

# Reduction of clinically irrelevant alarms in patient monitoring by adaptive time delays

Felix Schmid<sup>1</sup> · Matthias S. Goepfert<sup>1</sup> · Frank Franz<sup>2</sup> · David Laule<sup>2</sup> · Beate Reiter<sup>3</sup> · Alwin E. Goetz<sup>1</sup> · Daniel A. Reuter<sup>1</sup>

Received: 21 February 2015 / Accepted: 24 November 2015 / Published online: 30 November 2015  
© Springer Science+Business Media Dordrecht 2015

**Abstract** The problem of high rates of false alarms in patient monitoring in anesthesiology and intensive care medicine is well known but remains unsolved. False alarms desensitize the medical staff, leading to ignored true alarms and reduced quality of patient care. A database of intra-operative monitoring data was analyzed to find characteristic alarm patterns. The original data were re-evaluated to find relevant events and to rate the severity of these events. Based on this analysis an adaptive time delay was developed that individually delays the alarms depending on the grade of threshold deviation. The conventional threshold algorithm led to 4893 alarms. 3515 (71.84 %) of these alarms were annotated as clinically irrelevant. In total 81.0 % of all clinically irrelevant alarms were caused by only mild and/or brief threshold violations. We implemented the new algorithm for selected parameters. These parameters equipped with adaptive validation delays led to 1729 alarms. 931 (53.85 %) alarms were annotated as clinically irrelevant. 632 alarms indicated the 645 clinically relevant events. The positive predictive value of occurring alarms improved from 28.16 % (conventional algorithm) to 46.15 % (new algorithm). 13 events were missed. The false positive alarm reduction rate of the algorithm ranged from

33 to 86.75 %. The overall reduction was 73.51 %. The implementation of this algorithm may be able to suppress a large percentage of false alarms. The effect of this approach has not been demonstrated but shows promise for reducing alarm fatigue. Its safety needs to be proven in a prospective study.

**Keywords** Alarms · Monitoring · Alarm fatigue · Cardiac surgery

## 1 Introduction

The frequency of visual and auditory alarms in complex surgical and intensive care settings is very high [1]. High rates of false alarms (alarms without direct or indirect clinical consequence) can desensitize the attending staff to such warnings, known as “alarm fatigue” or the so-called “crying wolf” phenomenon. Desensitization can lead to ignored true alarms and reduced quality of care for the patient [2]. Clinically irrelevant alarms, defined as alarms without direct or indirect clinical consequence, are often caused by inadequate threshold settings or signal artifacts. Therefore, a reduction these alarms can be achieved by optimized threshold settings and artefact detection [3, 4].

A recent investigation of alarms during cardiac surgery revealed that overall, there was a frequency of 1.2 alarms/min, with 80 % of those alarms having no direct clinical consequence (i.e. clinically irrelevant alarms) [5].

Most monitoring devices have fixed time delays of defined length to suppress threshold alarms caused by very short threshold violations (e.g. by signal artifacts). If the threshold violation still exists after a few seconds, the alarm will be generated. The violation delay is fixed and can not be configured differently for each variable (e.g.

✉ Felix Schmid  
fe.schmid@uke.de

<sup>1</sup> Department of Anesthesiology, Hamburg-Eppendorf University Hospital, Martinistr. 52, 20246 Hamburg, Germany

<sup>2</sup> Drägerwerk AG & Co. KGaA, Moislinger Allee 53, 23542 Lübeck, Germany

<sup>3</sup> Department of Cardiovascular Surgery, Hamburg-Eppendorf University Hospital, Martinistr. 52, 20246 Hamburg, Germany

heart rate, blood pressure). Our first results showed that this approach—at least in highly complex surgery, such as cardiac surgery—performs well in suppressing alarms, which were caused by very short artifacts, but is insufficient, when alarms were caused by mild threshold violations.

Based on the results that 88.5 % of all clinically irrelevant alarms were caused by only mild and brief threshold violations we aimed to develop an alarming algorithm that is able to minimize the alarm frequency without decreasing the overall sensitivity of the patient monitoring.

