PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The proportion of clinically relevant alarms decreases as patient clinical severity decreases in intensive care units: A prospective, observational, clinical study
AUTHORS	Inokuchi, Ryota; Sato, Hajime; Nanjo, Yuko; Echigo, Masahiro; Tanaka, Aoi; Ishii, Takeshi; Matsubara, Takehiro; Doi, Kent; Gunshin, Masataka; Hiruma, Takahiro; Nakamura, Kensuke; Shinohara, Kazuaki; Kitsuta, Yoichi; Nakajima, Susumu; Umezu, Mituso; Yahagi, Naoki

VERSION 1 - REVIEW

REVIEWER	Maria M. Cvach DNP, RN, CCRN The Johns Hopkins Hosptial Baltimore, MD 21287 United States of America
REVIEW RETURNED	26-Jun-2013

THE STUDY	I don't think the title of this paper adequately represents your stated objectives. I suggest a that represents your primary and secondary outcome measures in the title. In the abstract, please make it clear that you are not reducing alarms, but suggesting a method which may be used to reduce alarms. You suggest that 28% of alarms can be reduced by evaluating their technical relevance. This is ambiguous and needs be be more clearly stated. You probably mean that the alarms can theoretically be reduced if an algorithm is used to evaluate technical relevance. There are 3 stated aims of this study: 1.Determine whether patient clinical severity affects the proportion of clinically relevant alarms 2. Determine if the proportion and number of clinically relevant alarms differ based on the type of monitoring device. 3. Identify methods for reducing clinically irrelevant alarms. Please address in your methodology, results and discussion, how you measured each of these aims. The study is good, but disorganized and hard to follow because it doesn't carry through on each aim in an organized manner. There are grammatical errors throughout this document. Content can be shorted and/or written more succinctly. For instance: During 2352 person-hours of continuous monitoring can be written as During 2,352 patient-monitored hours. On page 13, line 7 your refer to alarm grade, but never mention what that means. Do you mean alarm type? There are also upper and lowercase as well as sentence structure issues throughout this document. On page 14

	you indicate that you ..."could not see the duration of the sounding alarm"... not clear. Page 16, line 14-17 is unclear. Page 19, line 16-18 has grammatical issues. In the abstract on page 9,line 29 you indicate the study was between Feb and March 2012, yet in the body of the paper, (page 15, line 33) it states the study took place between Jan and Feb, 2012. Which one is correct? I know it is difficult to come up with a definition of clinical relevance, but I think you need to define, as clearly as possible, how you determined clinical relevance. What do you mean by immediate diagnostic or therapeutic decision? How did you determine this? What does it mean that the total number of alarms can be considerably reduced by evaluating their technical relevance? How does this reduce the number of alarms? How can a clinician practically do this? Is this a recommendation for manufacturers of devices or for the clinician? Explain what you mean by this. I believe you are supporting a case for multi-parameter alarms, but this is not clearly stated. There are many more references that could have been included in this paper. See Cvach, M. Monitor Alarm Fatigue: Integrative Review. BI&T 2012 Jul-Aug;46(4):268-77. doi: 10.2345/0899-8205-46.4.268 for an up-to-date reference list of articles published between 2000 and 2011 as well as seminal articles. You do include classic articles (Chambrin, Lawless, Tsien) which is very good.
RESULTS & CONCLUSIONS	Table 3 is unclear and not intuitive. What was this adjusted to? What was the Y intercept? All tables need to be clearly labeled so a reader can interpret them with minimal instructions. On the Alarm relevant figure, label X and Y axis more clearly. Discussion needs to clearly indicate what information you have collected that is different than what is already known or how your information adds to the existing body of knowledge on this topic. In your conclusion, you state that you anticipate the development of an algorithm which evaluates technical relevance. This is the first time I see mention of an algorithm. Need to explain this in the body of the document.
GENERAL COMMENTS	Thank you for allowing me to review this study. I think it is an important study that documents new information about alarms in relation to patient severity (SOFA score) and distiguishes types of devices that alarm most frequently. In addition, this study offers alarm annotation related to technical relevance. I think it will advise the development of better monitoring algorithms. I hope the authors will rewrite in a more organized manner that flows more succinctly with the study aims.

