Measuring Alarm System Quality in Intensive Care Units

Work in progress

Dirk Hüske-Kraus
Philips Medical Systems Böblingen, Germany
dirk.hueske-kraus@philips.com

Marc Wilken, Rainer Röhrig
Carl von Ossietzky University, Oldenburg, Germany

Abstract — Despite a vast and growing number of publications on alarm fatigue there seem to be few attempts to quantify in a systematic way the root causes or consequences of alarm fatigue. The paper describes such an attempt, based on the notion of quality dimensions of an alarm system. The concept as well as concrete instances of these dimensions were developed in a scoping review including focus group interviews and subsequent iterative refinement workshops. The resulting model goes beyond currently used statistics for describing alarm system quality and appears to be useful for the construction of tools to support clinical users in the management of their alarm system.

Keywords—alarm fatigue; medical devices; clinical alarms; patient safety; ICU; alarm system quality;

I. INTRODUCTION

Alarms from medical devices are known to be a potential source of inefficiencies and ineffectiveness on an Intensive Care Unit (ICU). Many of them do not require any reaction from the caregiver (are "non-actionable"). Non-actionable alarms also can lead to a desensitization of staff, impacting patient safety via inadequate reactions to alarms [1]. In the literature, quite high numbers of patient deaths are attributed to inadequate reactions to alarms [2]. But also the sheer number of alarms, up to 350 alarms per monitored bed day (AMBD) [3], can have negative impacts, e.g. unnecessary workload, chronic overload and acute cognitive stress. The term "alarm fatigue" (AF) is widely used to denote a condition characterized by a desensitization to alarms due to overload [4],[5]. Further possible impacts of excessive alarms are frequent interruptions of care tasks, disruptions of patients' rest during night and an overall increase of the noise level on an ICU (see fig. 1).

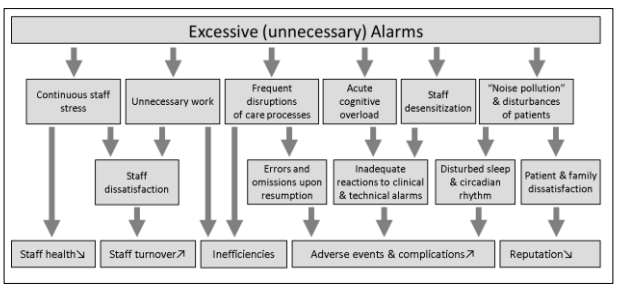

Figure 1: Possible impacts of excessive alarms, modified after [6]

The possible causes of excessive alarms are manifold: They range from an unfortunate floor layout and insufficient monitoring equipment, over inadequate consumables and sensors as well as suboptimal configurations to inefficient processes and workflows, lack of capabilities and competencies as well as learned habits and lack of discipline. Fig.Figure 2 shows an overview of root cause categories for excessive alarms.

Oftentimes, one will find a mix of several of these factors on any given ICU, in particular since some of them are causally connected. For instance, a monitor's configuration which creates many nuisance alarms may lead to a "learned habit" of acknowledging alarms without proper evaluation.

In order to fully apprehend the complex topic of alarm fatigue, its causes and consequences, it is important to broaden the focus from individual caregivers and particular phenomena like missed alarms to the entire sociotechnical system ICU, with staff, devices and consumables, the infrastructure, but also processes, policies and guidelines and, last, but not least, the patients. We will use the term "alarm system" to refer to the sociotechnical subsystem of an ICU that has to do with patient surveillance and alarm generation and handling.

Improving the alarm management of an ICU requires deep insights into which of the aforementioned factors are present and impacting the overall alarm system quality (ASQ).