## 2 Methods

### 2.1 Analysis of the database

We re-analyzed a pre-existing multi-parameter database including 25 cases (124 h) of intraoperative patient and alarm data during elective cardiac surgery [5]. Patient data collection for this study was approved by the local IRB.

This database contained digital recordings of all physiological signals and all alarms from the patient monitor and anesthesia device. Additionally, video recordings of the anesthesia workplace were collected from two different angles. The videos allowed retrospective annotation and correlation with the clinical situation and assessment of the masked anesthesiologist's reaction to the alarms. Due to the masking of the medical staff in the OR we only considered real reactions of the staff because the intent to react is not sufficient detectable. All alarm events were annotated by two independent experienced anesthesiologists and categorized as technically true/false and clinically relevant/irrelevant and, according to its clinical importance as warning, serious or life threatening [5] (see Table 1). An event was defined from the beginning until the end of a clinically relevant threshold violation and could consist of several alarms Table 2 shows the results of the categorization to events and the length of these events (see Table 2).

### 2.2 The algorithm

Based on the aforementioned data files with a frequency of one value per second, we developed an automated alarm filter for adult patients based on an adaptive alarm delay to remove alarms caused by clinically irrelevant mild threshold violations of short duration. In the present study a mild violation was defined as a mean violation of <4 % beyond the threshold.

Starting from the assumption that the lesser the violation the lesser the clinical importance, the delay for minor violations is long and decreases if the threshold violation

**Table 1** Terms contained in the text and their definition according to their use in the study

Term	Definition
Degrees of threshold violation	<i>Mild</i> a mean violation of <4 % beyond the threshold <i>Brief</i> a threshold violation with a duration lower than 16 s
Clinical relevant alarms	Alarms with direct or indirect clinical consequence [a]
Clinical irrelevant alarms	Alarms without direct or indirect clinical consequence [b]
Event	Duration from the beginning until the end of a clinically relevant threshold violation [c]
Alarm limits	<i>Warning</i> lowest priority, the threshold is set by the user [d] <i>Serious</i> medium priority, calculated 30 % above the upper warning threshold/30 % below the lower warning limit [e] <i>Life threatening</i> highest priority, an implemented clinical consensus-based limit [f]

The contained indices also mark these terms in Tables 2, 5 and Fig. 1

**Table 2** The amount of all clinically relevant annotated events subclassified by their relevance and the corresponding mean duration (start until the end of the present threshold violation) including the SD

	All annotated events [c]	Annotated events graded as		
		Warning	Warning	Warning
Quantity	645	105	462	78
Mean duration (s)	208 (±331)	208 (±332)	206 (±328)	199 (±316)

**Table 3** The differentiation of all conventional alarms according to their clinical relevance and the amount and duration of the threshold violations (mild and brief)

Limit	Source	Upper limit heart rate (bpm)	Time-delay (s)
Warning	User	90	60
Serious	Calculated	123	15
Safety	Knowledge-based	200	6

increases (see Table 3). The adaptive alarm delay should be able to distinguish between three different grades of severity of threshold violations. The user sets a threshold for the lowest grade (warning threshold) and the monitor then automatically sets two additional thresholds: first, clinical consensus based maximum as safety threshold, and second, an interpolated “serious threshold” (30 % above the upper warning threshold/below the lower warning threshold; see Table 4). To achieve consensus, four experienced physicians discussed the parameterization and all

