



Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: <http://www.elsevier.com/locate/ergon>

Mayday, Mayday, Mayday: Using salivary cortisol to detect distress (and eustress!) in critical incident training

Stephanie Balters^{a,*}, Joseph W. Geeseman^{a,b}, Ann-Kristin Tveten^c, Hans Petter Hildre^d, Wendy Ju^e, Martin Steinert^f

^a School of Medicine, Stanford University, USA

^b United States Navy, USA

^c Department of Biological Sciences, Norwegian University of Science and Technology, Norway

^d Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Norway

^e Jacobs Technion-Cornell Institute, Cornell Tech, USA

^f Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Norway

ARTICLE INFO

Keywords:

Critical application design

Distress and Eustress

Emergency

Salivary cortisol

Heart rate (variability)

Performance

ABSTRACT

The understanding of stress and its impact on human performance is crucial to mitigate human error in the face of a threat. This is especially the case for critical incidents on a ship bridge, where human error can easily lead to severe danger for crew, cargo, and other vessels. To overcome the current limitations of robust objective stress measures that reliably detect (di-)stress under highly noisy conditions, we set out to explore whether salivary cortisol – the stress biomarker in medicine and psychology – is a valuable complementary assessment tool in a high-stress/emergency context. In a controlled within-subjects experiment ($N = 12$) using a ship bridge simulator, we measured stress levels under three conditions (80 min each): baseline, low stress (open water navigation task in autopilot), and high stress (open water emergency scenario). We sampled salivary cortisol at 10 min intervals in conjunction with heart rate (variability) monitoring, and subjective stress assessments from both participants and expert evaluators. Results validate salivary cortisol as a successful tool for detecting distress. Unlike the other stress measures, salivary cortisol strongly correlated with expert stress assessments ($r = 0.856$) and overt stress behavior like instances of freezing and missing response cues. Surprisingly, data further revealed decreased salivary cortisol across periods of self-assessed improved performance (i.e., eustress). In fact, data suggests an inverted u-relationship between performance and salivary cortisol. The findings have direct implications for the vast field of emergency training, and serve as a first important validation and benchmark to proceed with real life applications.

1. Introduction

In 2014, the 6,825-ton ferry Sewol capsized. More than 300 people, most of them school children on a field trip, died aboard the ship. The official investigation revealed that the catastrophe was a result of human error. The safety officer and helmsman (i.e., person who steers the boat) enmeshed in a sequence of navigation errors, which eventually culminated in a sudden, extreme turning maneuver causing the ship to capsize (Safety4Sea, 2017). Crucially, the captain and crew of the ship panicked and abandoned the ship, fleeing on rescue boats rather than helping to evacuate the ship's passengers. Stress, and distress in particular, greatly contributed not only to the sinking, but also to the extent of the death

toll.

As critical application and system designers, we are interested in detecting, monitoring, and predicting human performance and stress in emergency situations to design and engineer systems and interfaces that mitigate human error and facilitate human performance (Balters and Steinert, 2017). Indeed, stress is a fundamental component to those situations, but excessive stress can cause people to perform poorly at the very moment that they need to perform well. Alerts and control interfaces for support systems need to be designed to reduce stress, or at least, to enable better performance in the face of exceeding stress. For example, during critical moments, an interface can filter information to include only high-priority information to avoid information overload or

* Corresponding author.

E-mail address: balters@stanford.edu (S. Balters).

<https://doi.org/10.1016/j.ergon.2020.102975>

Received 21 November 2019; Received in revised form 27 March 2020; Accepted 11 May 2020

Available online 13 June 2020

0169-8141/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

a critical alarm sound may shift to a less jarring, but still noticeable, visual stimulus once acknowledged. The understanding of stress during critical emergency incidents is limited. Much of what we know about stress detection and psycho-physiological responses might not hold true for highly critical emergency situations.

In order to systematically shed light onto these unknown areas, we require reliable stress measurement tools. Crucially, however, a ship bridge is a critical environment for most noise-sensitive physiological measures. We must consider that during critical marine events, ship bridge operators are highly mobile and active. For example, they provide a high proportion of mission-critical communications – they walk between widely-spaced control stations on the ship bridge; they frequently change their body posture; and they rapidly move their heads and arms in order to control and interact with the instrument cluster – all of which has been shown to induce (severe) artefacts in a variety of different physiological measures such as electrocardiogram (ECG) or electrodermal activity (EDA) (Healey and Picard, 2005; Hernandez et al., 2011; Salvendy, 2012). Our community requires, therefore, complementary robust objective stress measures that are not overly sensitive to body motion and changes in the measurement environment.

Salivary cortisol is such an objective and robust measure of stress that is often considered *the* golden standard stress biomarker in clinical sciences and psychology. Notably, cortisol has been shown to be substantially released only upon threatening psychological stressors. Therewith, cortisol becomes a unique tool that allows to capture the valence dimension of psychological stress (e.g., “negative threat” versus “positive challenge”). In other words, it allows researchers to physiologically, hence objectively, measure a subjective response to a psychological stressor. Despite this valuable characteristic, its application, however, has not reached prominence in the communities around Human Factors and Ergonomics (HFE), Human-Computer Interaction (HCI), and Human-Machine Interaction (HMI). Potential reasons for its slow adoption might be that: cortisol is generally only released in (major) threatening situations, hence a majority of HFE/HCI/HMI stressors might not be strong enough to elicit a cortisol response; there is a considerable time delay between stressful event and peak cortisol response; and the associated efforts required for biochemical analysis of cortisol can be cost-prohibitive and have not yet allowed for real-time monitoring. Recent advances in wearable technology may provide the capability for real-time analysis of cortisol concentrations via sweat (<https://news.stanford.edu/2018/07/20/wearable-device-measures-cortisol-sweat/>), and accordingly, may provide a useful tool in assessing the onset of a distress response to a slowly unfolding maritime emergency. A realization of this technology could transform cortisol into a complementary unobtrusive, constant, robust, cheap, and foremostly, objective (di-)stress measure in critical incident training and operations. As a first vital step into this direction, it is essential to test whether the stress response of professionals within a training setting is high enough to elicit cortisol. With a focus on this challenge, we set out to explore following research questions:

RQ1: Is salivary cortisol a valid metric for measuring stress in critical incident training and scenarios?

RQ2: Does salivary cortisol provide any additional objective information about the psycho-physiological (di-)stress response in comparison to more accessible measures such as heart rate and heart rate variability?

In the efforts to answering these questions, our specific contributions are three-fold:

1. We validate salivary cortisol as a tool for critical incident training and operations. We provide the community with an important first stressor-benchmark for future studies.
2. We demonstrate that salivary cortisol provides essential complementary information in conjunction with participants' self-

assessments and other physiological sensors. Salivary cortisol correlated with expert stress assessment and instances of non-coping behavior such as freezing or sporadic movements.

3. We revealed an unexpected finding indicating that during “positive stress” (i.e., eustress), salivary cortisol decreased.

As marine education and (re-) certification is almost exclusively done in simulator environments, the results of this study have direct implication for the marine education sector; and can serve as first indicators for real world applications.

2. Background and prior work

The stress response is an evolutionary mechanism that mobilizes physical resources to help humans cope with challenges and life-threatening situations (Cohen et al., 2007). The American Psychological Association differentiates between acute stress, episodic acute stress, and chronic stress (APA, 2019). In that sense, we understand stress as a psycho-physiological response to a stressor. While acute stress is a short-term response to a stressor from which individuals almost fully recover, experiencing frequent and constantly elevated levels of stress (episodic acute and chronic stress) is associated with a variety of pathophysiological risks including cardiovascular diseases and immune deficiencies, impaired quality of life, and shortened life expectancy (Lazarus, 1966; McEwen, 1998; Cohen et al., 2007). In this study, we investigate instances of acute stress.

In emergency response situations (Hiltz et al., 2011; Starbird and Palen, 2013; Bendak and Rashid, 2020) where people are controlling potentially dangerous machines (Jiang et al., 2004; Healey and Picard, 2005), or when people are engaging in high stakes decision-making or complex coordination tasks (MacKay, 1999; Langan-Fox et al., 2009), stress may have unfavorable consequences such as loss of concentration, risky and irrational decision-making, or choking under pressure (BBC, 2012; CNN, 2015; Guardian, 2014). It has been repeatedly proposed that the relationship between stress/arousal and performance follows an inverted u-shape (Hebb, 1955; Teigen, 1994). In other words, as a person experiences mild to moderate stress/arousal, their performance on a given task will improve (i.e., eustress) until the stress/arousal exceeds their idiosyncratic threshold at which time their performance decreases (i.e., distress). This phenomenon is yet to be explored in highly critical (life-threatening) emergency situations, in which humans anecdotally report about instant “breaking points” rather than continuous decreases of performance.

The underlying mechanisms amongst the psychological, physiological, and behavioral responses to a stressor are highly complex, and in this paper, we aim to reveal a novel relationship between the physiological response to a high stress situation and human performance. While the central aim of this paper is to test for the viability of salivary cortisol, our findings evoked a discussion on the relationship between stress and performance. We would therefore like to introduce common language throughout this article. Following Selye's definition (Selye, 1974, 1976), we understand distress as a state when stress is not resolved by physiological and/or psychological coping or adaptation, and is characterized with negative perception (i.e., “threat”) and a decrease in performance outcome. Eustress on the other hand may enhance one's functions and is characterized by positive perception (i.e., “challenge”) and enhanced performance outcome.

2.1. Psycho-physiological stress measurements

Stress responses have been captured using subjective self-reports and/or objective physiological signals. For example, subjective stress levels have been queried via the Perceived Stress Scale (PSS) (Roberti et al., 2006), the Circumplex Model of Affect (Russell, 1980), and/or Affect Grid (Russell et al., 1989) that sample stress as a combination of arousal and valence state within a nine-increment Cartesian grid

structure. Questionnaires have the advantage of being an inexpensive, easy, and established way to collect empirical data. On the other hand, surveys may be subject to error if they are performed mid-experiment because they interrupt the task, or surveys may risk “affective memory” error if performed after the completion of a task. Moreover, questionnaires are subjective and risk unintended variance due to differences in individual’s semantic interpretation of the word “stress”. For example, sometimes stress is understood to be an external stressor, and other times it may be understood as the physiological and/or psychological response to a stressor (Le Fevre et al., 2003). Analysis and debriefing of real-life critical emergency or highly stressful events currently rely on post-event self-reports from the individuals that experienced the stressors (Kowalski-Trakofler et al., 2003). Unfortunately, self-reporting, especially during highly stressful events, is prone to memory errors and is not a reliable means for understanding the stress response (Schacter, 1999).

With the intent to shed more “objective” light on the underlying mechanisms of stress, researchers developed the capability to collect and analyze real-time physiological responses in recent years. Within the context of affective computing/engineering, physiological measures have been applied to gauge the autonomic stress response, including pupillary response (Barreto et al., 2007; Pedrotti et al., 2014; Liu and Yu, 2017), facial temperature measure (Baltaci and Gokcay, 2016; Yun et al., 2009), blood pressure (Hjortskov et al., 2004), heart rate (HR) and heart rate variability (HRV) (Bernardi et al., 2000; Healey and Picard, 2005; Shakouri et al., 2018), breathing rate (Paredes et al., 2018b; Balters et al., 2018) and galvanic skin response/electrodermal activity (EDA) (Healey and Picard, 2005). The major advantages of these physiological tools are their objective nature and their relative non-intrusiveness, which allows continuous data collection. Shortcomings of these metrics are their sensitivity to confounding factors that sometimes introduce measurement errors, such as the effects of changing (external) light on the eye and pupillary response, body movement induced noise in heart rate and heart rate variability data, or increase in skin conductance due to increased room temperature instead of a stressor stimulus.

Additionally, researchers have been able to detect stress responses via changes in the somatic nervous system. For example, psycho-physiological stress has been detected via the activation of the trapezius muscle (Lundberg et al., 1994; Wixted and O’Sullivan, 2018); and researchers have also indirectly measured stress based on changes in muscle activation, such as via the click force on a computer mouse (Wahlström, 2005; Sun et al., 2014), keyboard (Hernandez et al., 2014), or steering wheel (Paredes et al., 2018a). The mentioned somatic stress measurements are advantageous because they are non-intrusive since they are measurements collected directly from the objects on which people interact. These types of measurement tools are inherent parts of the environment, and therefore, avoid potential biasing effects that are more common with externally introduced measures.