REVIEWER	Matthias Görges, Postdoctoral Fellow University of British Columbia, Vancouver, Canada I have no conflicts of interest pertaining to this work.
REVIEW RETURNED	26-Jun-2013

THE STUDY	The authors describe the work of an observational study identifying the number of false and technically irrelevant alarms in a Japanese

	adult ICU setting. Additionally, they aim to correlate a patient's organ failure severity, as measured by the SOFA score, with the number of alarms. They find that only 6.7% of all alarms are clinically relevant and that the number of clinically relevant alarms decreases with improved patient's health (low SOFA score). General comments: It is interesting to see an alarm evaluation study using a Nihon Kohden patient monitoring system, which I haven't seen in the literature before. It is discouraging to read that the false alarm rate is comparable with values reported in previous studies. And interesting observation is mentioned only briefly, which is that the number of clinically alarms could be used in an algorithm to discontinue its use (or at least include it in the risk/benefit analysis). This would be an interesting area for future work. The authors need to be congratulated in completing an enormous effort (classifying 8000 alarms and reviewing many hours of video) – quite an accomplishment. Finally, what I am missing is the "what do we learn from this" / novelty aspect of the work which you may want to emphasize on more.
RESULTS & CONCLUSIONS	Specific comments: Title: The title is misleading as it implies that a "technical" solution to reduce the number of clinically irrelevant alarms is proposed or evaluated. However, none of this is performed and the author's main evaluation focuses on the correlation and implied causality of more clinically relevant alarm rates in sicker patients. Summary: P.8 L9: Define clinical severity here to be based on the SOFA score. P.8 L22: ART, ECG and SpO2 monitors are not monitors – they are sources for monitoring data. In this case the later one should probably be called PLETH P.8 L35: The meaning of the first sentence is completely unclear to me – are you suggesting that pilot study was preformed? P.8 L43-50: This is a duplicated copy of the limitations statement found two lines below. P.8 L54: This part belongs into the discussion of the manuscript. I assume it was moved here on accident as the manuscript's discussion doesn't list any limitations. Abstract: P.9 L24: patient illness severity sounds strange – maybe reword it? P.9 L38: I am not sure, where the 70% value is coming from. Figure 1 has 7552/8031 technical alarms which are 94%. Background: P.11 L 35: If this is your objective the title should reflect this. (see above) Materials and Methods: P.12 L20: Was the study registered with clinicaltrials.gov? If yes please state the number here. P.12 L40 How do nasogastric tubes and urinary catheters alarm? Also no mentioning of infusion pump alarms is made after this point. Hence, why did you not analyze their alarms. Finally, ventilators can be associated with large numbers of false alarms due to coughing and "fighting" the ventilator – why did you not record the number of these alarms in your data collection system? P13 L6 Did you mean alarm priority instead of alarm grade? P13 L14 Why did you exclude technical alarms? They can be

	frequent and prompt nursing intervention to be resolved. Also technical alarms can be caused by artifacts – e.g. motion or ambient light falling on a pulse oximetry sensor causing “Sensor Off” alarms. P13 L27 Did you set the diastolic pressure alarm threshold to 56? Most alarms are set based on systolic values (or preferably mean arterial pressures). P13 L32 Why did you pick a SpO2 low alarm setting at 95%. The interesting point at the Hemoglobin/Oxygen binding curve is at 92-93%. Also pulse oximeters are known to be influenced highly by motion artifacts, which can be avoided by lowering the alarming threshold or preferably by introducing short alarm delays. Hence, I am wondering how long were your alarm delays were? P14 L52 Did you mean “ordinal” values? Also if you didn’t test for normality it may be better to report median, IQR and ranges for your data instead of Mean and SD. – If you tested for it your analysis is fine but you should report that this is true. P15 L15 How did you correct for multiple use of data. P15 L11 The approach sounds reasonable, however, it exceeds my understanding of statistics to adequately address if this is performed correctly. The paper would benefit from a statistical reviewer’s comments here. Results: P15 L35 You excluded 7 patients with 7216 alarms. Wouldn’t it have been interesting to have less sick patients included in your study – hence not exclude patients with a stay <48hrs? P.16 L17 This is a nice statement as it puts your findings in context for a regular shift length. Did nurses in your ward work 8hr shift or were there also providers on a 12hr shift schedule? Discussion: P.17 L26 What is missing here is the “how?” a monitor automatically evaluates technical relevance. In hindsight this is easy to make and if a clinician is required to make the determination here it may be better to just have them present at the bedside? P.17 L48 Threshold alarms are known to have problems with transient changes and artifacts. P.18 L23 You may want to take a look at Borowski et al, Biomed Tech (Berl). 2011 Apr;56(2):73-83 (PMID 21366502) for some additional suggestion here. P.18 L23-28 This is an interesting and novel suggestion. If might be good to elaborate on it and how you would propose to make device removal algorithms/flowcharts. P.19 L20 Again you make an interesting observation here, which you may want to elaborate on as well. P.19 L23 I am missing a limitations section here – did you accidentally move it to the summary page and remove it here. Also if that is true it should be expanded significantly to cover areas like the removal of 7216 alarms from the analysis, the exclusion of ventilation and infusion pump alarms, and why the number of indeterminable alarms is >10% P.19 L29 The final part of the last sentence of the conclusion is confusing. Please rephrase it. Figures: Figure 1: Your caption is partially redundant with the figure, which may not be a bad thing. However, you may want to use the same definitions throughout. (e.g. “Not alarm relevant” vs. “not relevant” etc.). Finally, some of the wording in the figure is unclear, which is likely as you wanted to reduce the number of words here. I’d