**Table 4** Clinical consensus based alarm thresholds used for adaptive alarm algorithm

Algorithm	PARAM	ALL	FP	FN	TP	Sensitivity (%)	PPV (%)	FP reduction (%)
CVTH	ALL	4893	3515	0	1378	100.00	28.16	0.00
	HR	869	679	0	190	100.00	21.86	0.00
	PLS	265	176	0	89	100.00	33.58	0.00
	ART S	693	500	0	193	100.00	27.85	0.00
	ART M	1761	1099	0	662	100.00	37.59	0.00
	ART D	572	464	0	108	100.00	18.88	0.00
	CVP	532	468	0	64	100.00	12.03	0.00
	SpO <sub>2</sub>	126	72	0	54	100.00	42.86	0.00
	Ta	75	57	0	18	100.00	24.00	0.00
STH	ALL	1729	931	13	798	98.40	46.15	73.51
	HR	164	90	1	74	98.67	45.12	86.75
	PLS	100	44	1	56	98.25	56.00	75.00
	ART S	213	80	5	133	96.38	62.44	84.00
	ART M	633	257	1	376	99.73	59.40	76.62
	ART D	383	300	2	83	97.65	21.67	35.34
	CVP	120	93	0	27	100.00	22.50	80.13
	SpO <sub>2</sub>	61	29	2	32	94.12	52.46	59.72
	Ta	55	38	0	17	100.00	30.91	33.33
	PARAM	ALL	CVTH event-detection		STH event-detection		STH event-detection (%)	
			Not detected	Detected	Not detected	Not detected	Detected	Not detected
Annotations/ events	ALL	645	0	645	13	632	2.02	97.98
	HR	47	0	47	1	46	2.13	97.87
	PLS	43	0	43	1	42	2.33	97.67
	ART S	117	0	117	5	112	4.27	95.73
	ART M	313	0	313	1	312	0.32	99.68
	ART D	60	0	60	3	57	5.00	95.00
	CVP	21	0	21	0	21	0.00	100.00
	SpO <sub>2</sub>	28	0	28	2	26	7.14	92.86
	Ta	16	0	16	0	16	0.00	100.00

alarm limits reflecting most realistically the needs for the vast majority of adult patients known to us. Since changes in different variables are of different importance (e.g. a fall in blood pressure must be detected faster than a decrease of body temperature), we aimed to combine each parameter with distinct time delays. For example, for heart rate the warning limit is combined with a time delay of 60 s, the serious-limit is combined with a time delay of 15 s and violations of the safety limit are delayed for 6 s. So, if the user sets the upper heart rate threshold to 90 beats per minute (bpm), the monitor automatically calculates a serious limit of 123 bpm. The safety limit of 200 bpm is knowledge-based (derived from clinician consensus) and set automatically. The monitor then checks every new value of the heart rate for a threshold violation. If the threshold of the warning limit is violated, an internal counter will be incremented by 1 per second. For every

new heart rate sample, it will be checked, if the counter is higher than the corresponding delay. If this is applicable, an alarm is generated. If the heart rate decreases below 120 bpm, the counter will be decremented. If the heart rate rises again, the counter increases again by 1.0 per second. If an alarm is confirmed by the user, all counters are reset to zero (see Fig. 1).

### 2.3 Implementation of the algorithm for selected parameters

In this work we focused on selected parameters of the patient monitor. We implemented this adaptive time delay for the parameters heart rate (derived from electrocardiogram), pulse rate (derived from pulse oximetry), arterial blood pressure (systolic, diastolic and mean), central venous pressure, oxygen saturation, and body temperature.

## 2.4 Reassessment of the database

Finally, the trend data from each case were printed and re-evaluated by two independent experienced anesthesiologists to assess potential false negative alarms missed by the algorithm. Both anesthesiologists were blinded to the results of alarm generation, and had the task to evaluate in pseudo-real-time without knowledge of the patients future, if, according to their clinical experience, an alarm was needed or not. Further, both had to prioritize each alarm they deemed clinically relevant according to their clinical appraisal. Necessary alarms had to be divided in three different degrees. Grade 1: low priority (warning), Grade 2: medium priority (serious), Grade 3: high priority (life-threatening). All annotator decisions (relevant events and alarm grades) were marked on separate work sheets (see Table 1). In cases of any discordance between the two annotators, a third experienced anesthesiologist was consulted to define the need and the priority of the alarm. The selected alarms were then compared with the alarms generated by the new algorithm as well as with alarms generated by conventional threshold alarms. This comparison was facilitated by the first author. By this, it was possible to keep the clinicians blinded for the results. Additionally, the decisions of the two observers were evaluated for interrater reliability.

## 3 Results

The assessment of patient data revealed 645 clinical events during 124 h of intraoperative monitoring in 25 patients.

