

Original Paper

# Evaluating the Construct Validity of the Charité Alarm Fatigue Questionnaire using Confirmatory Factor Analysis

Maximilian Markus Wunderlich<sup>1</sup>, MSc; Henning Krampe<sup>2</sup>, PhD; Kristina Fuest<sup>3</sup>, Dr med; Dominik Leicht<sup>4</sup>, Dr med; Moriz Benedikt Probst<sup>5</sup>, Dr med; Julian Runge<sup>6</sup>, Dr med; Sebastian Schmid<sup>7</sup>, Dr med; Claudia Spies<sup>2</sup>, Prof Dr med; Björn Weiß<sup>2</sup>, Prof Dr med; Felix Balzer<sup>1\*</sup>, Prof Dr med, Dr rer nat; Akira-Sebastian Poncette<sup>1,2\*</sup>, PD, Dr med; CAFQa Study Group Germany

<sup>1</sup>Institute of Medical Informatics, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

<sup>2</sup>Department of Anesthesiology and Intensive Care Medicine CVK/CCM, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

<sup>3</sup>Department of Anaesthesiology & Intensive Care Medicine, Klinikum Rechts der Isar, School of Medicine, Technical University Munich, Munich, Germany

<sup>4</sup>Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Justus Liebig University, Giessen, Germany

<sup>5</sup>Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany

<sup>6</sup>Department for Anesthesiology, Surgical Intensive Care, Pain and Palliative Medicine, Marien Hospital Herne–Universitätsklinikum der Ruhr-Universität Bochum, Herne, Germany

<sup>7</sup>Department of Anesthesiology and Intensive Care Medicine, University Hospital Ulm, Ulm University, Ulm, Germany

\*these authors contributed equally

**Corresponding Author:**

Akira-Sebastian Poncette, PD, Dr med

Institute of Medical Informatics

Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin

Charitéplatz 1

Berlin, 10117

Germany

Phone: 49 30 450 581018

Email: [akira-sebastian.poncette@charite.de](mailto:akira-sebastian.poncette@charite.de)

## Abstract

**Background:** The Charité Alarm Fatigue Questionnaire (CAFQa) is a 9-item questionnaire that aims to standardize how alarm fatigue in nurses and physicians is measured. We previously hypothesized that it has 2 correlated scales, one on the psychosomatic effects of alarm fatigue and the other on staff's coping strategies in working with alarms.

**Objective:** We aimed to validate the hypothesized structure of the CAFQa and thus underpin the instrument's construct validity.

**Methods:** We conducted 2 independent studies with nurses and physicians from intensive care units in Germany (study 1: n=265; study 2: n=1212). Responses to the questionnaire were analyzed using confirmatory factor analysis with the unweighted least-squares algorithm based on polychoric covariances. Convergent validity was assessed by participants' estimation of their own alarm fatigue and exposure to false alarms as a percentage.

**Results:** In both studies, the  $\chi^2$  test reached statistical significance (study 1:  $\chi^2_{26}=44.9$ ;  $P=.01$ ; study 2:  $\chi^2_{26}=92.4$ ;  $P<.001$ ). Other fit indices suggested a good model fit (in both studies: root mean square error of approximation  $<0.05$ , standardized root mean squared residual  $<0.08$ , relative noncentrality index  $>0.95$ , Tucker-Lewis index  $>0.95$ , and comparative fit index  $>0.995$ ). Participants' mean scores correlated moderately with self-reported alarm fatigue (study 1:  $r=0.45$ ; study 2:  $r=0.53$ ) and weakly with self-perceived exposure to false alarms (study 1:  $r=0.3$ ; study 2:  $r=0.33$ ).

**Conclusions:** The questionnaire measures the construct of alarm fatigue as proposed in our previous study. Researchers and clinicians can rely on the CAFQa to measure the alarm fatigue of nurses and physicians.

**Trial Registration:** ClinicalTrials.gov NCT04994600; <https://www.clinicaltrials.gov/study/NCT04994600>

**Keywords:** patient monitoring; intensive care unit; alarm; alarms; validity; validation; safety; intensive; care; alarm fatigue; alarm management; patient safety; ICU; alarm system; alarm system quality; medical devices; clinical alarms; questionnaire; questionnaires; warning; factor analysis

## Introduction

### Background

Alarm fatigue is a phenomenon where health care workers in intensive care units (ICUs) become desensitized to alarms of medical devices [1]. It can make ICU staff feel stressed, and it is a substantial risk to patient safety, as it can lead to alarms being missed or acknowledged with delay [2]. When implementing interventions or IT solutions [3] that try to remedy alarm fatigue, clinicians and clinical alarm researchers need a reliable way to assess whether they were successful. However, they have not yet agreed on a standardized way of measuring alarm fatigue [4,5], even though it was recognized more than 2 decades ago [6].

Solely analyzing an ICU's alarm log data cannot serve as a measure of staff's alarm fatigue. While it is a valuable method for designing alarm management interventions [7], there is no clear association between the number of alarms on an ICU and staff's subjective alarm fatigue. For example, simply focusing on the number of alarms disregards their temporal distribution (eg, it might fatigue staff more if alarms came in random bursts than if they were evenly spaced out [8]). This could be one of the reasons why in their intervention study, Sowan et al [9] did not find that staff's alarm fatigue improved despite having managed to significantly reduce the number of alarms on their ICU.