	suggest only showing the key elements in the figure and referring the reader to the appropriate section in the manuscript for details. Figure 2: You are using yet another way of labeling your alarm classification categories. Maybe just use: “Alarm relevant”, “Alarm not relevant” and “Alarm not relevant but helpful” throughout the manuscript? Tables: Table 1: L24 These are variables, not monitors. Also what you are really reporting is the frequency of use. Table 1: L27 Consider removing all variables which were either 100% used (ECG and SpO2) or not used in any patient (CVP). You can mention that you excluded these for simplicity reasons in the table caption or the text. Table 2: L18 Why is the number of indeterminable alarms so high? – Address in discussion. Table 2: L34 If you didn’t test for normality it may be better to report Median, IQR and ranges. Table 3: You need to indicate which p values were significant in your regression model due to the multiple comparisons performed your critical alpha is likely not 0.05?
GENERAL COMMENTS	I would encourage you to revise the focus of the paper and resubmit it to this or another appropriate journal.

VERSION 1 – AUTHOR RESPONSE

Responses to Reviewer Comments

Thank you for your kind letter dated June 28, 2013 and the editors’ e-mail concerning the manuscript entitled “How do we reduce clinically irrelevant alarms in intensive care units? A prospective, observational, clinical study” [manuscript ID: bmjopen-2013-003354]. We are pleased to note the favorable comments of the reviewers and the editor, and have made appropriate corrections in accordance with these comments.

We thank the reviewers for their help suggestions and comments and hope that the revised manuscript is now suitable for publication as an Original Article in BMJ Open.

Our point-by-point responses to the comments of the reviewers are given below.

Reviewer: Maria M. Cvach DNP, RN, CCRN
The Johns Hopkins Hosptial
Baltimore, MD 21287
United States of America

Comment #1:
I don't think the title of this paper adequately represents your stated objectives. I suggest a that represents your primary and secondary outcome measures in the title.

Response #1:
We completely agree with your suggestion. We have revised the title as follows: “The proportion of clinically relevant alarms decrease as patient clinical severity decreases in intensive care units: A

prospective, observational, clinical study."

Comment #2:
In the abstract, please make it clear that you are not reducing alarms, but suggesting a method which may be used to reduce alarms. You suggest that 28% of alarms can be reduced by evaluating their technical relevance. This is ambiguous and needs be be more clearly stated. You probably mean that the alarms can theoretically be reduced if an algorithm is used to evaluate technical relevance.

Response #2:
We completely agree with your suggestion. As you have mentioned, we did not reduce alarm practically, but in theory. We have now stated this clearly throughout our manuscript.

Comment #3:
There are 3 stated aims of this study:
1.Determine whether patient clinical severity affects the proportion of clinically relevant alarms
2. Determine if the proportion and number of clinically relevant alarms differ based on the type of monitoring device.
3. Identify methods for reducing clinically irrelevant alarms.
Please address in your methodology, results and discussion, how you measured each of these aims. The study is good, but disorganized and hard to follow because it doesn't carry through on each aim in an organized manner.

Response #3:
In keeping with your suggestion, we have revised the focus of our article in the summary as well as in the manuscript text.