The differentiation of alarms according to their grade of severity is shown in Table 3. 490 events were related to invasive arterial blood pressure (117 systolic, 313 mean and 60 diastolic). The interrater reliability of the annotators was evaluated by calculation of the Cohens-Kappa coefficient. The annotators agreed in their assessments in 87 % ( $\kappa = 0.87$ ) of events, and the third experienced anesthesiologist was required in 13 % of events.

### 3.1 Alarm data with conventional and new alarm algorithm

The conventional threshold algorithm led to 4893 alarms. 3515 (71.84 %) of these alarms were annotated as clinically irrelevant. 1378 alarms indicated relevant threshold violations within 645 clinical events. None of those events was missed by the conventional alarm algorithm resulting in 100 % sensitivity. Table 4 compares the results of the conventional threshold to the new approach. It contains the numbers of total alarms, alarms by parameter, false positive and false negative alarms by parameter, as well as the

false alarm reduction by parameter with the new approach. It also contains the differentiation of the alarms as well as the detected clinical events and the undetected clinical events in both approaches.

In addition, the subsequent analysis showed that 2471 (50.5 %) of all threshold violations were very brief, having a duration <16 s, of which 2119 (85.8 %) were annotated as clinically irrelevant. Further detailed information is given in Table 2.

We implemented the adaptive validation delay for the selected parameters described above. The new algorithm with adaptive validation delays led to 1729 alarms of the selected parameters. 931 (53.85 %) alarms were annotated as clinically irrelevant. 632 alarms indicated the 645 clinically relevant events. 13 events were missed by the new algorithm.

The false positive alarm reduction rate of the new algorithm ranged from 33 % (body temperature) to 86.75 % (heart rate). The overall reduction was 73.51 %.

The new approach showed differences in alarm sensitivity between the different parameters from 94.12 (SpO<sub>2</sub>/oxygen saturation) to 100 % (body temperature and central venous pressure). This corresponds to an overall sensitivity of 98.4 % of the new algorithm.

The positive predictive values (the proportion of positive test results that are true positive) reached from 21.67 (diastolic blood pressure) to 62.44 % (systolic blood pressure) with an overall value of 46.15 % compared to a positive predictive value of 28.16 % of the conventional algorithm.

Detailed information of false positive alarm reduction, sensitivity and positive predictive values of all parameters are shown in Table 5.

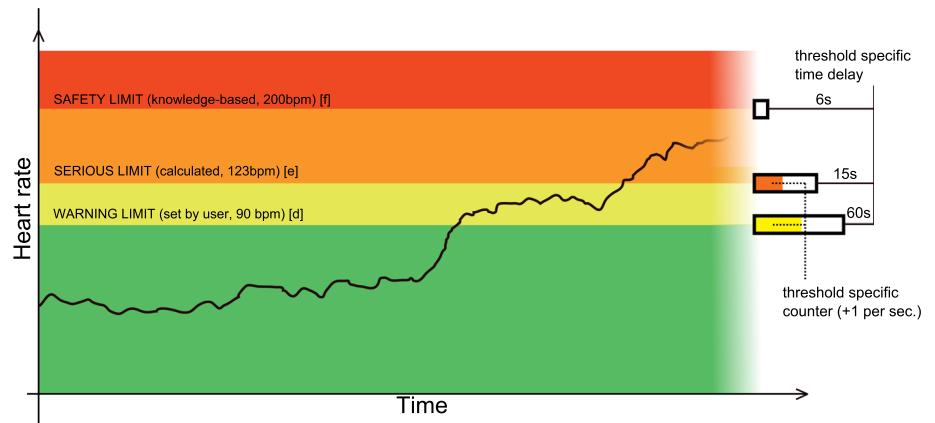
### 3.2 Missed alarms

The new algorithm missed 13 (2.02 %) of the 645 clinically relevant events. Parameters with the highest degree of missed alarms are oxygen saturation (7.14 %) and systolic blood pressure (4.27 %). In contrast to systolic blood pressure the mean arterial blood pressure only showed 0.32 % missed alarms. None of the body temperature and central venous pressure alarms was missed by the new algorithm. The distribution of all missed alarms to different parameters is shown in Table 5.