Therefore, we recently developed the Charité Alarm Fatigue Questionnaire (CAFQa), which is a 9-item questionnaire that measures alarm fatigue in nurses and physicians [10]. Using exploratory factor analysis, we identified 2 correlated factors: one revolving around the psychophysiological effects of alarms (eg, headaches and feelings of distraction), and one revolving around ICU staff's alarm management strategies (eg, customization of alarm limits). We named the former the "alarm stress scale" and the latter the "alarm coping scale." The alarm coping scale consists of items that are reversely scored. Hence, a high score on either scale is indicative of alarm fatigue.

When developing a new questionnaire, it is essential to establish construct validity, that is, whether the questionnaire truly measures what it attempts to measure. One way to test an instrument's construct validity is to administer it to a different sample and test whether the originally proposed factor structure reemerges using confirmatory factor analysis [11,12] (for a recent example see Canivez et al [13]).

### Aim

We aim to validate the exploratively derived factor structure of the CAFQa and thus underpin the instrument's construct validity.

## Methods

### Ethical Considerations

The ethical approval for this study was granted by the ethics committee of the Charité–Universitätsmedizin Berlin (EA4/218/20) and, if required, confirmed by the local ethics committee at the participating hospital. This study was conducted in compliance with the relevant guidelines and regulations. All participants voluntarily agreed to take part after being fully informed about the study. In study 1, as a reward for completing the questionnaire, we offered participants the chance to enter a draw where they could win a €50 (US \$53) voucher for online shopping. Participants were asked to consent to have their data collected, analyzed, and stored anonymously.

### Participants

In both studies, we included nurses, physicians, and nurses in training, while excluding other professions and non-ICU staff.

### Study 1

We recruited participants from 9 ICUs of 5 large German hospitals. The questionnaire was administered on the web using REDCap (Research Electronic Data Capture; Vanderbilt University) between October 2021 and July 2022.

### Study 2

Using a mailing list, we invited all members of the German Society of Anaesthesiology and Intensive Care Medicine [14] to fill out the web-based questionnaire (again using REDCap) between March 2023 and July 2023.

### Questionnaire

The questionnaire used in both studies was identical and consisted of all 9 items from the CAFQa [10] and 5 general questions about the alarm situation in participants' ICUs. These general questions were not part of the analysis for this report. All 14 items were pseudorandomly arranged and required responses on a Likert scale ranging from –2 (indicating "I do not agree at all") to 2 (indicating "I very much agree"). Items with negative valences were reverse scored. Demographic items asked participants about their average number of workdays in an intensive care or monitoring area, their number of years and months of ICU experience, their workplace (campus and unit), and their profession. We made small adjustments to the original wording of 2 items (items 8 and 9) to improve readability: In item 8 we used "situation" instead of "urgency." In item 9 we used the phrase "clinical pictures" instead of "clinical symptoms."

## Statistical Analysis

All analyses were conducted in R (version 4.2.1; R Foundation for Statistical Computing) using the following packages: *Tidyverse* [15], *reshape2* [16], *psych*, *semPlot* [17], and *lavaan* [18]. For study 1, we pooled the data from the participating hospitals.

## Missing Data

In accordance with Heymans and Eekhout [19], we used predictive mean matching via the *mice* package [20] to impute missing data that were assumed to be missing at random (MAR). We did not impute questionnaires that were either completely empty or terminated prematurely (presumably due to survey fatigue), as the assumption of MAR was not met in these cases. We assumed that survey fatigue occurred if a participant failed to respond to at least the final 20% of the questionnaire (ie, the last 3 or more of the 9 items of the CAFQa plus the 5 general questions). In total, 0.3% of the data were MAR.

## Testing Assumptions of Confirmatory Factor Analysis

In both studies, the results of the Mardia test indicated that the multivariate skew did not come from a normal distribution with  $P < .001$ . Outliers were identified using Mahalanobis distances, with none being detected in study 1 and 4 being detected in study 2 (for both studies:  $\chi^2_9$  cutoff=27.9;  $P < .001$ ). Visual inspection of the data from all 4 cases revealed no unusual response patterns. Given the large sample size, we decided not to remove any outliers. The Kaiser-Meyer-Olkin statistic [21] in study 1 was 0.76, and in study 2 it was 0.8. In both studies, the Bartlett test of sphericity [22] rejected the null hypothesis that the correlation matrix was an identity matrix (study 1:  $\chi^2_{36}=438.3$ ;  $P < .001$ ; study 2:  $\chi^2_{36}=2495.4$ ). There was no evidence of multicollinearity in either study as the determinant of both  $R$  matrices was greater than 0.00001 [23] and no correlations were greater than 10.71. Overall, these results suggest that the data of both studies were suitable for factor analysis.

## Confirmatory Factor Analysis

For both studies, we specified the model in line with our previous findings [10], with 2 correlated latent factors, labeled “alarm stress” and “alarm coping.” Items 1-5 were assigned to “alarm stress.” Items 5-9 were assigned to “alarm coping.” Since all CAFQa items are ordered categorical variables (due to being measured on a 5-point Likert scale) and because the Mardia tests indicated that the multivariate skew of both studies did not come from a normal distribution, we used the unweighted least-squares (ULS) algorithm based on polychoric covariances for estimating factor loadings [24-26]. We assessed the goodness-of-fit of the model using  $\chi^2$ , and the following fit indices in line with the cutoff criteria defined by Hu and Bentler [27]: root mean square error of approximation (RMSEA), relative noncentrality index (RNI),

Tucker-Lewis index (TLI), standardized root mean squared residual (SRMR), and comparative fit index (CFI).