Comment #4:
There are grammatical errors throughout this document. Content can be shorted and/or written more succinctly. For instance: During 2352 person-hours of continuous monitoring can be written as During 2,352 patient-monitored hours. On page 13, line 7 your refer to alarm grade, but never mention what that means. Do you mean alarm type? There are also upper and lowercase as well as sentence structure issues throughout this document. On page 14 you indicate that you ..."could not see the duration of the sounding alarm"... not clear. Page 16, line 14-17 is unclear. Page 19, line 16-18 has grammatical issues.

Response #4:
We apologize for the grammatical errors in the manuscript. The manuscript has now been carefully edited by a native English speaking editor. On page 13, line 7, we have changed the paragraph and added Table 1. On page 14, we have deleted the above-mentioned sentence. On Page 16, line 14-17, we have revised as follows: "During an 8-hour shift, on average, ICU nurses would hear a total of approximately 32 alarms, but only 2 were relevant." (Page 11, line 19). We have also corrected the sentence on Page 19, line 16-18.

Comment #5:
In the abstract on page 9,line 29 you indicate the study was between Feb and March 2012, yet in the body of the paper, (page 15, line 33) it states the study took place between Jan and Feb, 2012. Which one is correct?

Response #5:
We apologize for our oversight in the study period. The correct study period is January and February 2012.

Comment #6:

I know it is difficult to come up with a definition of clinical relevance, but I think you need to define, as clearly as possible, how you determined clinical relevance. What do you mean by immediate diagnostic or therapeutic decision? How did you determine this?

Response #6:
According to your suggestion, we have added the definition of clinical relevance in paragraph 1 in the Clinical Annotation section.

Comment #7:
What does it mean that the total number of alarms can be considerably reduced by evaluating their technical relevance? How does this reduce the number of alarms? How can a clinician practically do this? Is this a recommendation for manufacturers of devices or for the clinician? Explain what you mean by this. I believe you are supporting a case for multi-parameter alarms, but this is not clearly stated.

Response #7
To clarify the issue, we have added as follows: “We demonstrated that … the ART waveform with the data from the SpO2 monitor and ECG.” in paragraph 1 and “The number of ART monitor alarms and … the patient’s condition is not likely to change suddenly, the ART device may be removed.” in paragraph 3 in HOW CAN WE REDUCE THE NOISE IN THE ICU? Section.

Comment #8:
There are many more references that could have been included in this paper. See Cvach, M. Monitor Alarm Fatigue: Integrative Review. BI&T 2012 Jul-Aug;46(4):268-77. doi: 10.2345/0899-8205-46.4.268 for an up-to-date reference list of articles published between 2000 and 2011 as well as seminal articles. You do include classic articles (Chambrin, Lawless, Tsien) which is very good.

Response #8:
Thank you for these suggestions. We have added this reference and “Graham K , Cvach M. Alarm Fatigue: Standardizing Progressive Care Nurse's Utilization of Physiological Monitors and Decreasing Nuisance Alarms. American Journal of Critical Care Nursing. 2010; 19(1):28–35. ” in the reference list.

Comment #9:
Table 3 is unclear and not intuitive. What was this adjusted to? What was the Y intercept? All tables need to be clearly labeled so a reader can interpret them with minimal instructions. On the Alarm relevant figure, label X and Y axis more clearly.

Response #9:
We have included a new Table 4 that presents the results from the cross-sectional time-series analysis, with the number/ proportions of alarms as dependent variables and the severity score as the predictor. The models consider potential autocorrelations in errors. There was no adjustment made to the other variable (we did that only for nurse shift). The captions and note of the table have been revised for greater clarity.

Comment #10:
Discussion needs to clearly indicate what information you have collected that is different than what is already known or how your information adds to the existing body of knowledge on this topic.

Response #10:
Thank you for pointing this out. We have now revised the manuscript to clearly indicate how our findings add to the existing knowledge on this topic.

Comment #11:
In your conclusion, you state that you anticipate the development of an algorithm which evaluates technical relevance. This is the first time I see mention of an algorithm. Need to explain this in the body of the document.

Response #11:
We have revised our conclusion to make it more simple and clear.