2.2. Salivary cortisol as stress measurement

Salivary cortisol is a well-accepted tool, often *the* gold standard, in psycho-physiology and medicine to measure stress. Released by the adrenal cortex, cortisol is a lipid soluble steroid hormone that is released upon exposure to stressful environmental stimuli (Kirschbaum and Hellhammer, 1989; Dickerson and Kemeny, 2004). Some examples include public speaking/arithmetic tasks (Bassett et al., 1987; Kirschbaum et al., 1993) and final academic examinations (Kirschbaum and Hellhammer, 1989). Cortisol is also released in response to physiological stressors, such as prolonged exercise (Tsigos and Chrousos, 2002; MichaelEgan Alison and Carl, 2004). It is the main product of the hypothalamic-pituitary-adrenal (HPA) axis that, along with the sympathetic activation, is the main actor of the physiological stress response (Tsigos and Chrousos, 2002).

From a stress measurement point of view, cortisol serves as an

indicator of HPA axis activation, signaling that an individual experienced an environmental *stressor*. Internalized psychological stressors, however, require that specific characteristics of a psychological stressor are required to activate the HPA axis. The potential of objectively detecting distressing responses is therewith unique with cortisol measurements, as other tools such as EDA, HR, or HRV do not imply a valence distinction that are often found with “distressing psycho-social stimuli” (Kirschbaum and Hellhammer, 1989) and “social-evaluative threat” (Dickerson and Kemeny, 2004).

The use of salivary cortisol implies several methodological constraints, such as that a stress/threat threshold needs to be exceeded in order for cortisol to be released (Kirschbaum and Hellhammer, 1989; Dickerson and Kemeny, 2004); there is a delay in peak response (ranging approximately between 20 and 40 min depending on stressor-stimulus) for cortisol, resulting in long experimental procedures (Dickerson and Kemeny, 2004; Kirschbaum and Hellhammer, 2000); there exists bio-chemical interplay with other bodily functions related to circadian rhythm (Kirschbaum and Hellhammer, 2000; Dickerson and Kemeny, 2004), food intake (Huether and McCance, 2015; Pramanik, 2015), and caffeine and nicotine consumption (Lovallo et al., 2005; Kirschbaum and Hellhammer, 1989), requiring instructions/control prior to experimentation; and considerable efforts and costs for sampling and bio-chemical analysis of salivary probes. Compared to other physiological measures such as EDA or ECG, however, salivary cortisol is robust against physiological artefacts (e.g., resulting from body movements and/or external factors like lighting conditions or mild temperature fluctuation). Therefore, although it is a somewhat costly experimental methodology in regards to money and time, due to its uniqueness as objective distress marker and its robustness, salivary cortisol gains its usefulness for the emergency context described in this paper.

Salivary cortisol has only rarely been applied in the HCI/HMI context. More specifically, a single HCI/HMI study exists exploring a longitudinal assessment of chronic stress (Fujigaki and Mori, 1997). Even more broadly across salivary cortisol literature, real world emergency settings (acute stress), such as measuring cortisol levels of emergency personnel on duty (Yang et al., 2001; Baig et al., 2006; Regehr et al., 2008) or during military training (Taylor et al., 2012) are rare. To our knowledge, the work described in this paper is the first study with real-time measures using a controlled HFE/HCI/HMI-related critical-incident (i.e., “realistic”) stressor.

3. Method

To achieve a high level of ecological validity and meeting the standards of the certified ship simulator training facility, we designed the experimental methodology along with maritime experts at the facility. The aim was to design an environmental stressor that maximizes the corresponding stress response within the constraints of the experimental methodology. As our final experimental design, we converged onto a controlled experiment in which professional ship captains ($N = 12$) performed three tasks over the course of three days – a physiological baseline task, a low-arousal control task, and an emergency scenario on open water – each 80 min long. We sampled salivary cortisol at 10-min intervals, collected data from ECG and self-reports (level of arousal, valence, and stress), and experts provided evaluations as benchmark measures to validate our emergency scenario and other stimuli.

3.1. Participants

We recruited 13 male participants with Deck Officer Certificates (see Fig. 1). One participant interrupted the stress-inducing period 8 min in and elected to withdraw from the experiment. He later explained that the assessment situation was too risky in times of high lay-off rates in the Norwegian maritime sector. The age of the participants ranged from 24 to 56 years ($M = 40.5$ years; $SD = 10.87$), while the reported years of experience as a Watch Keeping Navigator ranged from 1 to 30 years (M



Fig. 1. Ship bridge simulator as locus for critical incident training.

= 12.5 years; $SD = 10.34$). The self-reported mean experience of navigating on the open sea, measured on a scale from 0% to 100% (in increments of 5%) was $M = 81.8\%$ ($SD = 21.71$), whereas the reported mean experience of simulator navigation was $M = 59.1\%$ ($SD = 24.88$). All participants stated to have normal or corrected-to-normal vision and normal hearing. They also indicated that they did not consume any regular medication (including antihistamine medications), did not have any trauma or psychological disorders including depression, and did not suffer from chronic augmented stress levels. Participants in this study were highly skilled officers who operated the world's most sophisticated and technically challenging vessels in the toughest conditions to be found in the maritime environment such as arctic oil platform support and cruise liners (e.g., responsible for 2500 passengers and 945 crew). We advertised recruitment on the university department's homepage and social media, and compensated participants with a 1000 NOK (approx. 115 USD) Amazon gift card.

3.2. Setting

We chose the Polaris Ship Bridge Simulator (Kongsberg Maritime AS) for the experiment because it provides a seated position that restricts body movements and related physiological data motion artefacts (see Fig. 2). The simulator consists of several systems with their original haptic and digital interfaces, including the (1) Steering System for both manual steering as well as autopilot, (2) Visual System with simulated binoculars on two display screens, (3) Electronic Chart Display and Information System (ECDIS) for route sailing, and (4) Polaris Radar System. Communication devices include (5) a telephone for external communication, (6) very high frequency (VHF) radio for communication with other maritime parties, and (7) handheld ultra-high frequency (UHF) radio for internal communication on-board such as with the engine room, deck, or galley. The visual simulation display system surrounds the ship bridge with a 180-degree field of view. Three projectors provide a seamless simulated navigation environment on a cylindrical display of approximately 4×10 m.



Fig. 2. The captain holds two communication devices in one hand while steering the vessel manually during the stress-inducing emergency situation.

Two wide-angle infrared cameras provide rear and side scenic views of the cabin, and a third camera records the operator's frontal close-up. Participants receive written instructions on a monitor in the front console. We placed a tablet for filling out a questionnaire on the middle console along with a folder including technical information about the vessel and a crew list.

3.3. Experimental procedure and protocol

The entire experiment consists of three different tasks – a physiological baseline task (Task 1), a control task (Task 2), and stressor-inducing task (Task 3) – each lasting for 80 min. To control for physiological preconditioning, the participants conduct the experiment on three consecutive days of the same week – a Tuesday, Wednesday, and Thursday – at exactly the same time of day with slots from either 12–2 pm, 2–4 pm, 4–6 pm, or 6–8 pm. Because of the circadian fluctuation of cortisol (Kirschbaum et al., 1993), we set the earliest start of an individual's experiment to 12 p.m. and the latest ended at 8 p.m. To avoid order effects, we randomize the tasks order between participants, resulting in six different order variations. All written and oral communication during the experiment is in English.

Before Study. Prior to the experiment, we instruct the participants not to eat or drink warm beverages (especially caffeine/energizing drinks and food), take warm baths/showers, sleep, brush their teeth (toothpaste includes citric acid, which stimulates saliva production and potentially alters the salivary cortisol concentration), perform heavy exercise, smoke or snus less than 1 h before the experiment, as these activities might influence their physiological reactivity during the experiment (Kirschbaum and Hellhammer, 1989; Dickerson and Kemeny, 2004; Huether and McCance, 2015). Furthermore, we ask the participants to complete a prescreening questionnaire inquiring about their demographics and on-sea experience, as well as possible physiological and/or psychological disorders, regular medications (e.g., antihistamine medications, beta-blockers, or contraceptives) or chronic high stress levels, as these are possible confounding parameters for salivary cortisol.

During Study. To ensure all participants undergo the same procedure throughout the three day experiment, we follow a strict protocol. On the first day, the participant is given the study and video consent form to read and sign. Each day, participants rinse their mouths out five times with plain water to remove any food residue. After the attachment of ECG electrodes, the experimenter (E1) instructs the participants to follow the descriptions on the instruction screen, and leaves the simulator space and joins the second experimenter (E2) in the instruction area (Fig. 5). For Tasks 2 and 3, participants undergo a familiarization process (on both days): After a brief description of the supply vessel's particulars (i.e., 85 m), participants conduct a standardized procedure of approximately 8 min, allowing the captain to familiarize himself with the vessel's steering, camera, and communication systems. Since we induce (di-) stress in professionals, certified trainers debrief the participants after the experiment per safety protocol.

3.4. Experimental tasks and stressors

Prior to the task instruction, participants give a baseline saliva sample 1 and fill out baseline questionnaire 1. During each task, the instruction screen prompts the participants to give a saliva sample and answer the related questionnaire every 10 min, with a total of nine saliva samples and questionnaire inputs each (Fig. 3).

Physiological Baseline Task 1. We designed Task 1 to generate a reference measurement. The simulator electronics are switched off (apart from the instruction screen), and E1 instructs participants to relax, stay seated, and keep their eyes open. We provide maritime literature in English – chosen to be emotionally neutral input – and invite the participants to read it if they desire.

Control Task 2. We designed Task 2 to be a low-arousal open water navigation task. The instruction screen prompts the participant to be the captain of a crew of 14 onboard the same supply vessel with which he has been familiarized. The vessel is enroute to Bergen from Ålesund in the Norwegian Sea, currently on the open water position (N 62° 25.4' E 005° 18.8' – <https://goo.gl/CnwSvP> – course 219°; speed 14.4 knots) with an approximate expected time of arrival (ETA) of 16 h. Weather conditions are stable: a westerly wind of 10 knots and cloudy sky with good visibility. The captain is assigned the navigation task of following the provided ECDIS route for open water passage to Bergen in autopilot mode (Fig. 4). The captain's tasks are thus reduced to observing the open sea environment, mainly with the help of the digital binoculars, and checking the vessel's instruments and radar. To provide realism for the task, we programmed two sailing boats to cross at minutes 5 and 65 at a distance of 10 km (only visible with radar). To control further for the impact of speaking on heart rate measurements due to underlying sinus arrhythmia (Berntson et al., 1993), we include a communication sequence, designed to be of low arousal/demand: at minute 22, E2 calls as the "Office" via the external phone reporting about an incomplete crew list. E2 asks the captain to read out the names and destination airports of 14 crew members from the crew list to book the flight tickets home for the crew. E2 strictly follows a written protocol.

Stress-inducing Task 3. We designed Task 3 to match the low-arousal open sea passage of Task 2, except for a stress inducing emergency sequence from minutes 22 to 30. Within the same weather conditions, the vessel is enroute to Kristiansund from Stavanger in the Norwegian Sea in autopilot mode, starting on an open water position west of Vaagsøy (N 62°00.0' E 004° 44.5' – <https://goo.gl/NdHLQ2> – course 024°, speed 14.4 knots) with an approximate ETA of 16 h.