## Convergent Validity

At the end of the questionnaire in both studies, we provided participants with a brief description of alarm fatigue and asked them to estimate their personal alarm fatigue as a percentage (0% indicating no alarm fatigue and 100% indicating extreme alarm fatigue). We also asked participants to provide their perceived rate of false alarms in their ICU as a percentage (0% indicating no false alarms, 100% indicating no true alarms). To measure convergent validity, we correlated the participants' mean scores on the questionnaire with the self-provided alarm fatigue and false alarm rate estimations (in total and per factor).

## Internal Consistency

As a measure of internal consistency, we report Cronbach coefficient  $\alpha$ , the McDonald coefficient  $\omega$  [28], and the mean interitem correlation for both factors.

## Results

### Participants

#### Study 1

We received 363 submissions. Among these, 23 came from participants who did not consent to have their data analyzed, 67 questionnaires were empty, and 8 showed signs of survey fatigue. Therefore, the sample size for this study was 265. The number of participants was roughly similar for each hospital (Giessen:  $n=43$ ; Herne:  $n=50$ ; Munich:  $n=64$ ; Ulm:  $n=57$ ; Vivantes:  $n=51$ ). Most participants were nurses ( $n=150$ , 56.6%) and 35.8% ( $n=95$ ) were physicians. A few participants ( $n=9$ , 3.4%) were supporting nurses, nurses in training, medical students, or interns, while 4.2% ( $n=11$ ) did not state their profession.

#### Study 2

Of the 1564 submissions we received, 69 came from participants who refused to consent to have their data processed and 223 were empty questionnaires. We suspected survey fatigue in 60 cases. Hence, the sample size of study 2 was 1212. Contrary to study 1, more participants were physicians ( $n=1002$ , 82.7%) than nurses ( $n=186$ , 15.3%). Again, the group of supporting nurses, nurses in training, medical students, and interns was a minority ( $n=6$ , 0.5%). Among the participants, 1.5% ( $n=18$ ) did not state their profession.

## Confirmatory Factor Analysis

Descriptive statistics of both studies are presented in Table 1 for each item.

**Table 1.** Descriptive statistics for each item and the pattern coefficients found in the confirmatory factor analysis of the 2-factor model in both studies. All loadings were statistically significant at  $P < .001$ .

Item	Description	Study 1					Study 2				
		Factor 1 (95% CI)	Factor 2 (95% CI)	Mean (SD)	Kurtosis	Skew	Factor 1 (95% CI)	Factor 2 (95% CI)	Mean (SD)	Kurtosis	Skew
1	With too many alarms on my ward, my work performance, and motivation decrease.	0.730 (0.643-0.818)	— <sup>a</sup>	0.47 (1.01)	-0.62	-0.3	0.677 (0.636-0.717)	—	0.51 (1.08)	-0.63	-0.37
2	Too many alarms trigger physical symptoms for me, e.g., nervousness, headaches, and sleep disturbances.	0.706 (0.612-0.800)	—	0.23 (1.26)	-1.13	-0.16	0.694 (0.653-0.735)	—	0.22 (1.20)	-1.00	-0.14
3	Alarms reduce my concentration and attention.	0.725 (0.635-0.814)	—	0.43 (1.07)	-0.91	-0.21	0.813 (0.779-0.846)	—	0.63 (1.03)	-0.59	-0.39
4	My or neighboring patients' alarms or crisis alarms frequently interrupt my workflow.	0.432 (0.318-0.547)	—	0.87 (0.83)	-0.41	-0.39	0.519 (0.469-0.570)	—	0.70 (0.87)	-0.28	-0.37
5	There are situations when alarms confuse me.	0.488 (0.384-0.593)	—	0.08 (1.09)	-0.73	-0.09	0.634 (0.592-0.676)	—	0.19 (1.07)	-0.85	-0.04
6	In my ward, procedural instruction on how to deal with alarms is regularly updated and shared with all staff. <sup>b</sup>	—	0.434 (0.270-0.598)	0.77 (1.24)	-0.68	-0.7	—	0.449 (0.375-0.523)	1.10 (1.02)	0.44	-1.06
7	Responsible personnel respond quickly and appropriately to alarms. <sup>b</sup>	—	0.587 (0.424-0.750)	-0.32 (0.82)	-0.12	-0.12	—	0.639 (0.567-0.711)	-0.39 (0.90)	-0.05	0.19
8	The acoustic and visual monitor alarms used on my ward floor and in my nurses' station allow me to assign the patient, the device, and the situation clearly. <sup>b</sup>	—	0.349 (0.182-0.517)	-0.46 (1.06)	-0.52	0.36	—	0.428 (0.359-0.498)	-0.36 (1.15)	-0.66	0.32
9	Alarm limits are regularly adjusted based on patients' clinical pictures (e.g., blood pressure limits for conditions after bypass surgery). <sup>b</sup>	—	0.581 (0.428-0.734)	-0.38 (0.93)	-0.24	0.26	—	0.575 (0.508-0.641)	-0.35 (0.94)	-0.12	0.40

<sup>a</sup>Not applicable.<sup>b</sup>Item with a negative valence that is reversely scored.