Comment #12:
Thank you for allowing me to review this study. I think it is an important study that documents new information about alarms in relation to patient severity (SOFA score) and distiguishes types of devices that alarm most frequently. In addition, this study offers alarm annotation related to technical relevance. I think it will advise the development of better monitoring algorithms. I hope the authors will rewrite in a more organized manner that flows more succinctly with the study aims.

Response #12:
We thank the reviewer for carefully reviewing our manuscript and for providing suggestions to further improve the quality of our manuscript. We hope that the revised manuscript now meets the expectations of the reviewer.

Reviewer: Matthias Görges, Postdoctoral Fellow
University of British Columbia, Vancouver, Canada

I have no conflicts of interest pertaining to this work.

The authors describe the work of an observational study identifying the number of false and technically irrelevant alarms in a Japanese adult ICU setting. Additionally, they aim to correlate a patient's organ failure severity, as measured by the SOFA score, with the number of alarms. They find that only 6.7% of all alarms are clinically relevant and that the number of clinically relevant alarms decreases with improved patient's health (low SOFA score).

General comments:
It is interesting to see an alarm evaluation study using a Nihon Kohden patient monitoring system, which I haven't seen in the literature before. It is discouraging to read that the false alarm rate is comparable with values reported in previous studies. And interesting observation is mentioned only briefly, which is that the number of clinically alarms could be used in an algorithm to discontinue its use (or at least include it in the risk/benefit analysis). This would be an interesting area for future work. The authors need to be congratulated in completing an enormous effort (classifying 8000 alarms and reviewing many hours of video) – quite an accomplishment. Finally, what I am missing is the "what do we learn from this" / novelty aspect of the work which you may want to emphasize on more.

Response:
We thank the reviewer for carefully reviewing our manuscript and for providing suggestions to further improve the quality of our manuscript. We have revised our manuscript according to the comments of the reviewer.

Comment #1:
Specific comments:
Title: The title is misleading as it implies that a "technical" solution to reduce the number of clinically irrelevant alarms is proposed or evaluated. However, none of this is performed and the author's main evaluation focuses on the correlation and implied causality of more clinically relevant alarm rates in sicker patients.

Response #1:
We completely agree with your suggestion. We have changed the title as follows: “The proportion of clinically relevant alarms decrease as patient clinical severity decreases in intensive care units: A prospective, observational, clinical study.”

Comment #2:
Summary:
P.8 L9: Define clinical severity here to be based on the SOFA score.

Response #2:
In keeping with your suggestion, we have revised “clinical severity affects the proportion of…” to “clinical severity, based on the sequential organ failure assessment (SOFA) score, affects the proportion of…”.

Comment #3:
P.8 L22: ART, ECG and SpO2 monitors are not monitors – they are sources for monitoring data. In this case the later one should probably be called PLETH

Response #3:
We apologize for the confusion. As you have mentioned, ART, ECG and SpO2 are devices, not monitors. We have now made changes to clearly differentiate monitors from devices in the revised manuscript.

Comment #4:
P.8 L35: The meaning of the first sentence is completely unclear to me – are you suggesting that pilot study was preformed?

Response #4:
We are very sorry for the confusion. We have revised the sentence as follows: “We evaluated the technical and clinical relevance of each alarm using 24-h video monitoring. This technique reduced bias introduced by bedside evaluation.”

Comment #5:
P.8 L43-50: This is a duplicated copy of the limitations statement found two lines below.
P.8 L54: This part belongs into the discussion of the manuscript. I assume it was moved here on accident as the manuscript’s discussion doesn’t list any limitations.

Response #5:
We have moved this part to the Discussion section of the manuscript.

Comment #6:
Abstract:
P.9 L24: patient illness severity sounds strange – maybe reword it?

Response #6:
We completely agree with your suggestion. We have revised the title to “patient severity based on the sequential organ failure assessment (SOFA) score”.

Comment #7:
P.9 L38: I am not sure, where the 70% value is coming from. Figure 1 has 7552/8031 technical alarms which are 94%.

Response #7:
We apologize for the confusion. We have corrected this part and have re-examined the manuscript carefully for such errors.

Comment #8:
Background:
P.11 L 35: If this is your objective the title should reflect this. (see above)

Response #8:
We completely agree with your suggestion. Hence, we have revised the title as mentioned in Response #1.

Comment #9:
Materials and Methods:
P.12 L20: Was the study registered with clinicaltrials.gov? If yes please state the number here.