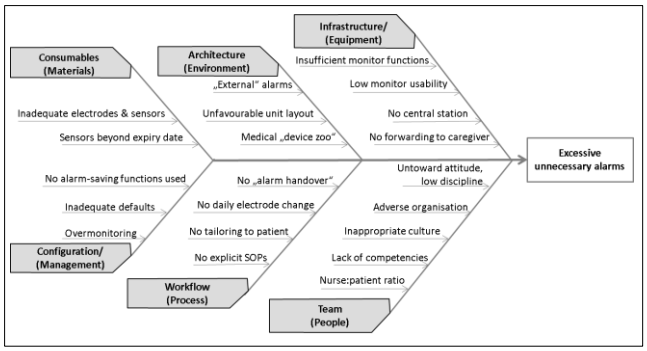

Figure 2:Possible causes of excessive alarms, modified after [6]

II. BACKGROUND

A. The current state

1) Alarms per monitored bed

In the vast literature on Alarm fatigue, there is only one quantitative measure which is almost always given, the number of alarms per monitored bed day (AMBD) (e.g. [5]). However, other than the fact that it is easy to calculate in daily routine, little more that can be said in favor of this metric:

- Other than for extreme values like 5 or 500, the measure does not map simply to a gradient from "excellent" to "acceptable" to "disastrous". Whether a value of 100 AMBD is a sign of high or low alarm system quality is not an easy question to answer.
- The AMBD-measure does neither reflect the care setting (e.g. neonatal vs. pediatric vs. adult ICU), the patient population (e.g. medical vs. surgical, neuro vs. cardiac), nor the safety culture which a hospital might embrace. It thus cannot be readily used to compare quality in a cross-unit or cross-institution fashion.
- But even as a relative measure, where an intervention has brought a unit from 200 down to 150 AMBD, the metric is not particularly usable: If the positive decline in sheer numbers masks the fact that the numbers of critical (or "red"') alarms has doubled, then certainly one should not regard the outcome of the intervention all too positive.
- The metric neither gives an indication whether healthcare professionals can still cope with the number of alarms, nor does it highlight possible root causes for a possibly excessive number of alarms.
- Finally, AMBD is only easy to calculate for the devices which are connected to a central station. While this is routinely the case for patient monitors, ventilators sometimes are and sometimes are not hooked to the central, and perfusion pumps most often are not. Aside from these three main contributors of alarms, there is a wide range of other medical devices, which for technical reasons are almost never routinely covered by alarm data analyses. The varying degree of device types included and the "blind spot" on devices not connected to a central station further limits the usability of AMBD as a means to gauge the quality of an alarm system.

Two other kinds of measures are sporadically used in the literature:

2) Significance

On the one hand side one my find measures of significance, i.e. the specificity or positive predictive value of alarms [7]. The obvious drawback of this metric is that it cannot be calculated routinely: It requires manual annotation, either by the caregivers at the point of care, or by retrospectively evaluating recorded biosignals and/or video recordings.

3) Responsiveness

On the other hand, one may find studies focusing on response times to alarms [8]. Metrics regarding response times (or response rates) certainly convey important information: A decline in responsiveness of clinical staff may indicate acute overload as well as a desensitization to alarms, and improved responsiveness after an intervention is likely to be a positive result of the intervention. The association of responsiveness with alarm load has been empirically demonstrated in video analyses [9]. However, response times just show one facet of the problem and need to be interpreted cautiously:

- Obviously, response times need to considered separately for different sources, types and severities of alarms. Less obviously, the alarm management policy of the institution as well as the physical unit layout may influence measurements negatively, in which case any interpretation as indicator of desensitization would be unjustified.
- Also, changes in the nurse-patient-ratio, for instance during a flu-epidemic, may temporarily skew the measurements.
- Most importantly, only measuring the timeliness of response times cannot gauge the adequacy of the response. Blindly acknowledging alarms without evaluation and taking clinical or technical corrective actions may yield better response times, but the overall reactivity of the alarm system would certainly not improved by such a behavior.