At minute 22, E1 initiates the emergency scenario with a pre-recorded distress call sent via marine VHF radio from the instructor station from a sinking sailing yacht with fifteen passengers, four of which are already in the water (for access to sound file: <https://goo.gl/KR4E6z>). Since the simulator vessel is the first vessel on sight of the emergency, the captain is therefore assigned the role of Scene Coordinator by the coastal radio coordinator. During the next 8 min, we designed the emergency scenario to keep the level of performance as demanding as possible – mainly by posing communication/event management challenges. The emergency tasks include management/coordination of different tasks/rescue procedures with different parties on different communication devices. These communication devices include: (1) VHF for communication with the vessel in distress, coastal radio station, rescue helicopter, and rescue craft; (2) handheld UHF

radio for communicating with and managing the captain's own crew, including those on deck, in the engine room, and in the galley (hospital); and (3) phone for external communication, such as with the press (due to open emergency communication channel) and office. Causally linked to the emergency scenario, the participant has to manage an engine failure on their own vessel at minute 25:30, which resolves at minute 29:00. Four simulator trainers enact the different parties with dedicated roles following script (see Fig. 5; for access to script: <https://goo.gl/frx6y7>). In order to simulate a reasonable closure of the emergency scenario, we include a rescue helicopter into the scenario. The emergency scenario ends with a helicopter rescuing the 15 persons from the water. We record two prompts of the rescue helicopter prior to the experiment in order to allow realistic background sounds (for access to sound file 1: <https://goo.gl/zin64r> and sound file 2: <https://goo.gl/5tP8sg>). Finally, the coastal radio station prompts the participants that a rescue craft is on its way to tow the abandoned sailing vessel. The participant is able to return to the actual route for the remaining 50 min.

3.5. Measures

We collect a total of six subjective and physiological stress measures during the experiment, namely (1) self-reported level of stress, (2) self-reported level of arousal, (3) self-reported level of valence, (4) heart rate, (5) heart rate variability, and (6) salivary cortisol. Additionally, we ask three marine trainers to grade participants' stress levels by noting overt stress behavior. Finally, a post-experimental questionnaire assesses the individual experiences during the experiment.

Self-reporting Questionnaires. At 10-min intervals across tasks, we administer questionnaires that inquire about (1) the level of perceived stress via a modified version of the Perceived Stress Scale (Roberti et al., 2006) "How stressed do you feel right now? (with 0% = not stressed and 100% very stressed)?" along with (2) the level of arousal and (3) the level of valence to obtain a combined measure of graded stress as defined by the Affect Grid (Russell et al., 1989), with "How aroused are you (with 0% = not aroused and 100% = very aroused)?" and "How pleasant do you feel (with 0% = very unpleasant and 100% = very pleasant)?"

Expert assessment. We ask three marine trainers to independently grade the stress level of individuals on a scale from 0% = "not stressed" and 100% = "very stressed" (with 10% increments) based on video recordings; and to further code for instances of overt stress behavior such as brief freezing, changes in voice pitch, excessive sweating, and rapid uncoordinated movements.

Electrocardiography (ECG) Sensor. We use the wireless Shimmer3 ECG sensor for electrocardiographic data gathering (Shimmer, 2019), with five lead configuration (Oster, 1999). If needed, E1 shaves corresponding chest areas, cleans the skin with alcohol pads, and tapes the cables onto the skin to avoid detachment. We sample at 512 Hz as recommended for heart rate variability analysis (Camm et al., 1996; Baumert et al., 2016).

Sampling Salivary Cortisol. We use the Salivette Code Blue for saliva collection (Sarstedt, 2019). We place the collectors on a storage box situated on the participants' left side within arm's reach in marked order 1–9. E1 familiarizes the participants with the collection procedure prior to the experiment. During the saliva collection time (90 s), the participants fill out the corresponding questionnaire on the tablet. We allow

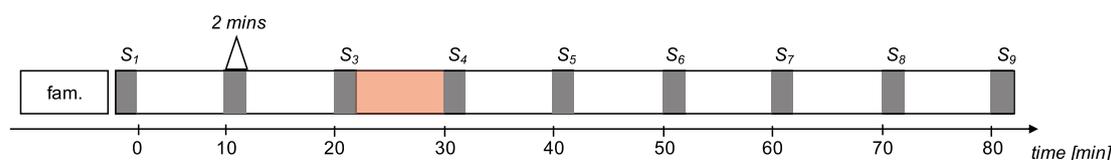


Fig. 3. Experimental timeline, with familiarization process, saliva collection time ($S_1 - S_9$), and communication (Task 2)/emergency sequence (Task 3) marked in red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

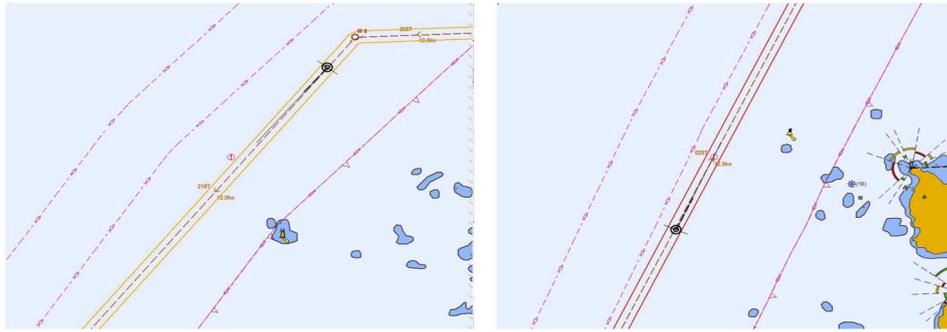


Fig. 4. Provided ECDIS route for task 2 (left) toward south-west and task 3 (right) direction north-east.

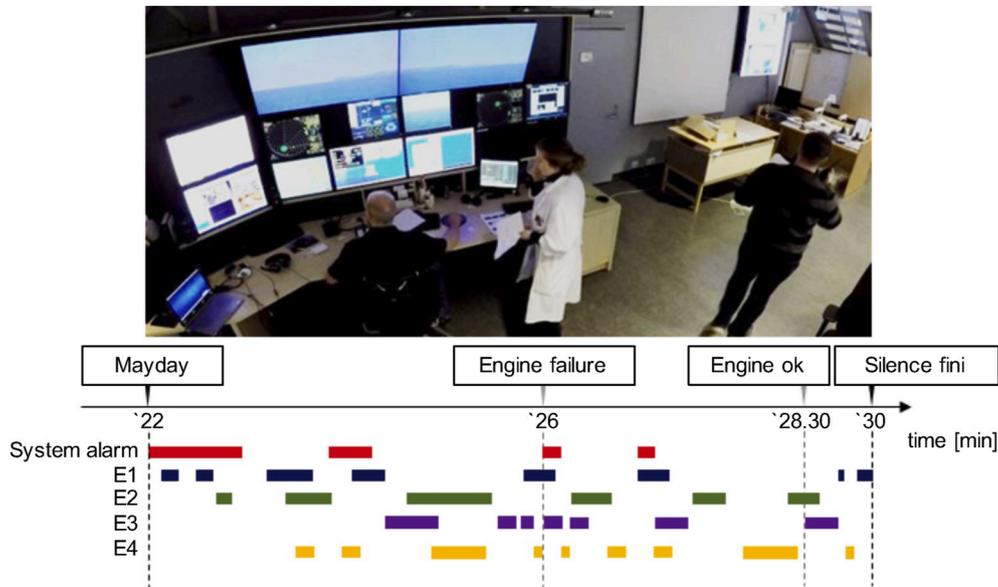


Fig. 5. The instruction area during the emergency task, and visualization of the “emergency script”, with communication sequences for each experimenter (E1-E4) enacting the roles of the coastal radio station, rescue helicopter, rescue vessel, and own crew (deck, engine, and galley (hospital)).

participants to drink plain water at room temperature directly after handling a probe to avoid diluting the subsequent saliva sample.

3.6. Hypotheses

Cortisol literature indicates that a certain “threat/distress threshold” needs to be overcome for cortisol to be released. For this experiment, we assess professionals that are highly trained in stress management, and to some extent, familiar with the stressor-type (e.g., emergency situations). Still, we expect that our experimental design elicits a salivary cortisol stress response in the individuals. Our hypotheses around RQ1 are as follows:

H1.1: There will be a delayed peak in salivary cortisol response upon the emergency event.

H1.2: This increase in salivary cortisol upon the emergency event is higher compared to the two other conditions.

Within our quest for supplementary measures that provide objective information about a subject’s state, we hypothesize that salivary cortisol positively correlates with subjective stress assessments. Our hypotheses around RQ2 are as follows:

H2.1: There is a positive correlation between subjective stress and salivary cortisol.

H2.2: There is a positive correlation between expert stress assessment and salivary cortisol.

4. Data processing

To prepare data for statistical analysis, we initially extract the self-report measures of *perceived stress*, *level of arousal*, and *level of valence* from the questionnaires.

We used Kubios Software to calculate *heart rate* and *heart rate variability* measurements (Tarvainen et al., 2009). ECG data of one participant was lost due to lost connectivity of electrodes during the emergency scenario when the participant produced extended body movements. For the remaining data, we set the artifact correction level in Kubios to “none” and applied a smoothing function with a regularization value of $\Lambda = 500$. We chose the Root Mean Square of the Successive Differences (RMSSD) as the measure for heart rate variability. An increase in RMSSD indicates a decrease in sympathetic nervous system activity (Tarvainen et al., 2009). Additionally, we averaged heart rate values (bpm) and RMSSD values (ms) for each 10-min interval.

After each experimental run, we processed the saliva probes (nine per task and participant) in the in-house biochemical laboratory. After centrifuging, we stored the probes in an $-80\text{ }^{\circ}\text{C}$ ultra-low temperature freezer. On the day of analysis, we thawed the probes and used the Cortisol Enzyme Immunoassay Kit (ArborAssay, 2019a) for colorimetric analysis. We followed the manufacturer’s protocol for analysis

(ArborAssay, 2019a) and measured optical absorbance values with the Synergy HTX Multi-Mode Microplate Reader at 450 nm in room temperature (21 °C). We inserted the absorbance measurements into the Arbor Assay online tool (ArborAssay, 2019b) to derive the salivary cortisol concentration (pg/mL) for each sample. We repeated the measurement twice to mitigate error, and averaged across the two measurements.

5. Analysis and results

Before testing the experimental hypotheses, we completed a stressor-stimulus check to verify whether or not the experimental manipulation was successful in inducing (di-) stress.

5.1. Stressor-stimulus check

In response to the stressor-stimulus, we expected the following changes in the measurements during the emergency task compared to the other two tasks:

- M1: an increase in perceived stress at t₃₀
- M2: an increase in arousal at t₃₀
- M3: a decrease in valence measurement at t₃₀
- M4: an increase in heart rate at t₃₀
- M5: a decrease in RMSSD at t₃₀

We ran non-parametric statistical tests since some measurements were not normally distributed. In Fig. 6, we present the relative time series and provide corresponding boxplots in the appendix (Figure A.11) to facilitate ease in interpreting the results. First, we ran Friedman's tests comparing initial time points (t₀) for each task to verify that there were no systematic differences at the beginning of each task. We found no differences for any measurement.

Next, we analyzed the period of time starting at the onset of the emergency scenario and the following 20 min (t₂₀-t₄₀) for each measurement (M1-M5). Finally, we ran another Friedman's test to determine significant differences in time points t₂₀ and t₃₀ values across the three tasks. Statistical results are summarized in Table 1. Though we conducted non-parametric tests, we added parametric Cohen's d to allow making inferences on effect sizes.

M1. Increase in perceived stress at t₃₀.

Table 1

Statistical results of stressor-stimulus check and H1.1 - H1.2. With "p" values after Bonferroni correction for multi comparison, and Cohen's d as effect size measure.

