

The proportion of clinically relevant alarms decreases as patient clinical severity decreases in intensive care units: a pilot study

Ryota Inokuchi,¹ Hajime Sato,² Yuko Nanjo,¹ Masahiro Echigo,³ Aoi Tanaka,¹ Takeshi Ishii,¹ Takehiro Matsubara,¹ Kent Doi,¹ Masataka Gunshin,¹ Takahiro Hiruma,¹ Kensuke Nakamura,¹ Kazuaki Shinohara,⁴ Yoichi Kitsu,¹ Susumu Nakajima,¹ Mitsuo Umezu,³ Naoki Yahagi¹

To cite: Inokuchi R, Sato H, Nanjo Y, et al. The proportion of clinically relevant alarms decreases as patient clinical severity decreases in intensive care units: a pilot study. *BMJ Open* 2013;3:e003354. doi:10.1136/bmjopen-2013-003354

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-003354>).

Received 7 June 2013
Revised 24 July 2013
Accepted 30 July 2013

For numbered affiliations see end of article.

Correspondence to
Dr Hajime Sato;
hsato-tky@umin.ac.jp

ABSTRACT

Objectives: To determine (1) the proportion and number of clinically relevant alarms based on the type of monitoring device; (2) whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms and to suggest; (3) methods for reducing clinically irrelevant alarms in an intensive care unit (ICU).

Design: A prospective, observational clinical study.
Setting: A medical ICU at the University of Tokyo Hospital in Tokyo, Japan.

Participants: All patients who were admitted directly to the ICU, aged ≥ 18 years, and not refused active treatment were registered between January and February 2012.

Methods: The alarms, alarm settings, alarm messages, waveforms and video recordings were acquired in real time and saved continuously. All alarms were annotated with respect to technical and clinical validity.

Results: 18 ICU patients were monitored. During 2697 patient-monitored hours, 11 591 alarms were annotated. Only 740 (6.4%) alarms were considered to be clinically relevant. The monitoring devices that triggered alarms the most often were the direct measurement of arterial pressure (33.5%), oxygen saturation (24.2%), and electrocardiogram (22.9%). The numbers of relevant alarms were 12.4% (direct measurement of arterial pressure), 2.4% (oxygen saturation) and 5.3% (electrocardiogram). Positive correlations were established between patient clinical severities and the proportion of relevant alarms. The total number of irrelevant alarms could be reduced by 21.4% by evaluating their technical relevance.

Conclusions: We demonstrated that (1) the types of devices that alarm the most frequently were direct measurements of arterial pressure, oxygen saturation and ECG, and most of those alarms were not clinically relevant; (2) the proportion of clinically relevant alarms decreased as the patients' status improved and (3) the irrelevance alarms can be considerably reduced by evaluating their technical relevance.

ARTICLE SUMMARY

Strengths and limitations of this study

- We evaluated the technical and clinical relevance of each alarm by using 24 h video monitoring. This technique reduced bias introduced by bedside evaluations.
- This study was limited by the small sample size (18 patients, total).

BACKGROUND

In an intensive care unit (ICU) setting, a large number of medical devices are attached to patients, generating numerous alarm signals every day. Several studies have demonstrated that most of these alarms are not clinically relevant¹⁻³ and tend to lower the attentiveness of the medical staff and, in turn, lower patient safety.^{4 5} In addition, alarm sounds are associated not only with patient delirium,⁶⁻¹⁰ which increases mortality,¹¹ but also with medical staff memory and judgement disturbances, decreased sensitivity and exhaustion.^{6 7} Many attempts have been made to reduce the number of clinically meaningless alarms by using statistical methods and artificial intelligence systems.^{5 12} Some examples include extending the time between the incident and the sounding of the alarm, shutting off alarms prior to performing procedures on patients, and calibrating machines to detect gradual changes in the patient condition. However, alarm devices having high sensitivity and specificity have not been developed because discrepancies remain between the priorities of equipment manufacturers, who are seeking devices with high sensitivity, and those of medical professionals, who desire machines with high specificity.

Previous studies have demonstrated that of the three types of alarms—threshold alarms, arrhythmia alarms and technical alarms—clinical relevance is the lowest for threshold alarms.¹³ However, the impact of patient clinical severity on the proportion of clinically relevant alarms remains unknown. Our objectives were (1) to determine if the number and proportion of clinically relevant alarms differ based on the type of monitoring device; (2) to determine whether patient clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of clinically relevant alarms and (3) to suggest methods for reducing clinically irrelevant alarms. To answer these questions, we used video monitors to collect 24 h continuous data from ICU patients.