### Study 1

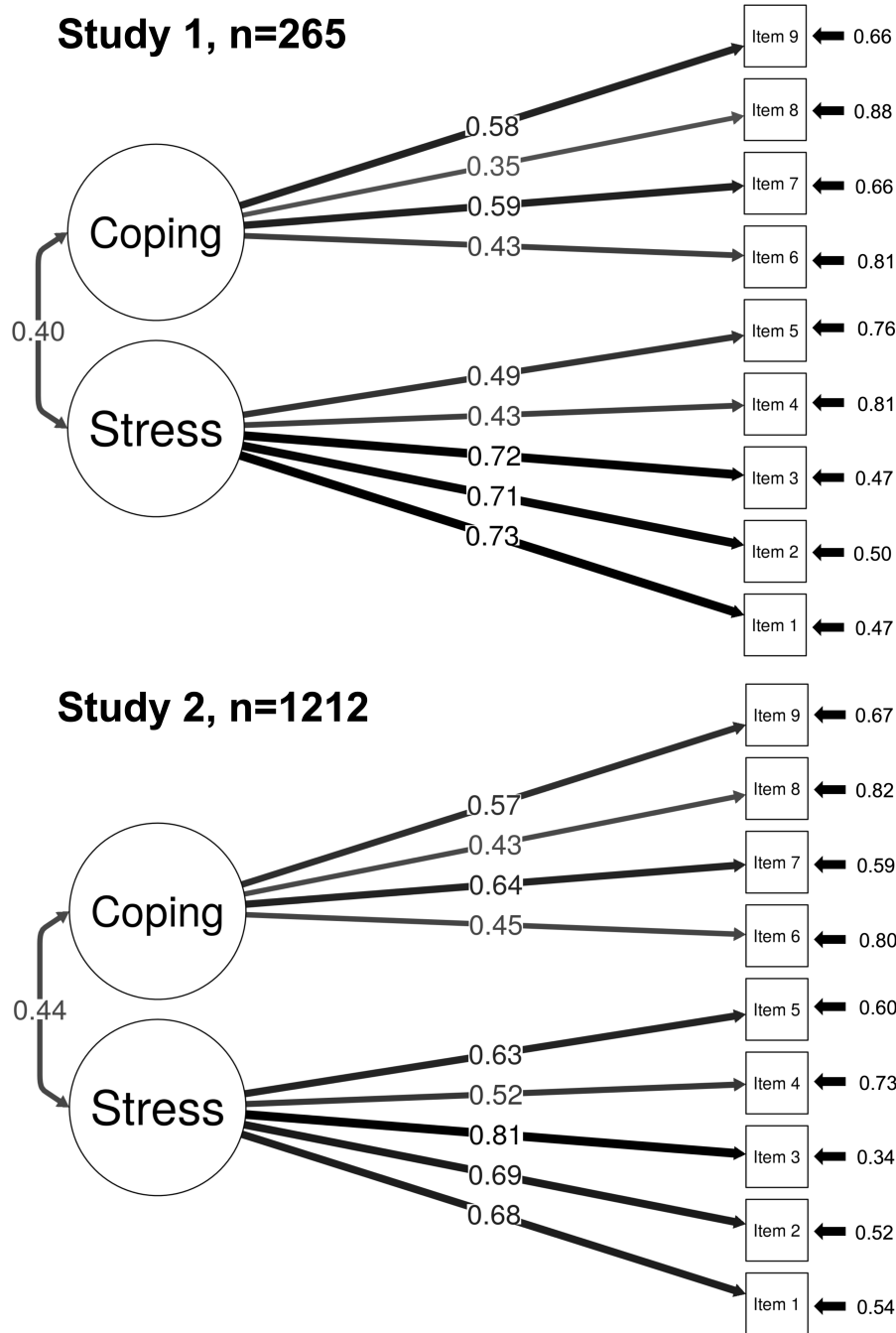
While the  $\chi^2$  test was significant at  $\alpha=.05$  with  $\chi^2_{26}=44.9$  ( $P=.01$ ), indicating that the model did not fit the data, all fit indices suggested a good model fit: RMSEA=0.03, SRMR=0.052, RNI=0.989, TLI=0.985, and CFI=0.989. All items loaded onto their hypothesized factors as expected, with factor loadings that were statistically significant at  $P<.001$ , ranging from 0.35 to 0.73. The factors were moderately correlated at 0.4 (95% CI 0.21-0.59;  $P<.001$ ).

### Study 2

As in study 1, the  $\chi^2$  test was significant ( $\chi^2_{26}=92.4$ ;  $P<.001$ ), indicating that the model did not fit the data, while

the fit indices showed a good model fit: RMSEA=0.046, SRMR=0.041, RNI=0.982, TLI=0.975, and CFI=0.982. Again, all items loaded onto their hypothesized factors as expected, with factor loadings that were statistically significant at  $P<.001$ , ranging from 0.43 to 0.81. The factors were moderately correlated at 0.44 (95% CI 0.36-0.51;  $P<.001$ ) (Figure 1).