Measure	Within emergency time series comparison	Between tasks comparison
Stress	$\chi^2(2) = 16.578, p < .001$ t20 and t30, $p < .001$, $d = 1.803$ t30 and t40, $p < .05$, $d = 1.360$	$\chi^2(2) = 18.200, p < .001$ control and emergency, $p < .01$, $d = 1.834$ baseline and emergency, $p < .01$, $d = 1.834$
Arousal	$\chi^2(2) = 19.400, p < .001$ t20 and t30, $p < .001$, $d = 1.352$ t30 and t40, $p < .05$, $d = 1.167$	$\chi^2(2) = 19.950, p < .001$ control and emergency, $p < .01$, $d = 2.237$ baseline and emergency, $p < .01$, $d = 2.312$
Valence	$\chi^2(2) = 9.243, p < .001$ t20 and t30, $p < .05$, $d = 0.650$	no
Heart rate	$\chi^2(2) = 16.545, p < .001$ t20 and t30, $p < .01$, $d = 0.750$ t30 and t40, $p < .05$, $d = 0.740$	$\chi^2(2) = 13.818, p < .001$ control and emergency, $p < .01$, $d = 1.983$ baseline and emergency, $p < .01$, $d = 2.066$
RMSSD	no	$\chi^2(2) = 8.909, p < .05$ control vs. emergency, $p < .01$, $d = 1.093$
Salivary cortisol	$\chi^2(2) = 19.500, p < .001$ t20 and tmax, $p < .01$, $d = 0.999$ tmax and t80, $p < .001$, $d = 1.708$	$\chi^2(2) = 18.500, p < .001$ control and emergency, $p < .05$, $d = 2.110$ baseline and emergency, $p < .001$, $d = 1.489$

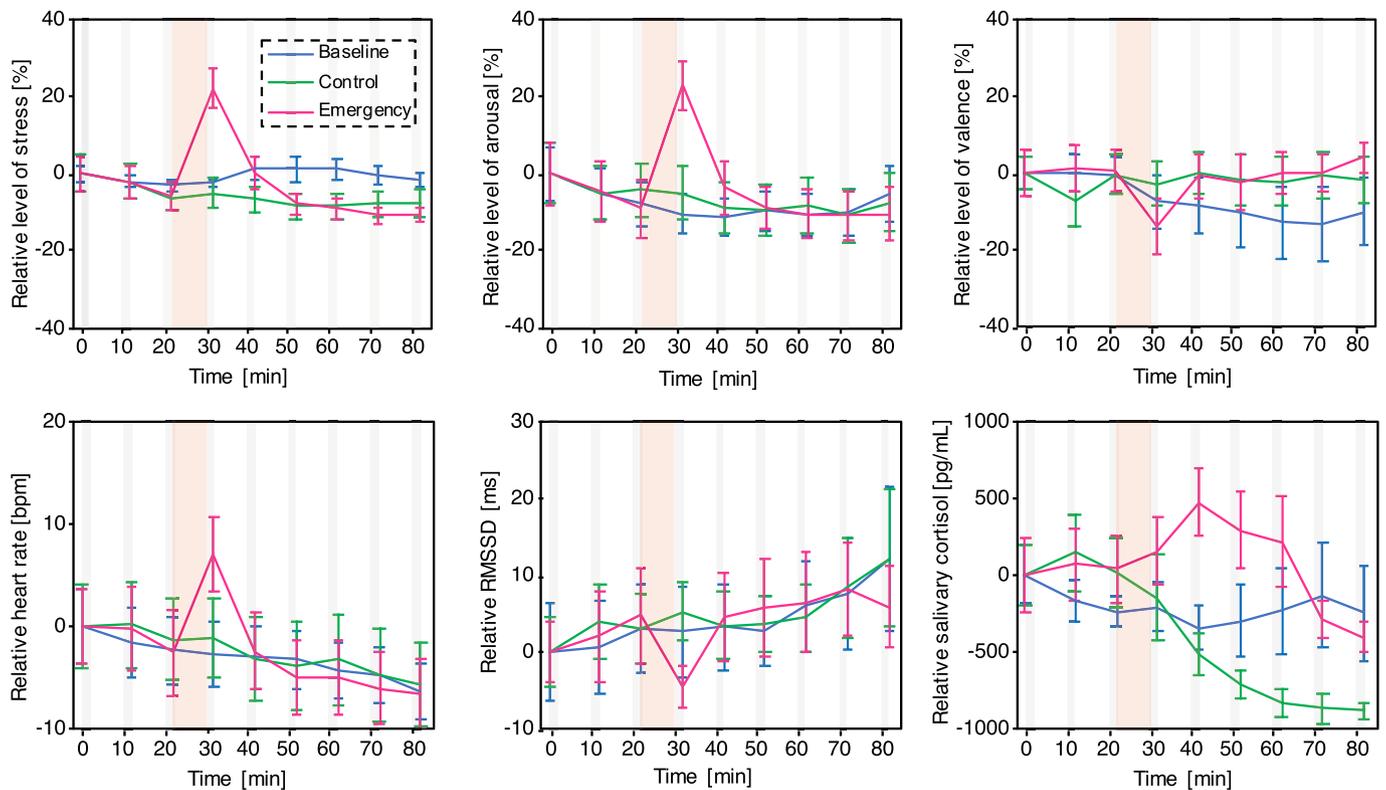


Fig. 6. Relative time series of the six measurements for the three tasks with standard errors of the means (baseline corrected with each measurements' value at t₀). Shaded red areas represent time of stressor-stimulus for Task 3, gray shaded area represents timing of questionnaires and salivary cortisol collection. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

series, subjective stress is increased at t_{30} compared to task compared to both control and baseline task. As expected, these results confirm a brief increase in perceived stress during both t_{20} and t_{40} . Subsequent comparisons between the tasks revealed significantly higher stress levels during the emergency the emergency task at t_{30} . The increase of about 30% (from $M = 8.33\%$, $SD = 12.85\%$ to $M = 36.25\%$, $SD = 17.72\%$), indicates a moderate increase in subjective perceived stress. Notably, four participants reported increases of 50–65%, indicating that some individuals had higher subjective stress responses than others.

M2. Increase in arousal at t_{30} . Similar to the findings above, we found higher arousal levels at t_{30} compared to both t_{20} (with increases from $M = 19.10\%$, $SD = 25.39\%$ to $M = 51.25\%$, $SD = 21.96\%$) and t_{40} ; as well as higher arousal during the emergency task compared both other tasks. Results further indicated a very strong correlation between stress and arousal ($r(25) = 0.886$, $p < .01$).

M3. Decrease in valence at t_{30} . Next, we found a decrease for the valence between t_{20} and t_{30} (from $M = 84.58\%$, $SD = 19.12\%$ to $M = 70.00\%$, $SD = 25.31\%$). Subsequent comparison between the tasks reveal no significant differences. Notably, the four individuals with the highest subjective stress responses have highest decreases in valence (between 20 and 50%). Overall, these results indicate that people perceived the stressor as more “distressing” than a neutral challenge.

M4. Increase in heart rate at t_{30} . Heart rate increased at t_{30} compared to both t_{20} (from $M = 77.39\%$, $SD = 13.71\%$ to $M = 87.00\%$, $SD = 11.85\%$) and t_{40} , and subsequent comparison between the tasks revealed higher heart rate during the emergency scenario compared to both baseline and the control task.

M5. Decrease in RMSSD compared to control task. We did not find significant differences in RMSSD across time, though a tendency was apparent (decreases from $M = 34.64\%$, $SD = 20.55\%$ to $M = 25.32\%$, $SD = 9.05\%$). When comparing tasks, however, we found lower RMSSD in the emergency task compared to the control task. Notably, RMSSD shows high variance in the control task along with a brief (yet not significant) increase at t_{30} . Although trending in the expected direction, we cannot confirm a significant decrease in RMSSD during the emergency task. There is a significant difference between control and emergency task at t_{30} likely due to the increasing trajectory of the control task.

The combined results suggest that the stressor-stimulus was able to elicit a moderate distress response across the cohort, and in four individuals in particular. All Cohen’s d values indicate large effect sizes.

5.2. H1: salivary cortisol response

Cortisol data were not normally distributed; therefore, we used non-parametric tests to investigate hypotheses H1.1 and H1.2. In accordance with the previous analyses, we initially tested whether or not there was an increase in salivary cortisol during the emergency scenario. To account for expected idiosyncratic differences in peak response delay (20–40 min after stressor onset), we identified each individuals’ maximum response in salivary cortisol between time points t_{30} - t_{60} . We compared these values with their own pre-stressor time point t_{20} and final experiment time point t_{80} due to expected variability in delay to individuals returning to baseline. Friedman’s tests revealed increased salivary cortisol at t_{max} ($Mdn = 1871.00$) compared to t_{20} ($Mdn = 932.60$) and t_{80} ($Mdn = 676.70$) with no difference in salivary cortisol between t_{20} and t_{80} . The delay in peak response varied between 10 and 40 min after stressor (i.e., emergency scenario) onset, with a median delay of 20 min. To further test for differences between tasks, we calculated the differences between the timepoint of peak response and pre-stressor level for each task. Friedman’s tests comparing tasks indicated increased salivary cortisol during the emergency task ($Mdn = 664.05$) compared to baseline ($Mdn = -148.45$) as well as the control task ($Mdn = -300.45$). Further, comparable to previous studies (Kirschbaum and Hellhammer, 1989; Dickerson and Kemeny, 2004), we observed a delayed peak in salivary cortisol between 10 and 40 min after stressor-stimulus onset ($Mdn = 20$). All Cohen’s d parameters indicate

large effect sizes (see Table 1).

The results show a significant increase in salivary cortisol during and after the emergency scenario with a delayed peak response between 10 and 40 min after stressor onset. This increase in cortisol during the emergency scenario is statistically higher than the control and baseline tasks. We can therewith confirm both hypotheses H1.1 and H1.2.

5.3. H2: correlation with subjective stress

In an attempt to further assess the viability of salivary cortisol as stress measure in the field, we investigated the relationship between cortisol concentrations and subjective stress assessment from both participants and expert observers. The expert assessment on observed stress levels had an inter-rater reliability of 84%. Again, we used individuals’ cortisol peak responses (that ranged between 10 and 40 min with median 20) and calculated the relative changes compared to their pre-stressor time point t_{20} . Preliminary analysis revealed a linear relationship (see Fig. 7) with normally distributed data (Shapiro-Wilk’s test). We removed one outlier from the data. Pearson’s tests revealed no correlation between subjective grading and change in cortisol. A very strong positive correlation, however, exists between expert grading and salivary cortisol, $r(9) = 0.856$, $p < .01$.

Interestingly, these data suggest that individuals may grade their subjective stress response lower ($M = 38.6$, $SD = 16.4$) than expert assessment ($M = 50.6$, $SD = 18.9$). This difference is not significant ($p = .121$), but the trend is apparent. Notably, the four participants with the highest cortisol readings ($M = 1269.12$ pg/mL, $SD = 190.41$ pg/mL compared to $M = 400.49$ pg/mL, $SD = 252.30$ pg/mL of the remainders) are also the participants with the highest expert stress rating ($M = 69.2\%$, $SD = 1.7\%$ compared to $M = 40.0\%$, $SD = 13.74\%$). Crucially, experts’ notes revealed that these four participants expressed instances of overt non-coping stress behavior such as freezing; rapid uncoordinated movements; changes in voice pitch, clearness of messages, and language use (from English to Norwegian); and missing to respond to prompts. We found neither correlations for salivary cortisol nor expert gradings and the two physiological measures (heart rate and heart rate variability).

Overall, opposing hypothesis H2.1 salivary cortisol shows no correlation with subjective stress assessment. There is, however, a strong positive correlation with expert grading. We can confirm hypothesis H2.2.

5.4. Exploring: decreased cortisol = eustress?

While conducting the analyses for the previous section, an interesting trend emerged for the control task: cortisol levels seemed to

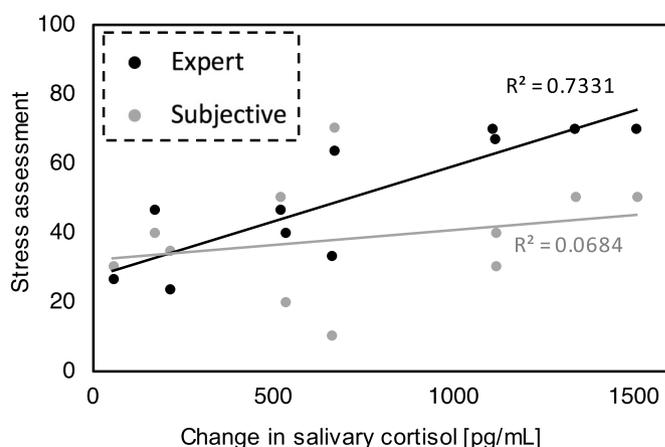


Fig. 7. Correlation between salivary cortisol response and subjective stress assessment and expert observer stress evaluation.

decrease over time. This motivated us to use linear mixed-effects modeling (LMM) to investigate whether it is possible to systematically quantify this trend via curve fitting. We completed the LMM analyses in R using the lme4 package (Nelder and Wedderburn, 1972). As data did not meet the General Linear Model (GLM) requirement of normality, we applied the appropriate link-function to transform the residuals. Box-Cox analyses provided the best link function for our subsequent analyses (see Appendix Figure A.12). The nested model comparisons indicated that the best fitting model for salivary cortisol response over time across task conditions, included an interaction with a quadratic fit for time and task condition, and task condition also as a random factor (i.e., idiosyncratic fit) for participants (see Table 2). This model fits the data well (see Fig. 8) with some participants exhibiting considerable deviation from the model's predicted salivary cortisol response pattern across conditions (see Fig. 9). For example, most participants (~75%) demonstrate a general decrease in cortisol across time with some more pronounced than others, but P02, P08, and P12 show an initial decrease in cortisol during the baseline task, but their cortisol levels rebound to above their initial t_0 cortisol levels. Overall, the results from LMM comparisons indicate that during the control task, salivary cortisol decreased for nearly the entirety of the task (see Fig. 8).