MATERIALS AND METHODS

Study setting and patient population

This study was conducted in a 6-bed, mixed ICU at the University of Tokyo Hospital, where patients are mainly admitted following ambulance transport. The study ICU is organised in an 'I' shape, with two individual patient rooms on the west side and two double patient rooms on the east side, with a central monitoring station. The doors to the patient rooms are left open unless procedures are being performed or privacy is required. The unit is staffed with one nurse for every two patients. Most patients monitored during the study had sepsis, respiratory failure, acute respiratory distress syndrome, multisystem organ failure, renal failure, heart failure or trauma.

The following inclusion criteria were used to enrol patients in the study: (1) admitted directly to the University of Tokyo Hospital mixed ICU, not stepped-down from other ICUs and (2) age ≥ 18 years. Patients were excluded if they were (1) already admitted to this ICU or (2) the patient refused active treatment. This study was approved by the Ethics Committee of the University of Tokyo Hospital, and all patients or their family provided signed informed consent before the beginning of the recordings.

Data collection

General patient information, such as age, gender and disease, was recorded. All patients were continuously videotaped using a network of cameras (JVC-Kenwood, V.NET@Web, Tokyo, Japan), attached to the ceiling above each bed, to record patient and/or system manipulations. Each patient was monitored for heart rate, invasive or closely monitored non-invasive arterial blood pressure, respiratory rate, oxygen saturation (SpO₂), end-tidal carbon dioxide (ETCO₂) and temperature. In addition, any changes in the equipment used for each patient were recorded throughout the study period. In addition, the acute physiology and chronic health evaluation (APACHE II) score¹⁴ was calculated for each patient within 24 h of admission, and the SOFA score¹⁵ was calculated every 8 h. Patient data were

pseudonymised and the electronic files and videos were stored in locked, encrypted hard drives.

Alarm systems and settings

During the study period, all patients were monitored with a standard cardiovascular monitoring system (BSM-9101 & CNS-9701, Nihon Koden, Tokyo, Japan). The numerical measurements, waveforms, alarms, alarm settings and alarm messages were acquired in real time and saved continuously (CNS-9600 & CAP-2100, Nihon Koden). The alarm information consisted of the parameter causing the alarm and the alarm message (table 1). The alarm messages were divided into three types: threshold alarms, arrhythmia alarms and technical alarms. The technical alarms indicated technical problems, such as a disconnected probe.

The initial alarm limits and every modification of these during the observation period were registered with corresponding time stamps and automatically recorded (CNS-9600 & CAP-2100, Nihon Koden). Chambrin *et al*¹ determined the initial limits for heart rate and systolic arterial pressure by using the rule, 'initial value observed during a stable period $\pm 30\%$ '. This rule was used in this study as well. When the prehospital patient heart rates and arterial pressures were not obtained, the initial limits were 156/56 mm Hg (120/80 $\pm 30\%$) for systolic arterial pressure/diastolic pressure and 78 and 43 bpm (60 $\pm 30\%$) for upper and lower heart rate limits, respectively. In addition, the SpO₂ limit was 93%, except for patients with chronic obstructive pulmonary disease or acute respiratory distress syndrome, where the limit was 90%; a temperature limit of 38.3°C was also used. After these initial settings, the alarm limits could be modified; any changes were automatically recorded.

Technical annotations

After completion of the data collection for a particular patient, two nurses and two intensivists, with at least 6 years' experience in intensive care medicine, annotated the data. The two nurses first analysed the technical validity of the alarms, and divided the alarms into three categories, *technically true*, *technically false* and *indeterminable*. They referred to the multimonitoring wave shapes or pulse rate when the monitor described alarm messages, rather than using the video record. Alarms were classified as *technically false*, unnecessary alarms if the monitor referred to other waveforms or pulse rates at the same time.