## 4 Discussion

The implementation of a novel, adaptive alarm delay working as a “soft threshold” reduced the number of false positive alarms by 73.5 %. However, the new algorithm missed 13 of 645 clinically relevant events (2.02 %).

**Fig. 1** Taking the upper heart rate violation as an example, the figure illustrates the functionality of the new algorithm approach that uses time delays and threshold specific counters for alarm generation



**Table 5** Overall results after the analysis of both alarm generation algorithms regarding their alarming performance

	Clinically relevant conventional alarms [a] (%)	Clinically irrelevant conventional alarms [b] (%)	All conventional alarms (%)
All threshold violations	1378 (28.2)	3515 (71.8)	4893 (100.0)
Mild threshold violations (<4 % beyond threshold)	458 (16.7)	2282 (83.3)	2740 (56.9)
Brief threshold violations (duration <16 s)	352 (14.5)	2119 (85.5)	2471 (50.5)
Brief and mild threshold violations	203 (11.5)	1555 (88.5)	1758 (35.9)

The situation of great numbers of false positive alarms that often even exceed the numbers of correct alarms is a frequently described but still unsolved problem in complex perioperative settings, as well as in the ICU. False alarm rates up to 94 % are described for both the OR and the ICU monitoring. Several approaches have been described for the reduction of these high rates of false alarms using median filters, statistic-based thresholds, or time delays in monitoring software. [6–8].

The analysis of our database showed that 96 % of all patient related alarms were caused by threshold violations (remaining 4 % are arrhythmia alarms). The majority of false positive alarms were characterized by mild threshold violations of short duration, i.e. 1–20 s, as illustrated in Fig. 2.

Görges et al. introduced a 14 s delay for monitoring and ventilation alarms in a medical ICU to reduce alarms caused by artifacts and threshold violations of only short duration. This reduced the rate of false alarms by 50 %. A delay of 19 s reduced ineffective alarms by 67 % [11]. Simple delays as used by Görges are nowadays part of several available patient monitors.

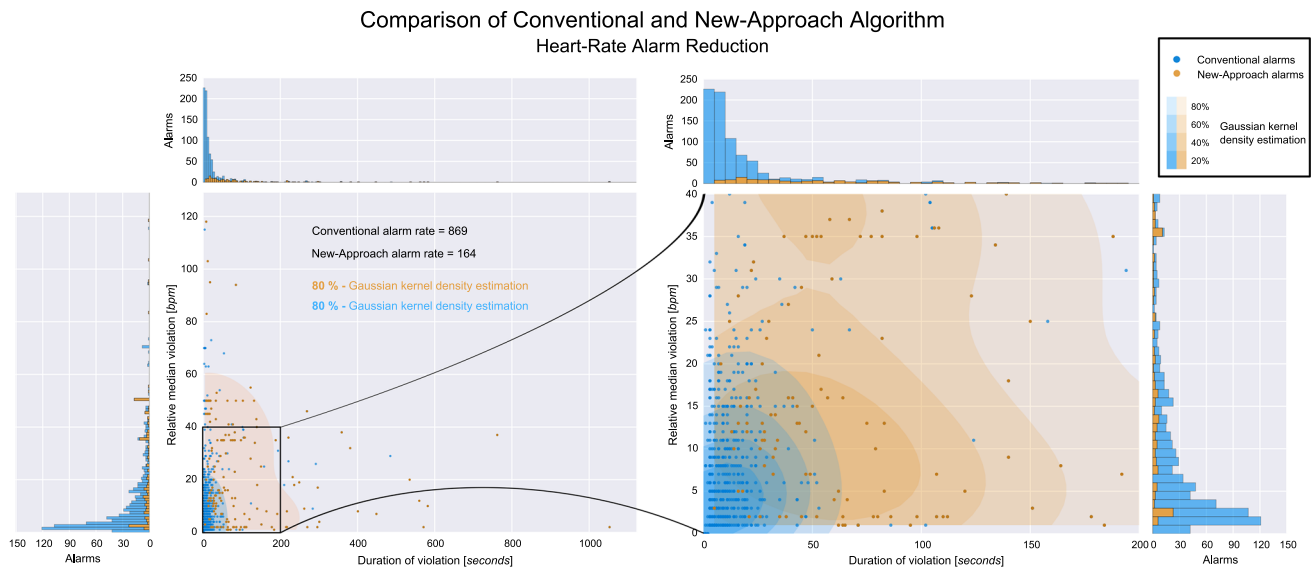
But a simple time delay keeps the problem of protraction of severe problems that need fast interventions. Soft thresholds bring additional safety and flexibility to that approach. First, severe deviations were alarmed faster because a simple delay does not distinguish mild and severe deviations. While a severe heart rate violation is alarmed after 6 s using soft thresholds, a simple delay waits