**Figure 1.** Diagram of the final model from each study. Factors are shown in circles (factor 1: alarm stress; factor 2: alarm coping), and items 1-9 in squares. In both studies, the variance of the factors was constrained to 1. The arrows connecting the factors with the items are the factor loadings and arrows pointing toward the items show the residuals. The arrows between the factors show their correlation.



**Convergent Validity**

In study 1, the participants’ mean scores on the questionnaire correlated moderately with self-reported alarm fatigue ( $r_{242}=0.45$ , 95% CI 0.34-0.54;  $P<.001$ ) and weakly with the

perceived percentage of false alarms ( $r_{247}=0.3$ , 95% CI 0.18 to  $-0.41$ ;  $P<.001$ ). Similar patterns were observed in study 2 (Table 2 provides full details).

**Table 2.** Correlation coefficients, *P* values, and CIs used to investigate the convergent validity in each study.

Study and correlation measure	<i>r</i> ( <i>df</i> ; 95% CI)	<i>P</i> value
<b>Study 1</b>		
MS-SRAF <sup>ab</sup>	0.45 (242; 0.34-0.54)	<.001
F1S-SRAF <sup>c</sup>	0.42 (242; 0.31-0.52)	<.001
F2S-SRAF <sup>d</sup>	0.29 (242; 0.17-0.4)	<.001

Study and correlation measure	<i>r</i> ( <i>df</i> ; 95% CI)	<i>P</i> value
MS-PPFA <sup>e</sup>	0.30 (247; 0.18-0.41)	<.001
F1S-PPFA	0.20 (247; 0.08-0.32)	.002
F2S-PPFA	0.29 (247; 0.17-0.4)	<.001
<b>Study 2</b>		
MS-SRAF	0.53 (1180; 0.49-0.57)	<.001
F1S-SRAF	0.49 (1180; 0.45-0.53)	<.001
F2S-SRAF	0.34 (1180; 0.29-0.39)	<.001
MS-PPFA	0.33 (1182; 0.28-0.38)	<.001
F1S-PPFA	0.26 (1182; 0.21-0.32)	<.001
F2S-PPFA	0.28 (1182; 0.23-0.33)	<.001

<sup>a</sup>MS: mean score on the questionnaire.

<sup>b</sup>SRAF: self-reported alarm fatigue.

<sup>c</sup>F1S: scores on factor 1.

<sup>d</sup>F2S: scores on factor 2.

<sup>e</sup>PPFA: perceived percentage of false alarms.

## Internal Consistency

In study 1, the Cronbach  $\alpha$  of factor 1 was 0.72, and it was 0.49 for factor 2. Cronbach  $\alpha$  across factors was 0.67. The mean interitem correlation on factor 1 was 0.38, and it was 0.23 on factor 2. The McDonald coefficient  $\omega$  for factor 1 was 0.77, and for factor 2 it was 0.55. The overall coefficient  $\omega$  for the assessment was 0.8.

Results were similar in study 2: the Cronbach  $\alpha$  was 0.77 for factor 1 and 0.55 for factor 2. Cronbach  $\alpha$  across factors was 0.72. The mean interitem correlation was 0.44 on factor 1 and 0.27 on factor 2. The McDonald coefficient  $\omega$  was 0.8 for factor 1 and 0.59 for factor 2. The overall coefficient  $\omega$  for the assessment was 0.85.

## Discussion

We aimed to underpin the construct validity of the CAFQa by submitting the exploratively derived factor structure from our previous study to confirmatory factor analysis in 2 independent studies. While the  $\chi^2$  test rejected the model in both studies, all fit indices indicated a good model fit. The factor loadings ranged from 0.35 to 0.73 in study 1 and from 0.43 to 0.81 in study 2; all were statistically significant. Overall, these results support the hypothesized factor structure. The questionnaire seems to measure the construct of alarm fatigue as proposed in our previous work [10].

The  $\chi^2$  test is known to be sensitive to large sample sizes [29], which might explain its statistical significance. We did not modify the model because all fit indices indicated a good fit and because model modifications, no matter how small or plausible, can make a model less generalizable.

In both studies, the first factor, that is, the alarm stress scale, and the overall questionnaire were internally consistent. However, the second factor, that is, the alarm coping scale, seems to have issues with its internal consistency. Here, Cronbach  $\alpha$  and McDonald  $\omega$  were 0.49 and 0.55 in study 1, respectively, and 0.55 and 0.59 in study 2, respectively. A

similar pattern can be found in our previous study, where the Cronbach  $\alpha$  of factor 2 was 0.57 [10]. An internally consistent questionnaire is desirable. However, it can also mean that items are very similar. It was our ambition to create a questionnaire that is brief while measuring the many facets of alarm fatigue. Future studies using the CAFQa should routinely assess the internal consistency of both factors. If the second factor continues to show medium internal consistency, research should be done on how it can be improved (eg, by adding additional items, which would come at the cost of a longer questionnaire).