The classifications were defined, in detail, according to the following criteria. For ECG, SpO₂, direct measurements of arterial pressure and ETCO₂, if the waveform was obviously an artefact produced by movements or procedures, the alarm was determined to be *technically false*. For waveforms in which the origin of the artefact(s) or arrhythmia(s) was uncertain, other waveforms or pulse rates (eg, a direct measurement of arterial pressure (ART) or SpO₂) at the time of alarm generation were also referenced. Alarms that did not meet any of

Table 1 The alarm information consisted of the parameter causing the alarm and the alarm message

| Devices | Threshold alarm | Arrhythmia alarm | Technical alarm |
|--|--|---|---|
| ECG | Bradycardia Tachycardia | Asystole ST (II) change Ventricular fibrillation Ventricular tachycardia Ventricular premature contraction run | Check electrodes cannot analyse |
| Oxygen saturation (SpO ₂) | SpO ₂ | | Not connected Check probe Check probe site Cannot detect pulse |
| Direct measurement of arterial pressure (ART) | ART (systolic) ART (diastolic) ART (mean) | | Not connected Check sensor Check label |
| Non-invasive blood pressure (NIBP) | NIBP (systolic) NIBP (diastolic) NIBP (mean) | | Cuff occlusion Not connected Module failure Measurement-out Cannot detect pulse |
| Capnometer | ETCO ₂ CO ₂ (APNEA) | | Not connected Check sensor |
| Thermometer | Tblad T2 | | Not connected Check sensor |
| Central venous pressure monitor Ventilator Other | VENT | | Check sensor Check sensor System failure |

ETCO₂, end-tidal carbon dioxide; Tblad, bladder temperature.

the above criteria were considered *technically true*. All technical evaluations that could not be determined from the relevant monitor's waveform recording were defined as *indeterminable*. For temperature alarms, all upper and lower limits of the temperature alarms were defined as *technically true*. Finally, for non-invasive blood pressure (NIBP) determinations, if an apparently abnormal value was obtained for the NIBP measurement, the patient's movements and concurrent procedures were also considered. Other values, for example, ART or SpO₂ were also referenced as they may have triggered the upper and lower limit alarms. In such instances, these alarms were considered *technically false*.

Clinical annotations

After the technical analyses, the two physicians divided the alarms into three types. These types were relevant alarms, helpful alarms that were not relevant and irrelevant alarms; these were classified by referring to the video and medical records. In this study, an alarm was defined as relevant when an immediate clinical examination plus diagnostic or therapeutic decision (eg, ECG, echocardiography or drug administration) were necessary. When the situation required clinical examination but did not require a diagnostic or therapeutic decision, it was classified as a helpful alarm but not relevant.

Intensivists determining the clinical relevance could see the result of technical validity.

Statistical analyses

All included patient characteristics were described using means and SDs for continuous variables, along with medians and ranges. After obtaining the descriptive statistics regarding the alarm counts and their proportions, the bivariate relationship of the alarms (the total number of alarms and the proportions of relevant alarms) to patient (SOFA) scores was examined by fitting cross-sectional, time-series models for panel data. Alarms from different monitoring devices were examined separately and together. In a preliminary analysis, the numbers and proportions of alarm types were regressed against SOFA scores by fitting either fixed-effects or random-effects models, using the Hausman test. The Hausman test indicated that the random-effects estimates were consistently more appropriate than the fixed-effects estimates.¹⁶ Therefore, the results obtained by the random-effects model were adopted. The interpretation of the statistical significance of relationships was made following multiple comparisons using the Bonferroni method.¹⁷ The NIBP data were not suited for univariate analysis because the amount of data and statistical power were inadequate.

Table 2 Study population baseline characteristics

| Subject description (n=18) | Mean±SD | | |
|--|--------------------|---------------|---------------|
| Age | 69.2±14.0 | | |
| Male/female | 10/8 (55.6%/44.4%) | | |
| | | ICU admission | ICU discharge |
| APACHE score | 18.5±8.3 | | |
| SOFA score | 6.2±3.8 | | 4.1±3.2 |
| The equipment rate of monitoring devices | | | |
| Direct measurement of arterial pressure (%) | 77.8 | | 33.3 |
| Electrocardiogram (%) | 100 | | 100 |
| Oxygen saturation (%) | 100 | | 100 |
| End-tidal CO ₂ (ETCO ₂) (%) | 61.1 | | 44.4 |
| Bladder temperature (%) | 100 | | 94.4 |
| Indirect blood pressure measurement (%) | 100 | | 100 |

APACHE, acute physiology and chronic health evaluation; SOFA, sequential organ failure assessment

The intraobserver and interobserver variabilities between the two physicians performing the clinical annotations of alarms, and the two nurses performing the technical annotations of the alarms were judged by a κ test.¹⁸ To evaluate the intraobserver variability, 300 alarm situations were reannotated by the same observer after a period of approximately 6 months. Statistical analyses were conducted using STATA Special Edition V.12.1 (StataCorp, College Station, Texas, USA).