14–19 s before an alarm is generated. This may result in improved patient safety when using soft thresholds but needs to be proven in further studies. Second, the adaptability offers the possibility of a prolonged delay (more than 14 s) in cases of only moderate and thus clinically irrelevant deviations. This keeps the advantage of further reduction of clinically irrelevant alarms.

Additionally, this new approach is of very low complexity and needs no further user-training, because the user has, like in the conventional set up, to determine only one threshold, and now the proposed novel algorithm adds the two others. This may support high acceptance, because it is easy to use and results in decreased annoyance and lower workload for the attending staff [9, 10]. Users prefer devices that are easy to use and easy to understand in every situation. Many developed approaches have never been introduced to clinical practice because of missing user acceptance [11].

The demonstrated algorithm uses several thresholds and categories (serious, warning, life-threatening) that were developed on a consensus based discussion of four experienced anesthesiologists according to clinical reality. However these values remain subjective and may be not suitable for all kinds of patients.

When analyzing reduction rates for each individual alarm category, reduction rates ranging from 33 % for body temperature to 86.8 % for heart rate alarms were found. The low rate of 33 % in body temperature alarms seemed to be of secondary importance for two reasons. First, it was a rare event compared to the others. Second, and more relevant,



**Fig. 2** Left overview of all generated conventional (*blue dots*) and adaptive (*orange dots*) heart rate threshold alarms. In the plot the duration (start until end of threshold violation) depending on the relative median violation of the threshold is shown. The histograms compare the absolute number of alarms generated within specific ranges of duration and relative median violation between both algorithms. Most of the conventional heart rate alarms generated due

to small and brief violations were eliminated by the new algorithm approach. 80 % of all alarms are located in the *blue* or *orange* region respectively fitted by a two dimensional Gaussian kernel density estimation. The magnification of the integrated inserted box of heart rate threshold alarms with small violation and brief duration with a finer granularity of the density estimation is shown *right*

body temperature alarms are seldom caused by artifacts or other events of short duration. Regarding heart rate and blood pressure being the parameters with the greatest number of alarms in this observation, the new approach was able to reduce the alarms by 86.75 %. The effect is visible in Fig. 2. This might be due to the fact that these kinds of alarms are often caused by artifacts or short physiological changes like fluctuations of blood pressure with respiratory cycle [4]. Another cause could be the selected population of cardiac surgery patients. In offline simulations, the new algorithm did not alarm 13 of the 645 events that the annotators marked as clinically relevant. So these alarms could be assumed as actionable alarms. These 13 events corresponded to nine unique situations (e.g. a single blood pressure increase event may cause a violation of thresholds for both systolic and mean arterial pressure at the same time point). In all missed events the values returned into the set threshold a few seconds before the violation counter exceeded the corresponding alarm delay. Since this was an offline simulation on pre-existing values generated by a conventional alarm system we assume that the attending anesthesiologist would have treated the underlying situation before the violation counter exceeded the corresponding alarm delay. However, formally this data show apparent 13 false negative alarms. Due to the limitation of an offline simulation on a data base of 25 patients it is not possible to conclude outcome differences.