Response #9:
We did not register this study at the website, because it was performed in Japan.

Comment #10:
P.12 L40 How do nasogastric tubes and urinary catheters alarm? Also no mentioning of infusion pump alarms is made after this point. Hence, why did you not analyze their alarms. Finally, ventilators can be associated with large numbers of false alarms due to coughing and “fighting” the ventilator – why did you not record the number of these alarms in your data collection system?

Response #10:
You are right that we did not monitor nasogastric tubes and urinary catheters, which are devices. We have now made changes to clearly differentiate monitors from devices in the revised manuscript. As you have commented0, ventilators can be deeply associated with false alarms. However, our system did not collect ventilators alarm messages in detail; hence, we did not analyze the clinical relevance of ventilator alarms. We have now mentioned this as a study limitation.

Comment #11:
P13 L6 Did you mean alarm priority instead of alarm grade?

Response #11:
We are very sorry for the confusion. We have deleted the phrase “the alarm grade” and have added Table 1 to explain the alarm message in detail.

Comment #12:
P13 L14 Why did you exclude technical alarms? They can be frequent and prompt nursing intervention to be resolved. Also technical alarms can be caused by artifacts – e.g. motion or ambient light falling on a pulse oximetry sensor causing “Sensor Off” alarms.

Response #12:
In keeping with your suggestion, we reanalyzed our date by including technical alarms and reached similar conclusions. We have revised the manuscript accordingly.

Comment #13:
P13 L27 Did you set the diastolic pressure alarm threshold to 56? Most alarms are set based on systolic values (or preferably mean arterial pressures).

Response #13:

We set the diastolic pressure alarm threshold to 56 initially in our study if the patient's normal blood pressure was unknown. However, after these initial settings, the alarm limits could be modified; any changes were automatically recorded.

Comment #14:
P13 L32 Why did you pick a SpO2 low alarm setting at 95%. The interesting point at the Hemoglobin/Oxygen binding curve is at 92-93%. Also pulse oximeters are known to be influenced highly by motion artifacts, which can be avoided by lowering the alarming threshold or preferably by introducing short alarm delays. Hence, I am wondering how long were your alarm delays were?

Response #14:
We apologize for the oversight regarding the SpO2 low alarm setting. The correct SpO2 low alarm setting was 93%. We have now corrected this error and have thoroughly re-examined the manuscript for such errors. With regard to the alarm delay, we consulted the manufacturer Nihon Kohden and received a response that their SpO2 alarm delay is 4 seconds.

Comment #15:
P14 L52 Did you mean "ordinal" values? Also if you didn't test for normality it may be better to report median, IQR and ranges for your data instead of Mean and SD. – If you tested for it your analysis is fine but you should report that this is true.

Response #15:
We apologize for the confusion. The correct term should be "ordinal". However, because there was no "ordinal" variable (hence we did not conduct normality tests), we have removed the description from the text. We have added the median and ranges of some variables in Table 3 for clearer presentation of our data.

Comment #16:
P15 L15 How did you correct for multiple use of data.

Response #16:
Statistical significance attained after Bonferroni adjustment of significance level for multiple comparisons is now indicated in Table 4.

Comment #17:
P15 L11 The approach sounds reasonable, however, it exceeds my understanding of statistics to adequately address if this is performed correctly. The paper would benefit from a statistical reviewer's comments here.

Response #17:
The second and corresponding author is a statistical expert, who has re-checked the analytical methods of the study. The statistical method used in this study has now been cited in the manuscript.

Comment #18:
Results:
P15 L35 You excluded 7 patients with 7216 alarms. Wouldn't it have been interesting to have less sick patients included in your study – hence not exclude patients with a stay <48hrs?

Response #18:
We have reanalyzed our data by including patients with a stay <48 hrs and reached similar conclusions. We have revised the manuscript accordingly.

Comment #19:

P.16 L17 This is a nice statement as it puts your findings in context for a regular shift length. Did nurses in your ward work 8hr shift or were there also providers on a 12hr shift schedule?

Response #19:
The nurses in our hospital have a day shift of 8 hr or a night shift of 16 hr as their work schedule.

Comment #20:
Discussion:
P.17 L26 What is missing here is the “how?” a monitor automatically evaluates technical relevance. In hindsight this is easy to make and if a clinician is required to make the determination here it may be better to just have them present at the bedside?