RESULTS

Patient characteristics

Between January and February 2012, a total of 15 229 alarms were recorded for 20 patients. Two patients were excluded because of their poor clinical condition at the time of admission and of their families' lack of expected benefit from invasive treatment. Therefore, a total of 11 591 alarms for 18 patients were included in this study, corresponding to 2697 person-monitored hours. The observation time for the cases averaged 150±113 h. **Table 2** describes patient characteristics on admission. During their treatment in the ICU, 66.7% of the patients improved (SOFA scores decreased), while 22.2% deteriorated (SOFA scores increased). The ECG, SpO₂ and NIBP devices were attached to all ICU patients throughout their time in the ICU.

The interobserver variabilities in the technical and clinical annotations, as estimated by the κ coefficient, were 0.98 and 0.68. Similarly, the intraobserver validities were

1.95 and 0.73. These values are within the range of substantial (0.61–0.80) or almost perfect (0.81–1.00) agreement.

In addition, false-negative situations were not recorded during the 2697 patient-monitored hours.

Alarm classifications

A total of 11 591 alarms were included in the analysis, classified as *technically true* (71%), *technically false* (21.4%) and *indeterminable* (7.7%) alarms (**figure 1** and **table 3**). The overall contribution of each alarm type to the 11 591 alarms is shown in **table 3**. Only 6.4% of all alarms were relevant, whereas 32.8% were helpful alarms but not relevant, and 60.8% of all alarms were irrelevant. During an 8 h shift, on average, ICU nurses would hear a total of approximately 32 alarms, of which only two were relevant.

The monitoring devices that triggered alarms the most often were ART (33.5%), SpO₂ (24.2%) and ECG (22.9%; **figure 2**). The numbers of relevant alarms were 12.4% (ART), 2.4% (SpO₂) and 5.3% (ECG).

Effect of patient status on the alarms

The results of the cross-sectional time-series analysis are shown in **table 4**. ART demonstrated a positive correlation between the SOFA score and the proportion of relevant alarms, as well as between the SOFA score and the total number of alarms, and also between the SOFA score and the total number of relevant alarms. The SpO₂ and ECG monitors demonstrated positive

Figure 1 Technical and clinical annotations. After an evaluation of the technical relevance was made by two nurses, an evaluation of clinical relevance was made by two intensivists.

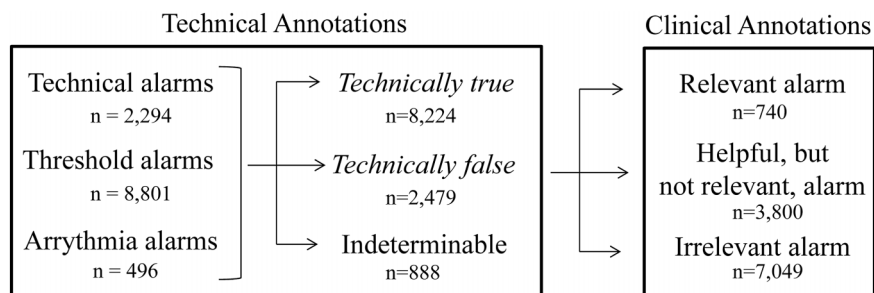


Table 3 The total number of all alarms and the number occurring every 8 h

| Alarms (/overall period: 2697 patient-monitored hours) | n | Per cent of total |
|--|-----------|-------------------|
| Total numbers | 11 591 | |
| Technical annotation | | |
| <i>Technically true</i> | 8224 | 71.0 |
| <i>Technically false</i> | 2479 | 21.4 |
| <i>Indeterminable</i> | 888 | 7.7 |
| Clinical annotation | | |
| Relevant alarm | 740 | 6.4 |
| Helpful, but not relevant, alarm | 3800 | 32.8 |
| Irrelevant alarm | 7049 | 60.8 |
| Indeterminable | 2 | 0.02 |
| Alarms (count/8 h) | Mean±SD | Median (ranges) |
| Total numbers | 31.8±28.6 | 23.5 (1–200) |
| Relevant alarm | 2.0±7.7 | 0 (0–60) |
| Helpful, but not relevant, alarm | 10.4±13.3 | 6 (0–178) |
| Irrelevant alarm | 19.4±20.9 | 13.5 (0–96) |
| Indeterminable | 0.005±0.1 | 0 (0–2) |

correlations only between the SOFA score and the proportion of relevant alarms.