The experimental procedure included questionnaires during both control and emergency tasks at time points t_{30} and t_{80} that asked about subjective performance "How successful were you in accomplishing what you were asked to do?" with values from 0 = "not at all" to 100 = "very much". See Fig. 10 (left) for this relationship. Wilcoxon signed-rank test revealed higher values in subjective performance between control ($Mdn = 90.0$) and emergency task ($Mdn = 45.0$) ($z = 2.983, p < .01$) for time points t_{30} and also for time points t_{80} with higher values for control ($Mdn = 90.0$) and emergency task ($Mdn = 72.5$) ($z = 2.943, p < .01$).

Following the definitions on *eustress* and *distress* in the introduction, our results suggest that participants experienced eustress during the entirety of the control task in contrast to the emergency task where they experienced distress. Overall, the unexpected result in this section

suggests that salivary cortisol decreases whilst experiencing eustress, and the model maintains this strong relationship at both the group and individual level (see Figs. 8 and 9).

In an additional step, we explored the differences in the relationship between subjective personal evaluations (i.e., eustress and distress) and changes in salivary cortisol for both the control and emergency task. We used the changes in salivary cortisol which we calculated for analyses in section 5.2, and plotted the values along with the subjective performance values eustress) and the emergency task (i.e., above identified as distress). As demonstrated with a gray-dotted line, this relationship at timepoint t_{30} . Fig. 10 (right), includes each participants' individual values for the control task (i.e., above identified as suggests an inverted u-shape consistent with other studies (Hebb, 1955; Teigen, 1994; Zettler and Lang, 2015). Crucially, the four individuals with overt stress behavior are seen in the far right corner of distress.

6. Discussion

In this section, we reflect on our contributions with respect to our research questions.

6.1. Salivary cortisol – a valid stress measure

The results revealed a significant increase in salivary cortisol during the emergency scenario, with an expected delayed peak response between 10 and 40 min after stressor onset. This validates salivary cortisol as a valid measure for critical incident training. To provide a clear example of the magnitude of the stress response, we benchmark our results with the most cited salivary cortisol study (Kirschbaum et al., 1993). The "Trier Social Stress Test (TSST)" comprises 10 min of public speaking paired with an arithmetic task (acute stressor) as well as 10 min of a preceding anticipation phase (stressor duration from minute 0 to minute 20). The salivary cortisol response induced by our emergency scenario was only 1/4–1/5 of the salivary cortisol response

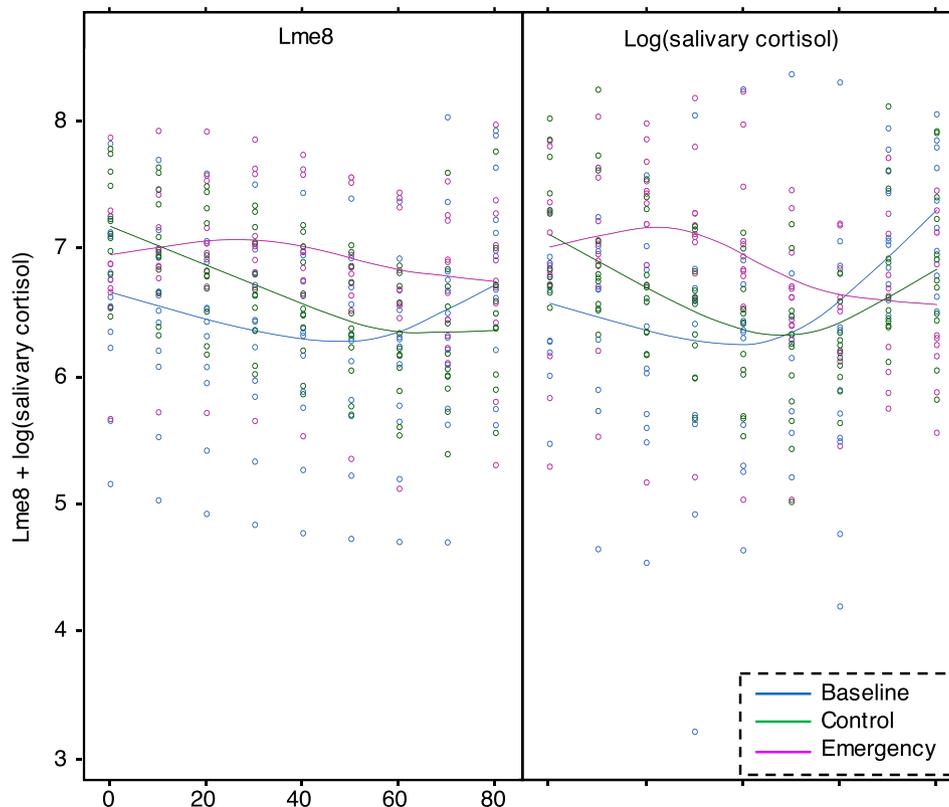


Fig. 8. Group analysis – Model fit for salivary cortisol across time and condition across participants.

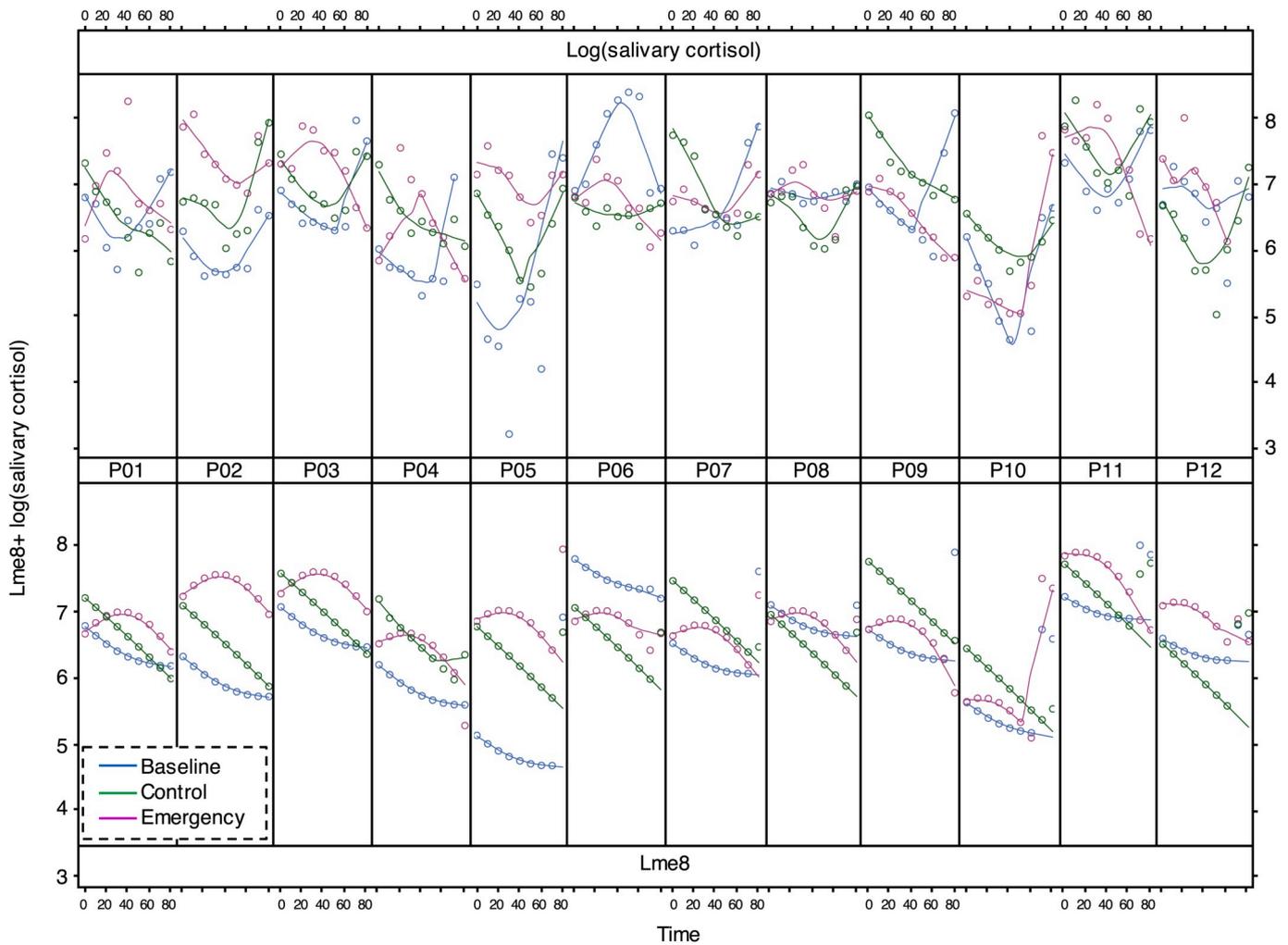


Fig. 9. Individual analysis – Model fit for salivary cortisol across time and condition for each participant. Raw data for individuals are in the top row and model fitted data are in the bottom row.

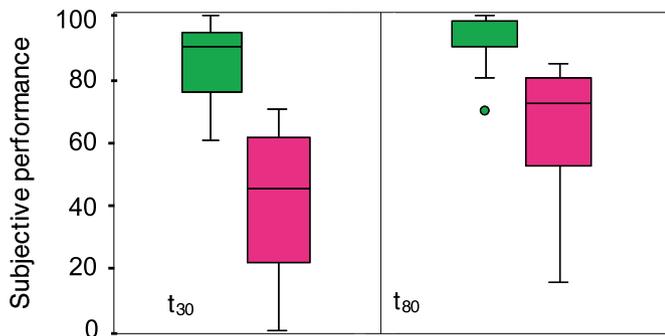


Fig. 10. Participants experienced significantly higher subjective performance during the entirety of the control task compared to the emergency task (left). The control task is therefore by definition a *eustress* scenario compared to the emergency task (i.e., *distress*). The relationship between performance and changes in salivary cortisol follow an inverted u-shape (right).

observed in the TSST study. A comparison in heart rate also shows this 1/4–1/5 ratio. Potential reasons for this reduced physiological response could stem from a combination of shorter stressor length, an absence of a direct negative social feedback, and a sense of maintained control during the emergency scenario from our professional participants (Dickerson and Kemeny, 2004; Kirschbaum et al., 1993). As scenarios of single individuals on a ship bridge are rare, future studies could investigate the

effects of direct social feedback whether through peers or assessing experts on the bridge (Liu and Yu, 2017; Liu et al., 2017).

With respect to simulator education and (re-) certification, the results of this study indicate that the stressor induced within a simulator environment is “threatening/distressing enough” to release salivary cortisol in experienced personnel. We provide, hence, proof that salivary cortisol is a valid tool for emergency training – for the maritime world and beyond (e.g., aviation, nuclear). This is particularly valuable since data processing revealed once more the critical sensitivity of electrode-based sensors for our measurement context – ECG data of one participant were completely lost. The robustness of cortisol is of value indeed. By means of our delayed description of experimental setup and stressor design, we provide an important foundation for stressor-stress response for future HFE/HCI/HMI critical incident experiments. Regarding salivary cortisol literature at large, this is, to our knowledge, the first emergency-scenario HFE/HCI/HMI stress study with an extended time series component (i. e., 10 min sequence).

