alarms indicate life-threatening situations. We identify that if further bradycardia alarms were inhibited for 30 s and desaturation and apnea for 60 s each, physiological alarms would decrease by 20%. While safety is of paramount concern when considering any alarm reduction strategy, it is noteworthy that at present following an alarm in our NICU, further alarms of the same category are inhibited in the handheld devices of nurses for a period of 45 s. This requires nurses to use the central monitor or the inter-bed communication to further evaluate the status of the baby. In such cases, sounding additional alarms of the same category at the central station is largely redundant. As shown in figure 3(b), the bradycardia curve starts to steeply level off at the 30 s mark. Therefore, inhibiting bradycardia alarms beyond this point yields only marginal reduction in alarms and is unmerited. This is a valuable finding pointing to the presence of flash-bradys, a situation in which the handheld devices do not transmit new bradycardia alarms for a period of 45 s. However, the presence of new bradycardia alarms during this time period points to the fact that the patient stabilized and deteriorated again. Meanwhile, desaturation and apnea alarms show clustering even beyond 60 s. We judge this time-window to be sufficient for nurses to arrive at the bedside, initiate evaluation and execute a response strategy, at least given the layout and characteristics of our NICU. Multiple alarms, originating from the same physiological deterioration, within this one-minute time window are deemed redundant.

The state-transition diagram and the phenotyping of alarms capture patterns originating from different alarm categories. The most valuable finding from the state-transition diagram is that 37% of all bradycardia alarms are followed by a desaturation and with a median time of just 42 s. This information raises the question of whether the clinical significance of such a deterioration cluster is different from the value of the individual alarms, considered alone. And if so, can we offer earlier predictions of such deterioration by analyzing continuous vital signs and thereby providing an increased window of opportunity for pre-emptive action?

While the state-transition diagram captures the ways in which alarms transit from one category to the next, alarm phenotyping captures frequencies of such sequences of more than two alarms. Phenotyping, as implemented, reduces the number of desaturation, bradycardia and apnea alarms into alarm sequences constituting 60% of the original alarms. The most striking feature of the alarm sequences is that nearly 50% and 15% of all sequences respectively consist of isolated alarms of desaturation and bradycardia and are potentially irreducible.

These findings lead to the following two questions: (i) what percentage of these alarms are clinically relevant?, and, (ii) are the vital signals leading up to a string of alarms (e.g. DDDD) distinctively different from those leading to an isolated alarm (e.g. D)? The first question requires a system capable of registering, either automatically or through the input of a nurse, whether an alarm is clinically relevant or not. In order to sufficiently address the second question, an answer to the first is a prerequisite. If the isolated instances of alarms are in fact clinically relevant, these alarms which constitute approximately 40% of the total alarms are

irreducible. If however these alarms are clinically impertinent, further research to explore signal processing or statistical techniques to reduce these isolated alarm occurrences are needed. Finally, the presence of long strings of desaturation alarms is noteworthy and suggests that the physiological mechanisms responsible for increasing SpO_2 levels are slow despite nursing intervention. It is worth mentioning that broadening alarm limits will increase risk but is unlikely to affect the frequencies of DD, DDD, DDDD, BB, BBB etc since the mean values of the SpO_2 and HR (figure 5(b)) for the alarms contributing to the alarm sequence are considerably lower than the desired zone ($>$85 for SpO_2 and $>$100 for HR).

Past research (Vergales *et al* 2014) has indicated that apnea is often under-detected by monitors and this is supported by our findings where apnea alarms constitute only 4% of all physiological alarms. Conspicuously, the ABD phenotype is uncommon with a prevalence of just 0.2%. Also, sequences containing only apnea alarms are not associated with reductions in HR or SpO_2. This suggests that sequences containing apnea alarms are either false alarms or very long apnea events that are in fact coupled with a slow decrease in both HR and oxygen saturation (Mohr *et al* 2015).

The strength of this study stems from the quantitative characterization of patterns in alarming obtained by analyzing a large amount of data. We show that alarming is not an entirely stochastic process, and that the deterministic relationship between alarms, now captured, can be exploited to safely reduce alarms by inhibiting further alarms of a cluster. Moreover, our study showcases alarm patterns that can be targeted for future studies.

This study also has some limitations which include the absence of the clinical relevance of alarms, which remains an open problem, absence of information on the duration of individual alarms and the fact that continuous physiological signals are not available. Furthermore, the influence of silencing of alarms by the nurses is lost. If alarms from these periods were present, it is likely that alarm clusters would be larger, transition times shorter and alarm sequences longer, leading to an increased synergy in the relationships amongst alarms, than that captured here. These limitations are however, adjunct to the primary aim of the study which is quantifying the patterns in alarming, as they occurred in the NICU.

Conclusions

We developed techniques to analyze and visualize life-threatening medical alarms in critically ill neonates. These techniques for quantitatively identifying patterns in alarming can be used for addressing alarm fatigue by reducing redundant alarms, gaining insight into the prevalent patterns of physiological deterioration, capturing the temporal relationships between alarms and identifying the simultaneous derangement of multiple parameters.

Acknowledgments

This research was performed within the framework of IMPULS perinatology.