All the devices demonstrated that the SOFA scores had statistically significant positive coefficients when regressed against the total number of relevant alarms ($p < 0.0001$), as well as against the total number of alarms ($p = 0.0061$) and the proportion of relevant alarms ($p < 0.0001$). The results indicated that as the SOFA score decreased, the number of alarms, the number of relevant alarms and the proportion of relevant alarms decreased; the converse was also true.

The inclusion of a regression variable that indicated whether an event occurred during a day or night shift, in the time-series model, indicated that the time of the alarm did not demonstrate a statistically significant relationship with the SOFA score.

Technical validity

Relevant alarms comprised those that were *technically true* and those that were *indeterminable*, but did not include those that were *technically false*. Thus, the irrelevant alarms could be reduced by 21.4% by evaluating their technical relevance.

DISCUSSION

General statement

ICU patients are surrounded by medical devices that regularly sound alarms, but most of the alarms are not clinically relevant.^{1–3} These irrelevant alarms cause a lower quality of patient care by distracting the medical staff^{4–7} and contributing to patient delirium.^{9, 10} Thus, attempts to reduce the number of clinically irrelevant alarms are important as solutions for this national problem are sought.¹⁹ The present study demonstrated that (1) the devices that alarm the most frequently are ART, SpO₂ and ECG; (2) the proportion of relevant alarms decreases as patient status improves and (3) the

irrelevant alarms can be reduced by combining the data for the waveforms or pulse rates of each device.

Prior to this study, Siebig *et al*¹³ were the first to record data with a 24 h video monitor, with the help of two physicians, to evaluate the clinical relevance of alarms. This technique reduced the possible bias introduced by bedside evaluations. The same method of evaluation was used in this study, with the added evaluation of alarm frequency for each device, and the determination of the fluctuations in alarm relevance and clinical severity for individual patients.

Alarm types and their relevance

The vast majority of alarms triggered in the ICU is either false alarms or are irrelevant for patient treatment. The present study shows that only 6.4% of all

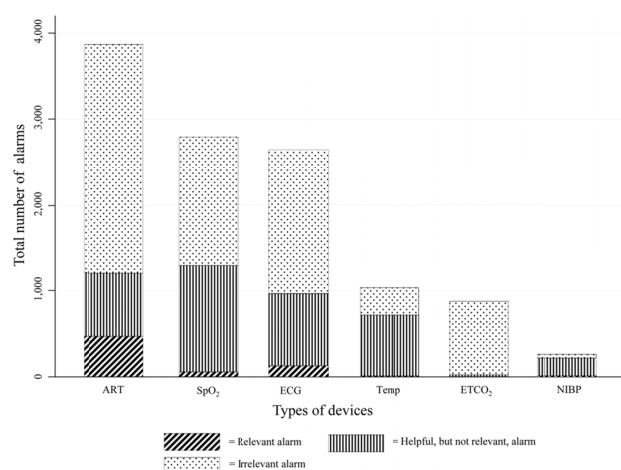


Figure 2 The numbers and types of different alarms. The monitoring devices that triggered alarms the most often were the ART, ECG and SpO₂ monitors. ART, direct measurement of arterial pressure; SpO₂, oxygen saturation; Temp; bladder temperature; ETCO₂, end-tidal carbon dioxide; NIBP, non-invasive blood pressure.

Table 4 Relationship of patient condition with alarm numbers and relevance

| Alarm types | Regression coefficients of severity score (SOFA) ^{†‡} | | | | | |
|---|--|---------|---------------------------------|----------|-------------------------------|----------|
| | Total number of alarms | p Value | Total number of relevant alarms | p Value | Percentage of relevant alarms | p Value |
| Direct measurement of arterial pressure | 1.8±0.5 | 0.0001* | 0.6±0.2 | <0.0001* | 2.2±0.6 | 0.0003* |
| Electrocardiogram | -0.4±0.4 | 0.3018 | 0.1±0.1 | 0.066 | 2.4±0.4 | <0.0001* |
| Oxygen saturation | 0.1±0.3 | 0.7191 | 0.05±0.03 | 0.167 | 0.7±0.2 | 0.0018* |
| Bladder temperature | 0.4±0.2 | 0.0166 | 0.002±0.01 | 0.8704 | -0.1±0.4 | 0.7307 |
| End-tidal CO ₂ | -0.02±0.2 | 0.9363 | 0.004±0.004 | 0.4143 | 0.4±0.2 | 0.0726 |

*Attained statistical significance ($p < 0.05$) after the adjustment for multiple comparisons by Bonferroni method.