With respect to real world scenarios, results suggest viability of salivary cortisol as a stress evaluation tool. Though the stress experienced in a lab setting is not as pronounced as a real-world critical incident, the use of simulators during experimentation can provide reliable approximations to the stress experienced in reality (Kowalski-Trakofler et al., 2003; Baddeley, 2000). This can be especially true when assessing professionals that frequently use simulators for training and (re-)certification, like persons working in sea faring, nuclear, or aviation professions (Geeseman, 2016). Given that real-life critical incident training will be a

major challenge due to risk, ethics, and cost, advances in passive, continuous cortisol measures (e.g., wearable device that measures cortisol in sweat) will one day provide new insights to human stress response during real-life critical events.

6.2. Correlation with overt stress behavior

The data revealed a significant correlation between salivary cortisol and experts' grading of stress. The individuals with the highest cortisol response (and expert stress grading) showed further overt stress behaviors like freezing, rapid uncoordinated movements, and missing to respond to prompts. Salivary cortisol could serve as an objective distress measure that replaces or complements expert grading. Importantly, neither the stress self-assessments nor the other physiological measures (i.e., heart rate and heart rate variability) were significantly correlated with salivary cortisol, nor with the expert grading of stress. A reason for the missing correlation with the other ECG-related physiological measures could be that an increase in heart rate has been shown to be lower during threat appraisal compared to challenge appraisal (Tomaka et al., 1997).

Data across participants overall demonstrate that participants underwent a moderate stress response which was high enough to elicit salivary cortisol across the majority of participants. From an operational perspective, however, it is important to detect an individual's response to a given stressor. Salivary cortisol measures in conjunction with expert stress assessments and observed critical behavior demonstrate that it is possible to distill those individuals that are in a critical state of distress. Due to participants' poor performance of subjective stress self-assessment and no significant correlation between expert stress behavior grading and other physiological tools like heart rate and heart rate variability, salivary cortisol becomes a valuable tool to detect distressed individuals.

An individual's stress response depends on many factors, such as the subject's affective state at the time of measurement, their knowledge/experience in terms of the stressors to be coped with, and their skills to deal with emergencies. Physiological responses to the perceived stress may further vary – between participants and, crucially, within the individuals across time. The data of this study showcases the divergence in responses to the same stressor. Using expert assessment as the baseline for stress and performance assessment is currently a standard practice for training and (re-) certification purposes for nautical and aviation professionals. Our data revealed that expert stress ratings had a tendency to be higher than subjective stress ratings. This phenomenon of overestimating one's own affective state and capabilities, is often observed in critical situations and environments (Matthews et al., 2011; Senf et al., 2010; Boe et al., 2015; Amado et al., 2014). Tichy (2004) named this phenomenon “over-optimism” (Tichy, 2004). In contrast, salivary cortisol correlates strongly with the expert stress assessment scale. By means of salivary cortisol we were able to distill those four individuals that underwent stress responses which manifested in poor task performance (i.e., overt stress behavior such as freezing, rapid uncoordinated movements, and missing to respond to prompts).

6.3. Decreased cortisol during eustress

Although not an original hypothesis, we noticed that individual model fits for cortisol response across time and condition for each participant followed fairly consistent patterns – granted, some idiosyncrasies emerged. These individual differences included different latencies to peak cortisol response and the magnitude of cortisol levels. Interestingly, individual data show that despite variations of individual's starting points across the three conditions, the best model fit reveals that cortisol nearly linearly decreased across the control condition for every individual (see lower portion of Fig. 9). We further investigated this occasional unexpected relationship and found that the participants that scored highest on self-evaluation of performance (i.e.,

eustress) in the control condition also had the largest decrease in salivary cortisol over time during the control condition (see P01, P04, P07, and P09).

Our results suggest that salivary cortisol could be a useful tool for evaluating eustress, especially when measured continuously and analyzed in a way that can reveal individual differences. At this stage, it is unclear whether the trajectory is induced by an “acute” eustress response, or whether the shape represents a prolonged eustress response. A first projection shows an inverted relationship between performance (i.e., eustress and distress) and changes in salivary cortisol. The four individuals with overt stress behavior are found at the far right corner of this u-shape. These are interesting first findings and we encourage further investigation into the relationships of cortisol, eustress, and performance. While we used subjective performance measures to separate between the states of eustress and distress, future studies could benefit from including additional objective measures of performance.

6.4. Strengths and limitations

To achieve a high level of realism and ecological validity, we co-designed and executed the experiment with marine training experts. We conducted the experiment in the largest ship training simulator in Norway – it is the main educational institution for cadets as well as for training and (re-)certification of professional mariners in Norway and Scandinavia. Furthermore, we only recruited highly experienced professional ship captains for this study. These professionals operate the world's most sophisticated and technically challenging vessels in the toughest conditions to be found in the maritime environment completing assignments like arctic oil platform support. We spent considerable time addressing confounding parameters that could alter salivary cortisol concentrations. For example, we controlled for stimuli that participants experienced in the hour prior to the start of the experiment and we conducted the experiment for each individual at the same time of day on three consecutive days in the middle of the week to control for circadian rhythm. We further screened participants for any chronic state (e.g., depression), that could affect a salivary cortisol response (Dickerson and Kemeny, 2004).

The main limitation of this study is the relatively small number of participants ($N = 12$). The time requirement of roughly 6 h per experimental session for each participant was the main contributing factor for the small sample size. Future experiments will benefit from a larger cohort to increase statistical power. Another limitation is that we chose not to apply supplementary skin conductance measures – the utility of the latter has been proven in prior field studies, however, with high portions of artefacts (Hernandez et al., 2011).

6.5. Designing a realistic stress experiment

Our design of an ecologically valid emergency task that included a stressor which allowed methodological repeatability for other researchers and statistical comparison between participants, required extensive piloting iterations and testing. To generate a benchmark for our stressor, we aimed for a 10 min stressor period to create an analogue of the Trier Social Stress Test study (Kirschbaum et al., 1993). In order to retain constant data acquisition every 10 min that included a questionnaire and salivary sampling, however, the final stressor length ended up only being 8 min long. Furthermore, due to the 60-min delayed recovery response of cortisol as found in the TSST study (Kirschbaum et al., 1993), we designed the experiment to continue 60 min after stressor onset.

Though the 8-min length of the emergency scenario was rather short for a maritime emergency scenario, this duration, was important from a methodological perspective in order to create a sharp and distinct “stressor-block”. Our analysis of the post-experimental questionnaire inquiring about the perceived realism of the navigation task, “How

realistic was the navigation task (with 0% = not at all and 100% = very much)?" revealed a reduction from 89.2% ($SD = 12.0$) for the control task to 54.2% ($SD = 27.4$) for the emergency task. This decrease in realism was primarily attributed to the shortness (and not the content) of the emergency task across participants (e.g., "the task was realistic but usually emergency situations last longer" – Participant 8). Future experiments will likely adjust the weight of the trade-off between realistic stressor length and related consequences for experimental methodology (e.g., expenses of time and manpower). In order to generalize the findings, we need further benchmark studies. As it is questionable that more standardized stressors such as the n-back task (Kane et al., 2007) and stroop-test (Bench et al., 1993) are threatening enough to release cortisol (Dickerson and Kemeny, 2004), it is important to create further "emergency benchmarks" to better understand which stressors are critical and which are not.

Our biggest challenge was to design a critical incident scenario which we could embed into a low-arousal maritime operation task that would elicit a stress response in professionals with a "sharp onset" and "sharp offset" within a 10-min window of time. Challenges included organizing a large team of professionals during the experimental design phase. For example, we were in constant discussion with trainers and mariners while programming the simulator during the four-week pilot phase, and we tested the scenarios with experts on-sight. Extensive planning of the emergency task required lengthy discussions of several ideas including other stress-inducing scenarios like docking to an oil platform, generating an engine failure with risk to slide onto shore, narrow passage navigation, and fire on board; yet, these critical incident scenarios had either an anticipatory component (hence no sharp onset) or no reasonable transfer back to normal navigation operations (hence no sharp offset). Another requirement that emerged while recruiting participants was the moral responsibility to properly debrief the participants after the experiment by trained and certified simulator trainers rather than ourselves. Overall, however, participants perceived our experimental methodology as realistic, and the experimental task induced negative stress in highly experienced and stress-trained professionals within our simulated critical incident scenario.

6.6. Conclusion

Salivary cortisol is a valid stress measure for critical incident training and experimentation. We provide an important benchmark of an ecologically valid stressor for future HFE/HCI/HMI critical-incident experiments. As mentioned above, this is, to our knowledge, the first study with an instantaneous (i.e., sharp onset and offset) HFE/HCI/HMI critical-incident stressor where data collection endures prior to and after the stressor onset. We found a very strong correlation between salivary cortisol and overt stress behaviors as assessed by training experts ($r = .856$). This is particularly crucial, because observations of overt stress behavior by peers is critical for the overall task performance – "a captain has to be in control at any point in time". Expert observers graded participants with the highest cortisol concentrations as exhibiting the highest number of overt stress behaviors showing instances of non-coping behaviors like freezing, rapid uncoordinated movements, changes in voice pitch and clearness of messages, and failing to respond to prompts. Combined with the inability to accurately self-assess during a subjective stress evaluation and other physiological tools, like heart rate and heart rate variability, did not correlate with expert stress grading of overt stress behavior, salivary cortisol becomes a valuable and unique tool to detect distress. Beyond the scope of distress, our data also suggest that decreased salivary cortisol over time could predict eustress – especially during during non-stressful, but still challenging tasks. Considering advances in wearable technology that could enable monitoring of cortisol concentrations via sweat in real-time, cortisol could become the "go-to" unobtrusive, constant, robust, cheap, and foremost objective (di-) stress measure in critical incident training and operations.

With respect to our research questions **RQ1** and **RQ2**, the results show evidence that (1) salivary cortisol is a valid metric for measuring stress in critical incident training, and (2) salivary cortisol provides additional objective information about the psycho-physiological stress response (e.g., distress and eustress) compared to more traditional measures such as heart rate and heart rate variability.

7. Application

The results of this study show that salivary cortisol is a valid and valuable tool for detecting *distress* in critical high-stress and emergency situations. The findings are directly applicable for the marine critical incident training context – that is not only for education purposes but also training and (re-) certification of professional ship captains. Released in only major distressing incidents, cortisol is suitable as a stress bio-marker for high-stress and emergency contexts. Due to the very strong correlation with expert stress assessment and overt stress behavior, salivary cortisol can play a unique role in stress measurement as it allows to capture the valence dimension of stress (i.e., distress vs. eustress). Recent advances in wearable technology, however, could resolve these issues and thrust cortisol into the forefront of stress monitoring with an unobtrusive, constant, robust, cheap, and most importantly, objective (di-)stress measurement in critical incident training and operations.

Two challenges for the use of salivary cortisol will remain. First, due to its bio-chemical interaction with other bodily functions related to circadian rhythm and/or food intake, these confounding variables require control prior to cortisol monitoring, and/or more holistic monitoring systems are required to account for false positive detection. The delay in peak cortisol response from stress onset is the second challenge for in-situ stress detection. In scenarios that include a sudden stressor onset, monitoring cortisol increase over time could be used as a validating measurement for other more instant measurements. Further research with big data, longitudinal data sampling, and machine learning might be able to untangle these two challenges.

Our findings that suggest a relationship between eustress and a decrease in salivary cortisol, could result in new and promising applications for the critical event and simulator training communities. For example, during normal training events, detecting a slight decrease in a trainee's baseline cortisol level could indicate that the student is highly engaged or "in the zone", and therefore, the training could be augmented to further challenge the student – improving their training experience. Alternatively, a static cortisol level could indicate non-engagement with a training task, demanding for adequate adjustments.

In future experiments and critical incident training – whether simulated or reality – we recommend including salivary cortisol as an augmenting measurement in assessing distress and as a possible mechanism to measure eustress. We recognize limitations in psychological and physiological measures, including salivary cortisol; therefore, we encourage combining multiple measurement tools to "triangulate" results to foster measurement validity.

Funding

This research has been supported by strategic funding of the Department of Ocean Operations and Civil Engineering and of the Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Officer Arnt Håkon Barmen for their support in the design and execution of this study.

We thank Master Mariner Arnt Myrheim-Holm and Navigation

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ergon.2020.102975>.