B. Desiderata for alarm management metrics

Metrics to quantify the quality of alarm systems in routine use – e.g. as ongoing quality control or for gauging the effects of certain interventions – need to satisfy certain requirements. In particular, the following properties are desirable:

- Calculability in routine, practicality: Only measures which can be obtained in a low-cost, low-effort manner are likely to be used outside dedicated research projects
- Multidimensionality, content validity: As there are many diverse possible contributors to, and also many different effects of excessive alarms, alarm metrics need to cover as many as possible of these.
- Clinical appropriateness, construct validity: Obviously the different severity levels of alarms, ignored by the AMBD metric, should be accounted for, if the burden of alarms is to be quantified. Also, the temporal course of alarms, i.e. a constant "base rate" vs. "clusters" or "bursts" of alarms should be accounted for, if a metric is not to ignore the possibility of acute overload.
- Face validity: Inasmuch as metrics are used to assess behavior and motivate behavioral change, it is imperative that they "make sense" for clinical front line staff. Only if there is some correlation between the metric and the clinicians' impression from their day-to-

day routine, will said metric be useful to trigger interventions and select the right targets for these.

- Gradients of granularity: Different use cases for ASQ metrics dictate that they need to be calculated for varying areas and intervals of interest: One may want to compare two different units, only look at isolation rooms, contrast day shift with night shift values or focus on surgical patients only. Thus the metrics need to provide meaningful and comparable results for subsets of beds or patients as well as (almost) arbitrary time intervals.
- Actionability: The metrics should enable one to
  - quantify a need for an intervention on the alarm system,
  - assess changes of ASQ over time, in particular after an intervention,
  - highlight presence of well-known possible root causes for low ASQ, and
  - indicate presence of well-known consequences of excessive alarms, like desensitization or acute overload.

C. Intended use of alarm system quality metrics

One of the goals of AlarmRedux, a project funded by the German Federal Ministry of Education and Research (BMBF), is to enable caregivers to manage the quality of their alarm system by themselves. As of today, this process requires thorough alarm data analytics, which only few external and expensive experts can do. The tools provided by vendors of medical devices typically lack sophistication, depth and breadth of analysis, user guidance and linkage to root causes. To overcome this, the AlarmRedux project targets an "alarm cockpit" with an intuitive visualization of ASQ metrics. To this end, one workstream of the project focused on identification of metrics for alarm system quality.

III. Methods

A scoping review, as described by Arksey and O'Malley [10], was conducted including a MEDLINE search and guided interviews, with ICU nurses from two German hospitals. A total of nine interviews was performed (5/4 per site), with an average duration of 45 minutes. The interviews were followed by a workshop with experts for clinical alarm management. Experts from medical device vendors, medical informatics scientists, intensivists and intensive care nurses participated in the workshop. Results were then refined in an iterative consensus process supported by telephone conferences.

In a next step, the metrics will be validated empirically. During several weeks, ICU nurses from two hospitals will record their impression once per shift and patient, using a short questionnaire. The subjective impressions and the input parameter values of the metrics for the respective time spans/patients will be subjected to a multiple regression. Parameters with the best fit to the subjective impression will then be used in a numeric model for the respective metrics.

IV. Results

There were two kinds of direct results from the scoping review: One was the need to introduce the notion of *quality dimensions*, i.e. particular ways in which the alarm system quality can be suboptimal, and the other one were concrete calculations from alarm data, which were supposed to bear some relation to ASQ. These calculations are henceforth referred to as "*calculated parameters*" or just "*parameters*".

A. Quality dimensions

In the cumbersome process of isolating quality dimensions which had to strike a balance between abstracting from what is technically measurable given a concrete technical infrastructure, but still remain practically useful, five dimensions were finally adopted. Each of the following dimensions corresponds to a metric.

1) Alarm load

This dimension covers all aspects of the number of alarms. Beyond the AMBD aspect, phenomena like the different weights of alarms by severity and type, but also "bursts" of alarms or the equality of the distribution of alarms over time and over different caregivers should impact the *alarm load* dimension of ASQ.