[†]Only the regression coefficients of severity scores on the (numbers and proportions of) alarms are shown, which were obtained by the cross-sectional time-series analyses (analysis conducted for each kind of alarm).

[‡]Constant terms were included in the random effect models obtained, but they are not shown. SOFA, sequential organ failure assessment.

alarms triggered in the ICU were relevant. These data are similar to the results of multiple prior studies from various institutions, which indicated that approximately 10% of alarms are relevant.^{1-3, 20} The number of alarms that were technically annotated as being *indeterminable* was 7.7%. When the amplitude of waveforms was small or when the arrhythmia indications and noises were mixed, the technical annotations were difficult.

The ART alarms had a positive correlation between the SOFA score and the number and proportion of relevant alarms. In contrast, the SpO₂ and ECG alarms only showed positive correlations between the SOFA score and the number of alarms. These findings indicate that the SpO₂ and ECG alarms sound regardless of the clinical severity. Therefore, the SpO₂ and ECG alarms are the primarily clinically irrelevant alarms, especially in patients with decreasing SOFA scores. However, this study revealed that the ECG and SpO₂ devices were attached to all ICU patients, for safety reasons, from the time of their ICU admission. Therefore, establishing criteria for removing these devices would be difficult.

How can we reduce the noise in the ICU?

We demonstrated that clinically irrelevant alarms were reduced by 21.4% by evaluating their theoretical technical relevance. When evaluating technical relevance, two nurses combined the data for waveforms or pulse rates for each device. After annotation, their intraobserver and interobserver correlations demonstrated almost perfect agreement and the relevant alarms comprised those that were *technically true* and *indeterminable*, but not those that were *technically false*. Thus, manufacturers can decrease the number of *technically false* alarms by combining the data from each device. In particular, the ART monitor is often used in the ICU setting, and a reduction in the number of clinically irrelevant alarms might be possible by combining the ART waveform with the data from the SpO₂ monitor and ECG.

The number of ART monitor alarms and the proportion of relevant alarms that were associated with the patient SOFA scores implied that there should be a

criterion established to remove this device when the SOFA score has decreased to some appropriate level. We found that when the SOFA scores were ≤ 2 , there were no relevant ART alarms. Thus, when the SOFA scores are ≤ 2 and the patient's condition is not likely to change suddenly, the ART device may be removed. As a general rule, if the sensitivity and specificity of a given test are constant, the positive predictive value (PPV) is assumed to increase as the (true) prevalence/incidence becomes higher. According to this rule, if alarms are being triggered constantly, then PPV is higher when the patient illness severity is higher. Thus, as the patient illness severity increases, the number of alarms increases, and these alarms include a large number of relevant alarms. In contrast, as the patient illness severity decreases, the number of alarms decreases, but these alarms include only a small number of relevant alarms. If the significance of medical treatment, measured by the alarms, is constant, the PPV would be more desirably held constant regardless of the patient's condition. Thus, when the patient illness severity is low, an increase in PPV is important, strictly according to the standards of sensitivity and specificity.

Why has this problem not resolved over the past decade?

The most serious problem encountered with these alarms was that although they provided PPVs (relevant alarms/all alarms), their sensitivity and specificity cannot be ascertained. These data cannot be ascertained because the evaluation of false negatives and true negatives are not possible in cases where the monitor does not alarm in clinical practice. Therefore, manufacturers need to produce alarmed devices that have higher sensitivities in order to avoid medical accidents. In this study, we did not detect false-negative situations. According to studies by Tsien³ and Siebig *et al*¹³ the sensitivity of the current alarms is close to 100%. However, their specificity, which is important for medical staff, could not be determined. Another reason for the failure to reduce the number of clinically irrelevant alarms is that physicians may be relatively insensitive to alarm problems because they do not stand by patient beds as often as nurses. Thus, physicians,

nurses, researchers and medical companies need to establish an evidence-based practice model and find a mutually acceptable solution to this matter.