Response #20:
We thank you for these suggestions. We have changed these throughout the manuscript.

Comment #21:
P.17 L48 Threshold alarms are known to have problems with transient changes and artifacts.

Response #21:
Thank you for pointing this out. We have deleted this sentence.

Comment #22:
P.18 L23 You may want to take a look at Borowski et al, Biomed Tech (Berl). 2011 Apr;56(2):73-83 (PMID 21366502) for some additional suggestion here.
P.18 L23-28 This is an interesting and novel suggestion. If might be good to elaborate on it and how you would propose to make device removal algorithms/flowcharts.

Response #22:
In keeping with your suggestion, we have revised the text: “When the SOFA scores were 2, there were no relevant ART alarms. Thus, when SOFA score are 2 and the patient’s condition is not likely to change suddenly, the ART device may be removed.”

Comment #23:
P.19 L20 Again you make an interesting observation here, which you may want to elaborate on as well.

Response #23:
We have changed “Another reason… this matter.” in paragraph 2 in Why do this problem not resolve over the 10 years? Section.

Comment #24:
P.19 L23 I am missing a limitations section here – did you accidentally move it to the summary page and remove it here. Also if that is true it should be expanded significantly to cover areas like the removal of 7216 alarms from the analysis, the exclusion of ventilation and infusion pump alarms, and why the number of indeterminable alarms is >10%

Response #24:
We have added the text “The number of alarms that were technically annotated as being indeterminable was 7.7%. When the amplitude of waveforms were small or when arrhythmia indications and noises were mixed, technical annotations were difficult.” (page 13, line 12) in ALARM TYPES AND THEIR RELEVANCE section.

In addition, we have added the limitation about the exclusion of ventilator and infusion pump alarms.

As you have mentioned, infusion pump alarms are very important; hence, we have added the reference “J Clin Monit Comput. 2012 Dec;26(6):429-36. doi: 10.1007/s10877-012-9370-0” in the reference list.

Comment #25:
P.19 L29 The final part of the last sentence of the conclusion is confusing. Please rephrase it.
Response #25:
We thank you for these suggestions and apologize for the confusion. We have rephrased the last sentence.

Comment #26:
Figures:
Figure 1: Your caption is partially redundant with the figure, which may not be a bad thing. However, you may want to use the same definitions throughout. (e.g. “Not alarm relevant” vs. “not relevant” etc.). Finally, some of the wording in the figure is unclear, which is likely as you wanted to reduce the number of words here. I’d suggest only showing the key elements in the figure and referring the reader to the appropriate section in the manuscript for details.

Response #26:
Thank you for pointing this out. We have changed the figure according to your suggestion.

Comment #27:
Figure 2: You are using yet another way of labeling your alarm classification categories. Maybe just use: “Alarm relevant”, “Alarm not relevant” and “Alarm not relevant but helpful” throughout the manuscript?

Response #27:
In keeping with your suggestion, we have used another way of labeling. We have changed “alarm relevant, helpful alarm but not alarm relevant, and not alarm relevant” throughout the manuscript.

Comment #28:
Tables:
Table 1: L24 These are variables, not monitors. Also what you are really reporting is the frequency of use.

Response #28:
We are sorry for the confusion. We have made changes to clearly differentiate monitors from devices in the revised manuscript.

Comment #29:
Table 1: L27 Consider removing all variables which were either 100% used (ECG and SpO2) or not used in any patient (CVP). You can mention that you excluded these for simplicity reasons in the table caption or the text.

Response #29:
We have changed the table according to your suggestion.

Comment #30:
Table 2: L18 Why is the number of indeterminable alarms so high? – Address in discussion.

Response:
Thank you for pointing this out. We have addressed this part in the Discussion section.

Comment #31:
Table 2: L34 If you didn’t test for normality it may be better to report Median, IQR and ranges.

Response #31:
We apologize for the confusion. Because there was no "ordinal" variable (hence we did not conduct normality tests), we have deleted the description from the text. In accordance with your suggestion, we have added the median and ranges of some variables in Table 3 for clearer presentation of our data.

Comment #32:
Table 3: You need to indicate which p values were significant in your regression model due to the multiple comparisons performed your critical alpha is likely not 0.05?

Response #32:
Statistical significance attained after Bonferroni adjustment of significance level for multiple comparisons is now indicated in Table 4.