Because no other measures of alarm fatigue exist that would allow us to estimate the CAFQa's convergent validity, we asked participants to rate their own alarm fatigue as well as the rate of false alarms they perceived in their daily work. Participants who had a high mean score on the questionnaire also rated themselves as more alarm fatigued ( $r=0.45$  in study 1 and  $r=.53$  in study 2). This positive correlation indicates the convergent validity of the questionnaire. Similarly, both studies demonstrated that participants with a high mean score on the questionnaire perceived more alarms to be false in their ICU. In study 1 this association was stronger for factor 2 than for factor 1. This makes sense since a high score on factor 2 (ie, the alarm coping scale) indicates that alarms are not properly managed (eg, by means of patient-specific threshold customizations), which typically leads to more false alarms [30]. However, study 2 could not replicate this pattern. Future research should find an answer to this question: Do ICU staff with a high perceived percentage of false alarms tend to develop stronger alarm fatigue, or do staff that are more alarm fatigued tend to perceive more alarms as being false?

## Limitations

The fit indices RMSEA, CFI, and TLI have been shown to overestimate model fit when using the ULS estimator [31,32], potentially leading researchers to accept a bad-fitting model. Yet, in our case, other fit indices indicated a good model fit. As in our previous work [10], the assumption that participants

can accurately reflect and express their own alarm fatigue as a percentage is likely flawed (otherwise, it would not be necessary to develop a questionnaire in the first place). However, most ICU nurses and physicians have heard of alarm fatigue, and we provided them with a brief recapitulation on the concept before having them answer the self-report item in each study. Likewise, it is also probably a flawed assumption that participants can accurately report the rate of false alarms, though Bliss et al [33] showed that participants were able to adapt their response frequencies to alarms based on to the perceived probability that an alarm was not false, thus suggesting that people might have an intuitive grasp of the rate of false alarms in their unit. All in all, we believe

that asking these self-rating questions is a valuable method for assessing convergent validity when no other instrument is available.

## Conclusion

Our results from 2 independent studies underpin the construct validity of the CAFQa. All items consistently loaded onto the factors, as we proposed in a previous publication [10]. When conducting research or quality improvement projects in ICUs, clinical alarm researchers and clinicians can rely on this instrument to measure, compare, and benchmark the alarm fatigue of nurses and physicians.

## Acknowledgments

ASP is a participant in the Charité Digital Clinician Scientist Program funded by the Charité–Universitätsmedizin Berlin and the Berlin Institute of Health. We are grateful to Nicolas Frey for supervising the logistical aspects of the survey distribution, to Sandro Amende-Wolf for his valuable clinical insights, and to Daniel Schulze for his critical examination of our statistical analysis. We are grateful to the German Society of Anaesthesiology and Intensive Care Medicine for the distribution of the questionnaire to its members. This article was extracted from the doctoral thesis of MMW.

## Data Availability

The data sets generated and analyzed during the current study are available in the Zenodo repository [34].

## Authors' Contributions

MMW, CS, FB, and ASP conceptualized the study. MMW, KF, DL, MBP, JR, and SS conducted the investigation. MMW and HK developed the study design and ensured methodological rigor. MMW, BW, and ASP contributed to the formal data analysis. MMW and ASP drafted the initial manuscript and contributed to its revision. MMW created all figures and tables and managed the project together with ASP. HK, KF, DL, MBP, JR, SS, CS, BW, FB, and ASP reviewed the manuscript. FB and ASP provided oversight and leadership throughout the study.

## Conflicts of Interest

None declared.

## Collaborators

Ulrich Frey (Department for Anesthesiology, Surgical Intensive Care, Pain and Palliative Medicine, Marien Hospital Herne–Universitätsklinikum der Ruhr-Universität Bochum, Herne, Germany), Christian von Heymann (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany), Jacqueline Holan (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany), Bettina Jungwirth (Department of Anesthesiology and Intensive Care Medicine, University Hospital Ulm, Ulm University, Ulm, Germany), Dragutin Popovic (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Vivantes Klinikum im Friedrichshain, Berlin, Germany), and Michael Sander (Department for Anaesthesiology, Intensive Care Medicine and Pain Therapy, Justus Liebig University, Giessen, Germany).

## References

1. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. *AACN Adv Crit Care*. 2013;24(4):378-386. [doi: [10.1097/NCI.0b013e3182a903f9](https://doi.org/10.1097/NCI.0b013e3182a903f9)] [Medline: [24153215](https://pubmed.ncbi.nlm.nih.gov/24153215/)]
2. Clinical alarm management: compendium. AAMI Foundation. 2015. URL: [https://www.aami.org/docs/default-source/foundation/alarms/alarm-compendium-2015.pdf?sfvrsn=2d2b53bd\\_2](https://www.aami.org/docs/default-source/foundation/alarms/alarm-compendium-2015.pdf?sfvrsn=2d2b53bd_2) [Accessed 2024-04-08]
3. Chromik J, Klopfenstein SAI, Pfitzner B, et al. Computational approaches to alleviate alarm fatigue in intensive care medicine: a systematic literature review. *Front Digit Health*. 2022;4:843747. [doi: [10.3389/fgdht.2022.843747](https://doi.org/10.3389/fgdht.2022.843747)] [Medline: [36052315](https://pubmed.ncbi.nlm.nih.gov/36052315/)]
4. Hüske-Kraus D, Wilken M, Röhrig R. Measuring alarm system quality in intensive care units. Presented at: Clusterkonferenz 2018; Apr 16 to 17, 2018; Freiburg, Germany. URL: <https://www.pflegeinnovationszentrum.de/wp-content/uploads/2018/12/17.-Measuring-Alarm-System-Quality-in-ICUs.pdf> [Accessed 2024-06-27]
5. Lewandowska K, Weisbrot M, Cieloszyk A, Mędrzycka-Dąbrowska W, Krupa S, Ozga D. Impact of alarm fatigue on the work of nurses in an intensive care environment—a systematic review. *Int J Environ Res Public Health*. Nov 13, 2020;17(22):8409. [doi: [10.3390/ijerph17228409](https://doi.org/10.3390/ijerph17228409)] [Medline: [33202907](https://pubmed.ncbi.nlm.nih.gov/33202907/)]