A. Appendix.

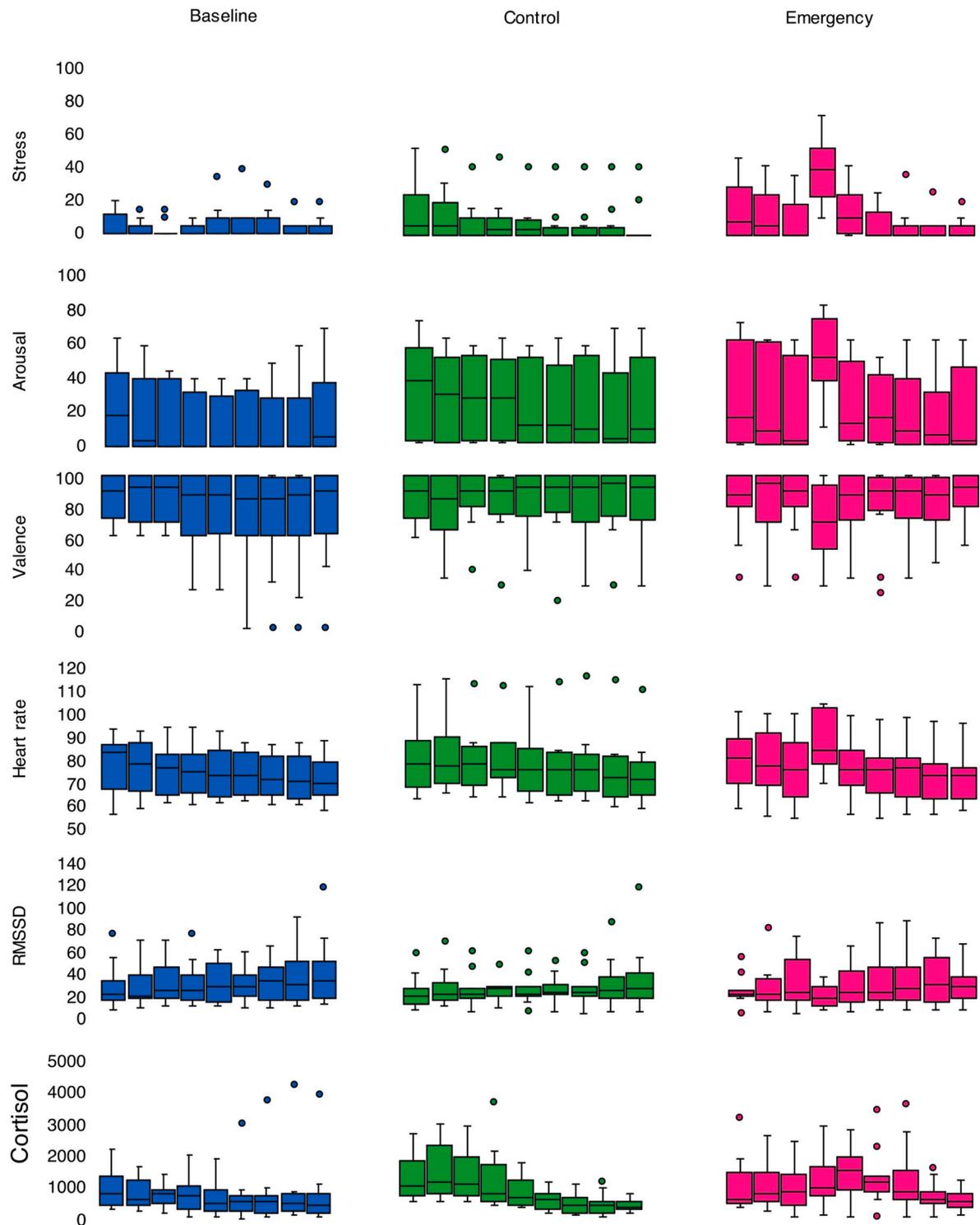


Figure A.11. Boxplots of time series of the six measurements for the three tasks.

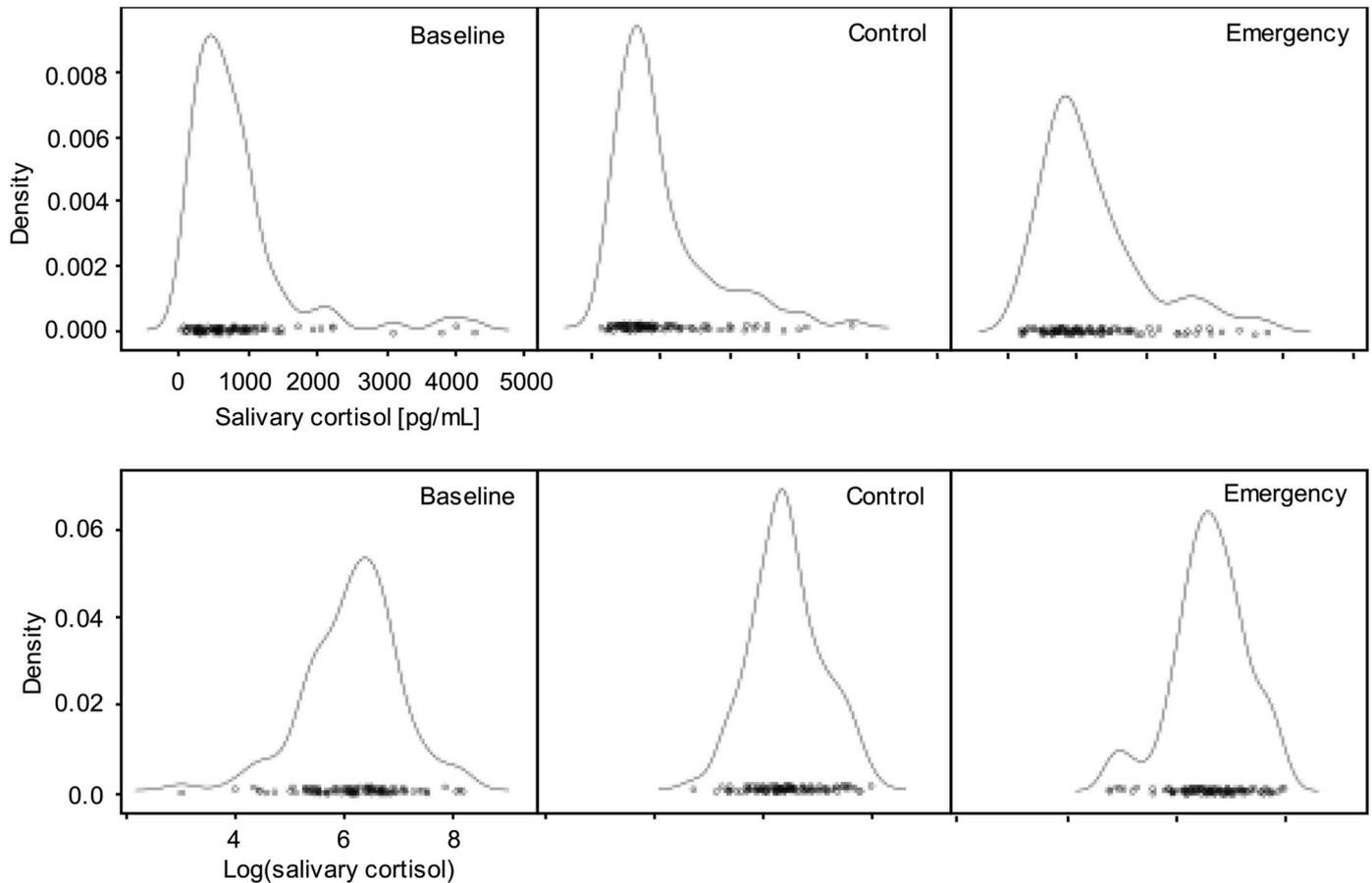


Figure A.12. Box-Cox transform.

Table 2
Model fits.

Models:									
lme1: log(cortisol) ~poly(time,2) * task + (1 participant)									
lme2: log(cortisol) ~poly(time,2) + task + (1 participant)									
lme3: log(cortisol) ~time * task + (1 participant)									
lme4: log(cortisol) ~time + task + (1 participant)									
lme5: log(cortisol) ~time + task + (time participant)									
lme6: log(cortisol) ~time + task + (task participant)									
lme7: log(cortisol) ~poly(time, 2) * task + (time participant)									
lme8: log(cortisol) ~poly(time, 2) * task + (task participant)									
lme9: log(cortisol) ~time * task + (time participant)									
lme10: log(cortisol) ~time * task + (task participant)									
Model	Df	AIC	BIC	Log likelihood	Deviance	Chi square	Chi df	p-value (>ChiSq)	
lme2	6	559.19	581.84	-273.59	547.19				
lme4	6	559.19	581.84	-273.59	547.19	0.000	0		1.000
lme1	8	545.00	575.19	-264.50	529.00	18.193	2		0.0001***
lme3	8	545.00	575.19	-264.50	529.00	0.000	0		1.000
lme5	8	559.40	589.60	-271.70	543.40	0.000	0		1.000
lme9	10	544.67	582.42	-262.34	524.67	18.728	2		8.57 × 10-05***
lme6	11	445.42	486.94	-211.71	423.42	101.249	1		<2.2 × 10-16***
lme10	13	415.74	464.81	-194.87	389.74	33.685	2		4.84 × 10-08***
lme7	16	542.90	603.30	-255.45	510.90	0.000	3		1.000
lme8	16	396.88	457.27	-182.44	364.88	146.028	0		<2.2 × 10-16***

References

Amado, S., Arikan, E., Kaça, G., Koyuncu, M., Turkan, B.N., 2014. How accurately do drivers evaluate their own driving behavior? an on-road observational study. *Accid. Anal. Prev.* 63, 65–73.

APA, 2019. The Different Kinds of Stress. American Psychological Association. <https://www.apa.org/helpcenter/stress-kinds>. created: 201910-21.

ArborAssay, 2019a. Arbor Assay Kit. <http://www.arborassays.com/product/k003-h-cortisol-eia-kit/>. created: 2019-06-13.

ArborAssay, 2019b. Arbor Assay Online. <https://www.myassays.com/arbor-assays-cortisol-enzyme-immunoassay-kit.assay>.