All annotating anesthesiologists were from the same university hospital. This could have led to a reduced

generalizability due to geographic or institution-specific practice patterns. Additionally, the masking of OR anesthesiologists does not allow objectivity in the evaluation of answering intent of the medical staff to occurring alarms, thus it is only possible to evaluate actions to alarms. This algorithm is tailored to reduce alarms that are caused by mild threshold violations of short duration. An alarm reduction by this approach would have several advantages. First, the working environment in the OR and on the ICU would become more silent. This may lead to reduced stress for the medical staff and their patients [12–14]. Reduced noise levels would allow better communication especially in emergency situations. Second, a reduction of errors caused by desensitization (“alarm fatigue”) is expected, but to our knowledge not reported in medical or technical literature. However, we cannot exclude the possibility of missed correct alarms with an analysis based on this pre-existing database. This requires a further, prospective evaluation, in the same clinical setting with a backup of conventional monitoring observed by an additional anesthesiologist to ensure patient safety.

Our study identified parameters like “CVP”, where alarms were reduced by 80 %, without missing any critical situation. On the other hand, the alarm rate of the parameter “diastolic blood pressure” was only reduced by 33 %, but resulted in two false negative alarms. Therefore, this tool of an adaptive alarm delay may not be ideal for every measured parameter.

## 5 Conclusion

The problem of false alarms in anesthesiology and intensive care medicine has been known for decades. False alarms cause disruption in clinical workflow, cause cognitive overload in clinicians, and lead to alarm fatigue. The majority of alarms without therapeutic consequence are triggered by mild threshold violations or violations of short duration. The presented algorithm, which implements adaptive violation delays for vital signs alarms, led to an overall reduction of false positive alarms by 73.5 %. The positive predictive value of occurring alarms improved from 28.16 (conventional algorithm) to 46.15 % (new algorithm). The safety of this new approach needs to be proven in a prospective study.

**Acknowledgments** This study was supported by an unrestricted Grant by Dräger & Co. KGaA Lübeck, Germany.

### Compliance with ethical standards

**Conflict of interest** Frank Franz and David Laule are employees of the Research Unit within Drägerwerk AG & Co. KGaA, which has affiliated companies selling products (e.g., patient monitors) proposed in the paper would be an advantageous feature which may have financial effects. All other authors report no conflict of interest.

## References

1. Seagull F, Sanderson P. Anesthesia alarms in context: an observational study. *Hum Factors*. 2001;43:66–78.
2. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med*. 1994;22:981–5.
3. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *Am J Crit Care*. 2010;19(1):28–34.
4. Takla G, Petre JH, Doyle DJ, Horibe M, Gopakumaran B. The problem of artifacts in patient monitor data during surgery: a clinical and methodological review. *Anesth Analg*. 2006;103(5):1196–204.
5. Schmid F, Goepfert MS, Kuhnt D, Eichhorn V, Diedrichs S, Reichenspurner H, Goetz AE, Reuter DA. The Wolf is crying in the OR: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. *Anesth Analg*. 2011;112(1):78–83.
6. Mäkivirta A, Koski E, Kari A, Sukuvaara T. The median filter as a pre-processor for a patient monitor limit alarm system in intensive care. *Comput Methods Programs Biomed*. 1991;34:139–44.
7. Connor CW, Gohil B, Harrison MJ. Triggering of systolic arterial pressure alarms using statistics-based versus threshold alarms. *Anaesthesia*. 2009;64:131–5.
8. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesth Analg*. 2009;108:1546–52.
9. Bentley S, Murphy F, Dudley H. Perceived noise in surgical wards and an intensive care area: an objective analysis. *BMJ*. 1977;2:1503–6.
10. Bell B. Monitor alarm fatigue. *Am J Crit Care*. 2010;19(1):38.
11. Schmid F, Goepfert M, Reuter DA. Patient monitoring alarms in the ICU and in the operating room. *Crit Care*. 2013;17:216.
12. Topf M, Dillon E. Noise-induced stress as a predictor of burnout in critical care nurses. *Heart Lung*. 1988;17:567–74.
13. Minckley BB. A study of noise and its relationship to patients discomfort in the recovery room. *Nurs Res*. 1968;17:247–50.
14. Hagerman I, Rasmanis G, Blomkvist V, Ulrich R, Eriksen CA, Theorell T. Influence of intensive coronary care acoustics on the quality of care and physiological state of patients. *Int J Cardiol*. 2005;98:267–70.