2) Avoidable alarms

The often cited cry-wolf effect, i.e. the effect that alarm types which often are non-actionable tend negatively impact the reactivity of staff, has been demonstrated both in a lab setting [11] and in ICU practice [12]. This dimension covers ASQ deficits due to alarms which require no consequence on the caregiver's side, but also "technical" alarms, which require an action, but could be avoided nonetheless.

3) Responsiveness/alarm handling

Suboptimal quality in this dimension not only shows in delayed reaction, but also in inadequate reactions like, for instance, turning off alarms altogether, muting alarm sounds or failure to take necessary corrective actions.

4) Sensing

The *Sensing* part of ASQ describes how suitable the technical infrastructure and its configuration is for a risk-adjusted patient surveillance at any given time. Inadequate sensors and consumables fall into this category as well as overmonitoring or undermonitoring of patients.

5) Exposure

A high quality alarm system will limit the exposure of people to alarms to those whose attention needs to be directed towards a particular state of affairs. This excludes patients and relatives, but also all caregivers not responsible for a particular patient at least most of the time. Thus differences in the ways alarms are distributed and signaled to caregivers will be seen in the *Exposure*-dimension of ASQ.

B. Calculated parameters

There were more than thirty suggestions for parameters which were regarded as highlighting particular aspects of ASQ.

They range from well-known ones, like AMBD or average response times to red or yellow alarms, to rather elaborate ones, introducing novel concepts like "alarm bursts", "red-after-yellow" or "proper pauses". To give an idea of the sophistication of these calculated parameters, two of them shall be explained in some detail:

1) Technical burst frequency

An alarm burst is a series of many alarms in a time interval, for instance "more than 30 alarms in one hour" for a single patient. It is evident that alarm bursts have a high potential to overload the responsible caregivers, but also annoy or even stress the rest of staff and patients. There are many reasons for alarm bursts: Criticality of a patient may be one, but also maladjusted alarm limits and failure to resolve a technical issue immediately. Technical burst frequency calculates how often alarm bursts occur with a dominant fraction of technical alarms during the burst.

2) Proper pause to pause ratio (PPPR)

Pauses are means to suspend all alarms for a short period of time, helpful for suppressing artifacts during a bedside maneuver, like the suctioning of a patient. It is vital that pauses are terminated once the particular maneuver is completed or artifacts are no longer to be expected. Since pauses self-terminate after a configurable time, one can determine how often the pause was actively suspended – constituting a "*proper pause*" – and how often the pause just continued for the default duration, maybe even with the caregiver having left the room, a potentially dangerous situation. The *proper pause to pause ratio* parameter calculates the ratio of *proper pauses* to all pauses and should ideally be near 1.0

C. Mapping calculated parameters to dimensions

In a final step, the parameters were mapped to the ASQ dimensions in a many-to-one fashion: AMBD and burst frequency, for instance, were mapped to alarm load whereas average response time to clinical red alarms was mapped to responsiveness and PPPR was mapped to the Sensing dimension. An excerpt of the mapping table with just eleven parameters is shown in fig.

Figure 3.

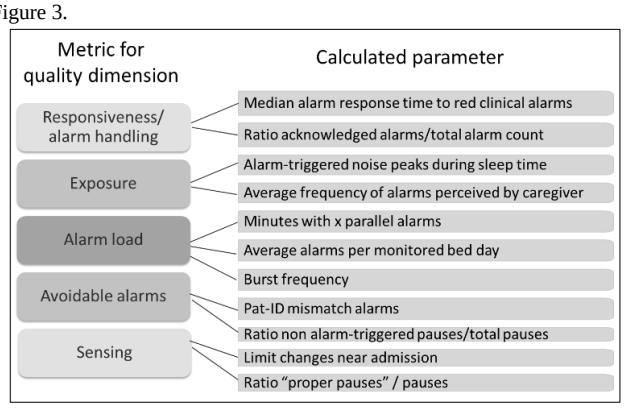

Figure 3: Mapping of eleven calculated parameters to metrics for the five quality dimensions

V. DISCUSSION AND OUTLOOK

In our view, the field of research into Alarm fatigue on ICUs suffers from two important deficits, the "measurability gap" and the "diagnosis-therapy gap".