6. Wears RL, Perry SJ. Human factors and ergonomics in the emergency department. *Ann Emerg Med.* Aug 2002;40(2):206-212. [doi: [10.1067/mem.2002.124900](https://doi.org/10.1067/mem.2002.124900)] [Medline: [12140500](https://pubmed.ncbi.nlm.nih.gov/12140500/)]
7. Poncette AS, Wunderlich MM, Spies C, et al. Patient monitoring alarms in an intensive care unit: observational study with do-it-yourself instructions. *J Med Internet Res.* May 28, 2021;23(5):e26494. [doi: [10.2196/26494](https://doi.org/10.2196/26494)] [Medline: [34047701](https://pubmed.ncbi.nlm.nih.gov/34047701/)]
8. Wilken M, Hüske-Kraus D, Röhrig R. Alarm fatigue: using alarm data from a patient data monitoring system on an intensive care unit to improve the alarm management. *Stud Health Technol Inform.* Sep 3, 2019;267:273-281. [doi: [10.3233/SHTI190838](https://doi.org/10.3233/SHTI190838)] [Medline: [31483282](https://pubmed.ncbi.nlm.nih.gov/31483282/)]
9. Sowan AK, Gomez TM, Tarriela AF, Reed CC, Paper BM. Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: a pilot project. *JMIR Hum Factors.* Jan 11, 2016;3(1):e1. [doi: [10.2196/humanfactors.5098](https://doi.org/10.2196/humanfactors.5098)] [Medline: [27036170](https://pubmed.ncbi.nlm.nih.gov/27036170/)]
10. Wunderlich MM, Amende-Wolf S, Krampe H, et al. A brief questionnaire for measuring alarm fatigue in nurses and physicians in intensive care units. *Sci Rep.* Aug 24, 2023;13(1):13860. [doi: [10.1038/s41598-023-40290-7](https://doi.org/10.1038/s41598-023-40290-7)] [Medline: [37620385](https://pubmed.ncbi.nlm.nih.gov/37620385/)]
11. DiStefano C, Hess B. Using confirmatory factor analysis for construct validation: an empirical review. *J Psychoeduc Assess.* Sep 2005;23(3):225-241. [doi: [10.1177/073428290502300303](https://doi.org/10.1177/073428290502300303)]
12. Bryant FB, Yarnold PR, Michelson EA. Statistical methodology: VIII. using confirmatory factor analysis (CFA) in emergency medicine research. *Acad Emerg Med.* Jan 1999;6(1):54-66. [doi: [10.1111/j.1553-2712.1999.tb00096.x](https://doi.org/10.1111/j.1553-2712.1999.tb00096.x)] [Medline: [9928978](https://pubmed.ncbi.nlm.nih.gov/9928978/)]
13. Canivez GL, McGill RJ, Dombrowski SC, Watkins MW, Pritchard AE, Jacobson LA. Construct validity of the WISC-V in clinical cases: exploratory and confirmatory factor analyses of the 10 primary subtests. *Assessment.* Mar 2020;27(2):274-296. [doi: [10.1177/1073191118811609](https://doi.org/10.1177/1073191118811609)] [Medline: [30516059](https://pubmed.ncbi.nlm.nih.gov/30516059/)]
14. Deutsche Gesellschaft für Anästhesiologie & Intensivmedizin. URL: <https://www.dgai.de/> [Accessed 2023-08-28]
15. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4(43):1686. [doi: [10.21105/joss.01686](https://doi.org/10.21105/joss.01686)]
16. Wickham H. Reshaping data with the reshape package. *J Stat Soft.* 2007;21(12):1-20. [doi: [10.18637/jss.v021.i12](https://doi.org/10.18637/jss.v021.i12)]
17. Epskamp S. semPlot: unified visualizations of structural equation models. *Struct Equ Model.* Jul 3, 2015;22(3):474-483. [doi: [10.1080/10705511.2014.937847](https://doi.org/10.1080/10705511.2014.937847)]
18. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw.* 2012;48:1-36. [doi: [10.18637/jss.v048.i02](https://doi.org/10.18637/jss.v048.i02)]
19. Heymans M, Eekhout I. Chapter 8: missing data in questionnaires. In: *Applied Missing Data Analysis With SPSS and (R) Studio.* Bookdown; 2019. URL: <https://bookdown.org/mwheyman/bookmi/missing-data-in-questionnaires.html> [Accessed 2024-06-27]
20. Van Buuren S, Groothuis-Oudshoorn GOK. Mice: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(1):1-67. [doi: [10.18637/jss.v045.i03](https://doi.org/10.18637/jss.v045.i03)]
21. Kaiser HF. An index of factorial simplicity. *Psychometrika.* Mar 1974;39(1):31-36. [doi: [10.1007/BF02291575](https://doi.org/10.1007/BF02291575)]
22. Watkins MW. Exploratory factor analysis: a guide to best practice. *J Black Psychol.* Apr 2018;44(3):219-246. [doi: [10.1177/0095798418771807](https://doi.org/10.1177/0095798418771807)]
23. Field A, Miles J, Field Z. *Discovering Statistics Using R.* SAGE Publications; 2012.
24. Flora DB, Labrish C, Chalmers RP. Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Front Psychol.* 2012;3:55. [doi: [10.3389/fpsyg.2012.00055](https://doi.org/10.3389/fpsyg.2012.00055)] [Medline: [22403561](https://pubmed.ncbi.nlm.nih.gov/22403561/)]
25. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D. Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. *Struct Equ Model.* Oct 6, 2009;16(4):625-641. [doi: [10.1080/10705510903203573](https://doi.org/10.1080/10705510903203573)]
26. Koğar H, Yılmaz Koğar E. Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Egit Psikol Olcme Deger Derg.* Dec 2015;6(2). [doi: [10.21031/epod.94857](https://doi.org/10.21031/epod.94857)]
27. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* Jan 1999;6(1):1-55. [doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)]
28. Kalkbrenner MT. Alpha, omega, and H internal consistency reliability estimates: reviewing these options and when to use them. *Couns Outcome Res Eval.* Jan 2, 2023;14(1):77-88. [doi: [10.1080/21501378.2021.1940118](https://doi.org/10.1080/21501378.2021.1940118)]
29. Alavi M, Visentin DC, Thapa DK, Hunt GE, Watson R, Cleary M. Chi-square for model fit in confirmatory factor analysis. *J Adv Nurs.* Sep 2020;76(9):2209-2211. [doi: [10.1111/jan.14399](https://doi.org/10.1111/jan.14399)] [Medline: [32323338](https://pubmed.ncbi.nlm.nih.gov/32323338/)]
30. Winters BD, Cvach MM, Bonafide CP, et al. Technological distractions (part 2): a summary of approaches to manage clinical alarms with intent to reduce alarm fatigue. *Crit Care Med.* Jan 2018;46(1):130-137. [doi: [10.1097/CCM.0000000000002803](https://doi.org/10.1097/CCM.0000000000002803)] [Medline: [29112077](https://pubmed.ncbi.nlm.nih.gov/29112077/)]