References

Ahlborn V, Bohnhorst B, Peter C S and Poets C F 2000 False alarms in very low birthweight infants: comparison between three intensive care monitoring systems *Acta Paediatr.* **89** 571–6

Albert N M, Murray T, Bena J F, Slifcak E, Roach J D, Spence J and Burkle A 2015 Differences in alarm events between disposable and reusable electrocardiography lead wires *Am. J. Crit. Care* **24** 67–74

Altimier L and Phillips R M 2013 The neonatal integrative developmental care model: seven neuroprotective core measures for family-centered developmental care *Newborn Infant Nurs. Rev.* **13** 9–22

Atzema C, Schull M J, Borgundvaag B, Slaughter G R and Lee C K 2006 ALARMED: adverse events in low-risk patients with chest pain receiving continuous electrocardiographic monitoring in the emergency department. A pilot study *Am. J. Emerg. Med.* **24** 62–7

Baumgartner B, Rodel K and Knoll A 2012 A data mining approach to reduce the false alarm rate of patient monitors *Proc. of the 34th Annual Int. Conf. of the IEEE EMBS* pp 5935–8

Bliss J P and Dunn M C 2000 Behavioural implications of alarm mistrust as a function of task workload *Ergonomics* **43** 1283–300

Blum J M and Tremper K K 2010 Alarms in the intensive care unit: too much of a good thing is dangerous: is it time to add some intelligence to alarms? *Crit. Care Med.* **38** 702–3

Bonzheim K A, Gebara R I, O'Hare B M, Ellis R D, Brand M A, Balar S D, Stockman R, Sciberras A M and Haines D E 2011 Communication strategies and timeliness of response to life critical telemetry alarms *Telemed. J. e-Health: Official J. Am. Telemed. Assoc.* **17** 241–6

Borowski M, Görges M, Fried R, Such O, Wrede C and Imhoff M 2011a Medical device alarms *Biomed. Tech./Biomed. Eng.* **56** 73–83

Borowski M, Siebig S, Wrede C and Imhoff M 2011b Reducing false alarms of intensive care online-monitoring systems: an evaluation of two signal extraction algorithms *Comput. Math. Methods Med.* **2011** 143480

Bustamante E A, Bliss J P and Anderson B L 2007 Effects of varying the threshold of alarm systems and workload on human performance *Ergonomics* **50** 1127–47

Chambrin M C 2001 Alarms in the intensive care unit: how can the number of false alarms be reduced? *Crit. Care-London* **5** 184–8

Cvach M 2012 Monitor alarm fatigue: an integrative review *Biomed. Instrum. Technol.* **46** 268–77

Cvach M M, Frank R J, Doyle P and Stevens Z K 2013 Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals *J. Nurs. Care Qual.* **29** 9–18

Drew B J, Harris P, Zègre-Hemsey J K, Mammone T, Schindler D, Salas-Boni R, Bai Y, Tinoco A, Ding Q and Hu X 2014 Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients *PLoS ONE* **9** e110274

Funk M, Clark J T, Bauld T J, Ott J C and Coss P 2014 Attitudes and practices related to clinical alarms *Am. J. Crit. Care* **23** 9–19

Görges M, Markewitz B A and Westenskow D R 2009 Improving alarm performance in the medical intensive care unit using delays and clinical context *Anesth. Analg.* **108** 1546–52

Graham K C and Cvach M 2010 Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms *Am. J. Crit. Care* **19** 28–34

Imhoff M and Kuhls S 2006 Alarm algorithms in critical monitoring *Anesth. Analg.* **102** 1525–37

Imhoff M, Kuhls S, Gather U and Fried R 2009 Smart alarms from medical devices in the OR and ICU *Best Pract. Res. Clin. Anaesthesiol.* **23** 39–50

Keller J P 2012 Clinical alarm hazards: a "top ten" health technology safety concern *J. Electrocardiol.* **45** 588–91

Konkani A, Oakley B and Bauld T J 2012 Reducing hospital noise: a review of medical device alarm management *Biomed. Instrum. Technol.* **46** 478–87

McCartney P R 2012 Clinical alarm management *MCN: Am. J. Matern. Child Nurs.* **37** 202

McMahon E, Wintermark P and Lahav A 2012 Auditory brain development in premature infants: the importance of early experience *Ann. N. Y. Acad. Sci.* **1252** 17–24

Mitka M 2013 Joint commission warns of alarm fatigue: multitude of alarms from monitoring devices problematic *JAMA* **309** 2315–6

Mohr M A *et al* 2015 Very long apnea events in preterm infants *J. Appl. Physiol.* **118** 558–68

O'Carroll T M 1986 Survey of alarms in an intensive therapy unit *Anaesthesia* **41** 742–4

Pul C, Mortel H P M E, Bogaart J J L, Mohns T and Andriessen P 2015 Safe patient monitoring is challenging but still feasible in a neonatal intensive care unit with single family rooms *Acta Paediatr.* **104** e247–54

Schmid F, Goepfert M S, Kuhnt D, Eichhorn V, Diedrichs S, Reichenspurner H, Goetz A E and Reuter D A 2011 The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery *Anesth. Analg.* **112** 78–83

Schoenberg R, Sands D Z and Safran C 1999 Making ICU alarms meaningful: a comparison of traditional versus trend-based algorithms *Proc. of the AMIA Symp., American Medical Informatics Association* p 379

Sendelbach S 2012 Alarm fatigue *Nurs. Clin. North Am.* **47** 375–82

Siebig S, Kuhls S, Imhoff M, Gather U, Schölmerich J and Wrede C E 2010 Intensive care unit alarms—how many do we need? *Crit. Care Med.* **38** 451–6

Vergales B D *et al* 2014 Accurate automated apnea analysis in preterm infants *Am. J. Perinatol.* **31** 157–62

Wachman E M and Lahav A 2011 The effects of noise on preterm infants in the NICU *Arch. Disease Child.-Fetal Neonatal Ed.* **96** F305–9

Walsh M C, Powers E and Fanaroff J 2015 The potential for harm from alarm fatigue in single-room NICUs *Acta Paediatr.* **104** 436–7

Weaver J M 2013 Alarm fatigue can decrease the safety of dental office sedation and anesthesia *Anesth. Prog.* **3006** 93–4