A. The measurability gap

Alarm fatigue, cognitive overload and desensitization, and its consequences, fatal and near-fatal events as well as staff burnout and the "second-victim effect", are almost invariably described qualitatively. The work by Cho et al. [13] targeting a scoring system for AF using on a questionnaire based on the HTF survey [14] was a first step to come to a quantification of alarm fatigue. There is, to the best of our knowledge, no published attempt to make the various facets of AF routinely measurable and comparable in a comprehensive manner, i.e. also taking into account contributing causes and its effects. We are convinced that one of the reasons for the measurability gap is an ill-guided interpretation of the term "alarm fatigue": Many researchers see it – or seem to see it – as a condition of the single healthcare worker, which makes it an issue of occupational medicine, makes a generalization to the unit level problematic and promotes a tendency to disregard dysfunctionalities on the level of technical infrastructure or organization. We argue that alarm fatigue should rather be regarded as a condition of the entire sociotechnical system ICU. In this way one can ascribe e.g. a tendency to react inadequately to alarms to an alarm system without pathologizing or blaming individuals. Also, it becomes feasible to measure said tendency without having to average over pathological conditions of individuals.

Another reason for the measurability gap may lie in the fact that the currently used calculated parameters AMBD and response times/rates constitute to narrow a focus with regard to the overall alarm system quality. The quality dimensions and associated metrics defined in and following the scoping review promise to be a step towards making alarm fatigue routinely measurable in a holistic way.

Our own experience from ICU projects shows that the caregivers' gut feeling about their ASQ is often not reflected by AMBD and response time measures. We hope that in the planned validations we can create a numeric model with a better fit of quantitative measurements to the caregivers' perceived stress and fatigue. We also stipulate that the metrics will provide a better basis for post-interventional pre-post comparisons as well as cross-unit benchmarks, though the latter may be hard to achieve given the fact that many structural factors (like floor-layout or nurse-to-patient ratio) influencing the measures may differ from unit to unit.

B. The diagnosis-therapy gap

Even in undeniable cases of alarm fatigue, for instance in situations where repeatedly inadequate or delayed reactions have led to patient deaths, the question of what should be done remains not an easy one to answer. It is a tell-tale sign that publications like [3],[15], the AACN "Alarm Management Performance Improvement Plan" or the AAMI "Clinical Alarm Management Compendium" [17] do list recommendations on

how to improve ASQ, but do not link them to observations or assessment results. Like in clinical medicine it is obvious that interventions without clear indications in order to address a complex syndrome are inefficient at best, and ineffective at worst. What seems to be missing is a middle layer which links observations and measurements to underlying root causes, so that specific interventions to target these root causes can be selectively implemented. In the AlarmRedux project, work on such a layer, the "pathophysiology of excessive alarms" has led to causal models of contributing factors for and impacts of excessive alarms (shown in figs. 1 and 2). The quality dimensions and metrics to gauge ASQ established so far cover a fair portion of the causal net, though not all of it. Inasmuch as they do, they should enable a better decision making as to which interventions should be taken to address specific problems in the alarm system. Once the validation of the metrics has been concluded, further work will be necessary to establish evidence-based interventions to improve alarm system quality.

C. Completeness and practicality of the quality dimensions and metrics

The proposed dimensions are a first attempt to depart from a single parameter (AMBD) approach and cover the multi-faceted topic of excessive alarms and alarm fatigue. They will undoubtedly be refined and extended in the future, as they are utilized to support caregivers in their management of their alarm system. In order to do this, certain technical hurdles need to be overcome:

Calculating parameters for the *Exposure*-metric requires – amongst others – noise level measurement equipment which is not routinely available. For the time being this metric remains unused. Moreover, routinely calculating the input parameters for the metrics requires access to alarm related data in a depth and breadth which is currently not provided by vendors of monitors and other medical devices. Only if vendors embrace the idea to enable end users to manage their alarm system quality and disclose the required data from their central station routinely to an alarm cockpit, only then can the elaborated metrics be put to the best use: Increasing patient safety and improving the working conditions for ICU staff.