31. Savalei V. Improving fit indices in structural equation modeling with categorical data. *Multivariate Behav Res.* 2021;56(3):390-407. [doi: [10.1080/00273171.2020.1717922](https://doi.org/10.1080/00273171.2020.1717922)] [Medline: [32054327](https://pubmed.ncbi.nlm.nih.gov/32054327/)]
32. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. *Behav Res Methods.* Feb 2019;51(1):409-428. [doi: [10.3758/s13428-018-1055-2](https://doi.org/10.3758/s13428-018-1055-2)] [Medline: [29869222](https://pubmed.ncbi.nlm.nih.gov/29869222/)]
33. Bliss JP, Gilson RD, Deaton JE. Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics.* Nov 1995;38(11):2300-2312. [doi: [10.1080/00140139508925269](https://doi.org/10.1080/00140139508925269)] [Medline: [7498189](https://pubmed.ncbi.nlm.nih.gov/7498189/)]
34. Dataset for evaluating the construct validity of the Charité Alarm Fatigue Questionnaire. Zenodo. URL: <https://zenodo.org/records/8296934> [Accessed 2024-06-27]

## Abbreviations

**CAFQa:** Charité Alarm Fatigue Questionnaire  
**CFA:** confirmatory factor analysis  
**CFI:** comparative fit index  
**EFA:** exploratory factor analysis  
**ICU:** intensive care unit  
**MAR:** missing at random  
**REDCap:** Research Electronic Data Capture  
**RMSEA:** root mean square error of approximation  
**RNI:** relative noncentrality index  
**SRAF:** self-reported alarm fatigue  
**SRMR:** standardized root mean squared residual  
**TLI:** Tucker-Lewis index  
**ULS:** unweighted least-squares

*Edited by Daniel Gooch; peer-reviewed by Joseph Schlesinger, Kaio Bin, Keisuke Takano; submitted 28.02.2024; final revised version received 02.05.2024; accepted 02.05.2024; published 08.08.2024*

### *Please cite as:*

*Wunderlich MM, Krampe H, Fuest K, Leicht D, Probst MB, Runge J, Schmid S, Spies C, Weiß B, Balzer F, Poncette AS, CAFQa Study Group Germany*

*Evaluating the Construct Validity of the Charité Alarm Fatigue Questionnaire using Confirmatory Factor Analysis*  
*JMIR Hum Factors* 2024;11:e57658

URL: <https://humanfactors.jmir.org/2024/1/e57658>

doi: [10.2196/57658](https://doi.org/10.2196/57658)

© Maximilian Markus Wunderlich, Henning Krampe, Kristina Fuest, Dominik Leicht, Moriz Benedikt Probst, Julian Runge, Sebastian Schmid, Claudia Spies, Björn Weiß, Felix Balzer, Akira-Sebastian Poncette, CAFQa Study Group Germany. Originally published in *JMIR Human Factors* (<https://humanfactors.jmir.org>), 08.08.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Human Factors*, is properly cited. The complete bibliographic information, a link to the original publication on <https://humanfactors.jmir.org>, as well as this copyright and license information must be included.