Study limitations

This study has several limitations. The first is that the sample size was small, with only 18 patients. The second limitation is that although a determination could be made regarding whether an alarm was *technically true* or *false*, a strict definition of the clinical annotations was more difficult. There are relevant alarms that require clinical examination, plus diagnostic or therapeutic decision, but this annotation may differ from a definition considered by intensivists. Finally, we did not analyse ventilator and infusion pump alarms, because detailed ventilator alarm messages were not recorded by our system; thus, annotation of their clinical relevance could not be performed. In addition, infusion pump alarms could not connect our system. These irrelevant alarms also need to be decreased,²¹ and should be the subject of a future study.

CONCLUSION

Excessive alarms in clinical settings are linked to lower medical attentiveness and poorer treatment environments. Manufacturers should work to decrease the number of *technically false* alarms by combining waveform data with the device measurement, especially for ART. Physicians should remove ART when patient conditions improve sufficiently and they are not likely to change suddenly.

Author affiliations

¹Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital, Bunkyo-ku, Tokyo, Japan

²Department of Health Policy and Technology Assessment, National Institute of Public Health, Wako, Saitama, Japan

³Cooperative Major in Advanced Biomedical Sciences, Joint Graduate School of Tokyo Women's Medical University and Waseda University, Shinjuku-ku, Tokyo, Japan

⁴Department of Emergency and Critical Care Medicine, Ohta Nishinouchi Hospital, Koriyama, Fukushima, Japan

Acknowledgements The authors are deeply grateful to Yugo Tamura for collecting data, and would like to thank Yohei Hashimoto, Kikuo Furuta and Hiroko Hagiwara for their support. The authors would also like to thank all participating intensive care unit members at the University of Tokyo Hospital for their support.

Contributors RI conceived of the study, RI and HS designed the analysis plan and performed the statistical analyses. RI wrote the first draft of the study, RI, YN, ME, AT, TI, TM, KD, MG, TH, KN, YK, SN and NY contributed to patient management. KS, MU, and NY critically reviewed the manuscript. All authors contributed to the design, interpretation of results and critical revision of the article for intellectually important content.

Funding This work was supported by a Grant-in-Aid for Young Scientists (C) (12710000424), and a Health Labour Sciences Research Grant.

Competing interests None.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The technical appendix, statistical code and dataset are available from the corresponding author at Dryad repository; a permanent, citable and open access home for the dataset will be provided.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

1. Chambrin MC, Ravaux P, Calvelo-Aros D, *et al*. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. *Intensive Care Med* 1999;25:1360–6.
2. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. *Crit Care Med* 1994;22:981–5.
3. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. *Crit Care Med* 1997;25:614–19.
4. Gorges 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.
5. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *Am J Crit Care* 2010;19:28–34.
6. Christensen M. Noise levels in a general intensive care unit: a descriptive study. *Nurs Crit Care* 2007;12:188–97.
7. Kam PC, Kam AC, Thompson JF. Noise pollution in the anaesthetic and intensive care environment. *Anaesthesia* 1994;49:982–6.
8. Kahn DM, Cook TE, Carlisle CC, *et al*. Identification and modification of environmental noise in an ICU setting. *Chest* 1998;114:535–40.
9. Zaal IJ, Spruyt CF, Peelen LM, *et al*. Intensive care unit environment may affect the course of delirium. *Intensive Care Med* 2012;39:481–8.
10. Radtke FM, Heymann A, Franck M, *et al*. How to implement monitoring tools for sedation, pain and delirium in the intensive care unit: an experimental cohort study. *Intensive Care Med* 2012;38:1974–81.
11. Ely EW, Shintani A, Truman B, *et al*. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. *JAMA* 2004;291:1753–62.
12. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. *Anesth Analg* 2006;102:1525–37.
13. Siebig S, Kuhls S, Imhoff M, *et al*. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care—a methodologic framework. *J Crit Care* 2010;25:128–35.
14. Knaus WA, Draper EA, Wagner DP, *et al*. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818–29.
15. Vincent JL, Moreno R, Takala J, *et al*. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22:707–10.
16. Greene W. *Econometric analysis*. 3rd edn. Prentice Hall, 1997.
17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JR Stat Soc* 1995;57:289–300.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
19. Cvach M. Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol* 2012;46:268–77.
20. Koski KJ, Marttila RJ. Transient global amnesia: incidence in an urban population. *Acta Neurol Scand* 1990;81:358–60.
21. Gorges M, Westenskow DR, Markewitz BA. Evaluation of an integrated intensive care unit monitoring display by critical care fellow physicians. *J Clin Monit Comput* 2012;26:429–36.