CONFLICT OF INTEREST

Dirk Hüske-Kraus is working at Philips Healthcare. The other authors state that they have no conflict of interests. All authors are part of the AlarmRedux-Project funded by the German Federal Ministry of Education and Research (Code: 16SV7501)

REFERENCES

[1] Borowski, M., Görges, M., Fried, R., Such, O., Wrede, C., & Imhoff, M. (2011). Medical device alarms. *Biomedizinische Technik/Biomedical Engineering*, *56*(2), 73-83.

[2] Lacker, C. (2011). Physiologic alarm management. *Pennsylvania Patient Safety Advisory*, *8*(3), 105-8.

[3] AAMI Foundation's HTSI (2012). Using Data to Drive Alarm System Improvement Efforts: The Johns Hopkins Hospital Experience. available from: www.aami.org/htsi/SI_Series/Johns_Hopkins_White_Paper.pdf, last accessed 5.5.2018

[4] ECRI Institute (2016). Top 10 health technology hazards for 2017. Health Devices.

[5] Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical instrumentation & technology*, *46*(4), 268-277.

[6] Wilken, M., Hüske-Kraus, D., Klausen, A., Koch, C., Schlauch, W., & Röhrig, R. (2017). Alarm Fatigue: Causes and Effects. *Studies in health technology and informatics*, *243*, 107-111.

[7] Imhoff, M., & Kuhls, S. (2006). Alarm algorithms in critical care monitoring. *Anesthesia & Analgesia*, *102*(5), 1525-1537.

[8] Voepel-Lewis, T., Parker, M. L., Burke, C. N., Hemberg, J., Perlin, L., Kai, S., & Ramachandran, S. K. (2013). Pulse oximetry desaturation alarms on a general postoperative adult unit: a prospective observational study of nurse response time. *International journal of nursing studies*, *50*(10), 1351-1358.

[9] Paine, C. W., Goel, V. V., Ely, E., Stave, C. D., Stemler, S., Zander, M., & Bonafide, C. P. (2016). Systematic review of physiologic monitor alarm characteristics and pragmatic interventions to reduce alarm frequency. *Journal of hospital medicine*, *11*(2), 136-144.

[10] Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

[11] Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics*, *38*(11), 2300-2312.

[12] Bonafide, C. P., Lin, R., Zander, M., Graham, C. S., Paine, C. W., Rock, W., ... & Localio, A. R. (2015). Association between exposure to nonactionable physiologic monitor alarms and response time in a children's hospital. *Journal of hospital medicine*, *10*(6), 345-351.

[13] Cho, O. M., Kim, H., Lee, Y. W., & Cho, I. (2016). Clinical alarms in intensive care units: Perceived obstacles of alarm management and alarm fatigue in nurses. *Healthcare informatics research*, *22*(1), 46-53.

[14] Ruppel, H., Funk, M., Clark, J. T., Gieras, I., David, Y., Bauld, T. J., ... & Holland, M. L. (2018). Attitudes and Practices Related to Clinical Alarms: A Follow-up Survey. *American Journal of Critical Care*, *27*(2), 114-123.

[15] Bell, L. (2010). Monitor alarm fatigue. *American Journal of Critical Care*, *19*(1), 38-38.

[16] American Association of Critical-care Nurses (2013): Alarm Management Performance Improvement Plan, available from: https://www.aacn.org/clinical-resources/clinical-toolkits/strategies-for-managing-alarm-fatigue

[17] AAMI foundation (2015). Clinical Alarm Management Compendium, available from: http://s3.amazonaws.com/rdcms-aami/files/production/public/FileDownloads/HTSI/Alarms/Alarm_Compendium_2015.pdf last accessed 6.5.2018