- Baddeley, A., 2000. Selective attention and performance in dangerous environments. *Journal of Human Performance in Extreme Environments* 5, 9.
- Baig, A., Siddiqui, I., Naqvi, H., Sabir, S., Jabbar, J., Shahid, M., 2006. Correlation of serum cortisol levels and stress among medical doctors working in emergency departments. *Journal of the College of Physicians and Surgeons–Pakistan: JCPSP* 16, 576–580.
- Baltaci, S., Gokcay, D., 2016. Stress detection in human–computer interaction: fusion of pupil dilation and facial temperature features. *Int. J. Hum. Comput. Interact.* 32, 956–966. Taylor & Francis.
- Balters, S., Murnane, E.L., Landay, J.A., Paredes, P.E., 2018. Breath booster! exploring in-car, fast-paced breathing interventions to enhance driver arousal state. In: *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pp. 128–137.
- Balters, S., Steinert, M., 2017. Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. *J. Intell. Manuf.* 28, 1585–1607.
- Barreto, A., Zhai, J., Adjouadi, M., 2007. Non-intrusive physiological monitoring for automated stress detection in human-computer interaction. *Hum. Comput. Interact.* 29–38.
- Bassett, J.R., Marshall, P.M., Spillane, R., 1987. The physiological measurement of acute stress (public speaking) in bank employees. *Int. J. Psychophysiol.* 5, 265–273. [https://doi.org/10.1016/0167-8760\(87\)90058-4](https://doi.org/10.1016/0167-8760(87)90058-4).
- Baumert, M., Porta, A., Vos, M.A., Malik, M., Couderc, J.P., Laguna, P., Piccirillo, G., Smith, G.L., Tereshchenko, L.G., Volders, P.G., 2016. Qt interval variability in body surface ecg: measurement, physiological basis, and clinical value: position statement and consensus guidance endorsed by the European heart rhythm association jointly with the esc working group on cardiac cellular electrophysiology. *Europace* 18, 925–944.
- BBC, 2012. No crown office action over capsizing of bourbon dolphin off shetland. <http://www.bbc.com/news/uk-scotland-north-east-orkney-shetland-18393125> created: 2019-06-13.
- Bench, C., Frith, C., Grasby, P., Friston, K., Paulesu, E., Frackowiak, R., Dolan, R.J., 1993. Investigations of the functional anatomy of attention using the stroop test. *Neuropsychologia* 31, 907–922.
- Bendak, S., Rashid, H.S., 2020. Fatigue in aviation: a systematic review of the literature. *Int. J. Ind. Ergon.* 76, 102928.
- Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., Sleight, P., 2000. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *J. Am. Coll. Cardiol.* 35, 1462–1469.
- Berntson, G.G., Cacioppo, J.T., Quigley, K.S., 1993. Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology* 30 (Wiley Online Library).
- Boe, O., Bang, H., Nilsen, F.A., 2015. Developing an observational instrument to measure character strengths during military field exercises. *Procedia Social and Behavioral Sciences* 197, 1126–1133.
- Camm, A.J., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R.J., Coumel, P., Fallen, E.L., Kennedy, H.L., Kleiger, R.E., et al., 1996. Heart rate variability: standards of measurement, physiological interpretation and clinical use. In: *Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology*, vol. 93.
- CNN, 2015. Sewol Ferry Captain Jailed for Murder of 304 Passengers. <http://www.cnn.com/2015/04/28/asia/sewol-ferry-captain-jailed-murder/index.html>. created: 2019-06-13.
- Cohen, S., Janicki-Deverts, D., Miller, G.E., 2007. Psychological stress and disease. *J. Am. Med. Assoc.* 298, 1685–1687.
- Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin. American Psychological Association* 130, 355.
- Fujigaki, Y., Mori, K., 1997. Longitudinal study of work stress among information system professionals. *Int. J. Hum. Comput. Interact.* 9, 369–381.
- Geeseman, J., 2016. Narrowing of attention and functional field of view for helicopter pilots in a degraded visual environment. In: *Proceedings from American Helicopter Society: 72nd Annual Forum*.
- Guardian, 2014. Costa Concordia captain: I brought ship close to shore as treat to passengers. <https://www.theguardian.com/world/2014/dec/02/costa-condordia-captain-brought-ship-close-to-shore-treat-for-passengers> created: 2019-10-21.
- Healey, J., Picard, R., 2005. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transport. Syst.* 6, 156–166.
- Hebb, D.O., 1955. Drives and the CNS (conceptual nervous system). *Psychol. Rev.* 62, 243.
- Hernandez, J., Morris, R.R., Picard, R.W., 2011. Call center stress recognition with person-specific models. *International Conference on Affective Computing and Intelligent Interaction* 125–134.
- Hernandez, J., Paredes, P., Roseway, A., Czerwinski, M., 2014. Under pressure: sensing stress of computer users. *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 51–60.
- Hiltz, S.R., Diaz, P., Mark, G., 2011. Introduction: social media and collaborative systems for crisis management. *ACM Trans. Comput. Hum. Interact.* 18, 18.
- Hjortskov, N., Rissén, D., Blangsted, A.K., Fallentin, N., Lundberg, U., Søgaard, K., 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* 92, 84–89. Springer.
- Huether, S.E., McCance, K.L., 2015. *Understanding Pathophysiology*. Elsevier Health Sciences.
- Jiang, X., Hong, J.I., Takayama, L.A., Landay, J.A., 2004. Ubiquitous computing for firefighters: field studies and prototypes of large displays for incident command. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 679–686.
- Kane, M.J., Conway, A.R., Miura, T.K., Colflesh, G.J., 2007. Working memory, attention control, and the n-back task: a question of construct validity. *J. Exp. Psychol. Learn. Mem. Cognit.* 33, 615.
- Kirschbaum, C., Hellhammer, D., 2000. Salivary Cortisol. *Encyclopedia of Stress* 3.379–383 (2000). PNEC Psychoneuroendocrinology: Supplement 1 25, S14–S14. OCLC: 4930595949.
- Kirschbaum, C., Hellhammer, D.H., 1989. Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology* 22. Karger Publishers).
- Kirschbaum, C., Pirke, K.M., Hellhammer, D.H., 1993. The 'trier social stress test'—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology* 28, 76–81.
- Kowalski-Trakofler, K.M., Vaught, C., Scharf, T., 2003. Judgment and decision making under stress: an overview for emergency managers. *Int. J. Emerg. Manag.* 1, 278–289.
- Langan-Fox, J., Canty, J.M., Sankey, M.J., 2009. Human–automation teams and adaptable control for future air traffic management. *Int. J. Ind. Ergon.* 39, 894–903.
- Lazarus, R.S., 1966. *Psychological Stress and the Coping Process*. McGraw-Hill.
- Le Fevre, M., Matheny, J., Kolt, G.S., 2003. Eustress, distress, and interpretation in occupational stress. *J. Manag. Psychol.* 18, 726–744.
- Liu, N., Yu, R., 2017. Influence of social presence on eye movements in visual search tasks. *Ergonomics* 60, 1667–1681.
- Liu, N., Yu, R., Yang, L., Lin, X., 2017. Gender composition mediates social facilitation effect in co-action condition. *Sci. Rep.* 7, 1–13.
- Lovullo, W.R., Whitsett, T.L., al'Absi, M., Sung, B.H., Vincent, A.S., Wilson, M.F., 2005. Caffeine stimulation of cortisol secretion across the waking hours in relation to caffeine intake levels. *Psychosom. Med.* 67 <https://doi.org/10.1097/01.psy.0000181270.20036.06>.
- Lundberg, U., Kadefors, R., Melin, B., Palmerud, G., Hassmén, P., Engström, M., Dohns, I. E., 1994. Psychophysiological stress and emg activity of the trapezius muscle. *Int. J. Behav. Med.* 1, 354–370.
- MacKay, W.E., 1999. Is paper safer? the role of paper flight strips in air traffic control. *ACM Trans. Comput. Hum. Interact.* 6, 311–340.
- Matthews, M.D., Eid, J., Johnsen, B.H., Boe, O.C., 2011. A comparison of expert ratings and self-assessments of situation awareness during a combat fatigue course. *Mil. Psychol.* 23, 125–136.
- McEwen, B.S., 1998. Stress, adaptation, and disease: allostasis and allostatic load. *Ann. N. Y. Acad. Sci.* 840, 33–44.
- Michael, McGuigan, Egan Alison, D., Carl, Foster, 2004. Salivary cortisol responses and perceived exertion during high intensity and low intensity bouts of resistance exercise. *J. Sports Sci. Med.* 3, 8–15. OCLC: 5535188712.
- Nelder, J.A., Wedderburn, R.W., 1972. Generalized linear models. *J. Roy. Stat. Soc.* 135, 370–384.
- Oster, C., 1999. Improving ecg trace quality. *Biomed. Instrum. Technol.* 34, 219–222.
- Paredes, P.E., Ordonez, F., Ju, W., Landay, J.A., 2018a. Fast & furious: detecting stress with a car steering wheel. <https://doi.org/10.1145/3173574.3174239>.
- Paredes, P.E., Zhou, Y., Hamdan, N.A.H., Balters, S., Murnane, E., Ju, W., Landay, J.A., 2018b. Just breathe: in-car interventions for guided slow breathing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 28.
- Pedrotti, M., Mirzaei, M.A., Tedesco, A., Chardonnet, J.R., Mérianne, F., Benedetto, S., Baccino, T., 2014. Automatic stress classification with pupil diameter analysis. *Int. J. Hum. Comput. Interact.* 30, 220–236. Taylor & Francis.
- Pramanik, D., 2015. *Principles of Physiology*. Jaypee Brothers Medical P, S.I. OCLC: 958098326.
- Regehr, C., LeBlanc, V., Jelley, R.B., Barath, I., 2008. Acute stress and performance in police recruits. *Stress and Health. Journal of the International Society for the Investigation of Stress* 24, 295–303.
- Roberti, J.W., Harrington, L.N., Storch, E.A., 2006. Further psychometric support for the 10-item version of the perceived stress scale. *J. Coll. Counsel.* 9, 135–147.
- Russell, J.A., 1980. A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. <https://doi.org/10.1037/h0077714>.
- Russell, J.A., Weiss, A., Mendelsohn, G.A., 1989. Affect Grid: a single-item scale of pleasure and arousal. *J. Pers. Soc. Psychol.* 57, 493–502. <https://doi.org/10.1037/0022-3514.57.3.493>.
- Safety4Sea, 2017. Sewol sinking: South Korea's ferry disaster. <https://safety4sea.com/cm-sewol-sinking-south-koreas-ferry-disaster/>. created: 2019-06-13.
- Salvendy, G., 2012. *Handbook of Human Factors and Ergonomics*. John Wiley & Sons.
- Sarstedt, 2019. *Salivette code Blue*. <https://www.sarstedt.com/en/products/diagnostic/salivasputum/product/51.1534.500/>. created: 2019-10-21.
- Schacter, D.L., 1999. The seven sins of memory: insights from psychology and cognitive neuroscience. *Am. Psychol.* 54, 182.
- Selye, H., 1974. *Stress without Distress*. New York.
- Selye, H., 1976. *Stress without Distress*, pp. 137–146.
- Senf, B., Brandt, H., Dignass, A., Kleinschmidt, R., Kaiser, J., 2010. Psychosocial distress in acute cancer patients assessed with an expert rating scale. *Support. Care Canc.* 18, 957–965.
- Shakouri, M., Ikuma, L.H., Aghazadeh, F., Nahmens, I., 2018. Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: the case of highway work zones. *Int. J. Ind. Ergon.* 66, 136–145.
- Shimmer, 2019. *Shimmer Sensing ECG. Shimmer Discovery in Motion*. <http://www.shimmersensing.com/products/shimmer3-ecg-sensor>. created: 2019-10-21.
- Starbird, K., Palen, L., 2013. Working and sustaining the virtual disaster desk. In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pp. 491–502.

- Sun, D., Paredes, P., Canny, J., 2014. Moustress: detecting stress from mouse motion. *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 61–70.
- Tarvainen, M.P., Niskanen, J.P., Lipponen, J.A., Ranta-aho, P.O., Karjalainen, P.A., 2009. Kubios HRV – A Software for Advanced Heart Rate Variability Analysis, pp. 1022–1025. https://doi.org/10.1007/978-3-540-89208-3_243.
- Taylor, M.K., Padilla, G.A., Stanfill, K.E., Markham, A.E., Khosravi, J.Y., Dial Ward, M.D., Koehler, M.M., 2012. Effects of dehydroepiandrosterone supplementation during stressful military training: a randomized, controlled, double-blind field study. *Stress* 15, 85–96.
- Teigen, K.H., 1994. Yerkes-dodson: a law for all seasons. *Theor. Psychol.* 4, 525–547.
- Tichy, G., 2004. The over-optimism among experts in assessment and foresight. *Technol. Forecast. Soc. Change* 71, 341–363.
- Tomaka, J., Blascovich, J., Kibler, J., Ernst, J.M., 1997. Cognitive and physiological antecedents of threat and challenge appraisal. *J. Pers. Soc. Psychol.* 73, 63. American Psychological Association.
- Tsigos, C., Chrousos, G.P., 2002. Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. *J. Psychosom. Res.* 53, 865–871.
- Wahlström, J., 2005. Ergonomics, musculoskeletal disorders and computer work. *Occup. Med.* 55 (Oxford University Press).
- Wixted, F., O’Sullivan, L., 2018. Effect of attention demand on upper trapezius muscle activity—a moderated mediation model. *Int. J. Ind. Ergon.* 66, 146–156.
- Yang, Y., Koh, D., Ng, V., Lee, F.C.Y., Chan, G., Dong, F., Chia, S.E., 2001. Salivary cortisol levels and work-related stress among emergency department nurses. *J. Occup. Environ. Med.* 43, 1011–1018.
- Yun, C., Shastri, D., Pavlidis, I., Deng, Z., 2009. O’game, can you feel my frustration?: improving user’s gaming experience via stresscam. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 2195–2204.
- Zettler, I., Lang, J.W., 2015. Employees’ political skill and job performance: an inverted u-shaped relation? *Appl. Psychol.* 64, 541–577.



Dr. Ann-Kristin Tveten is an associate professor in biomedical laboratory sciences/ biotechnology at Institute of biological sciences, NTNU. She is currently working within the field of applied biotechnology, establishing methodology for genetic analysis of microbiota and hormone analysis. Her background as biomedical laboratory scientist and PhD in molecular microbiology has given her a field of expertise within biomedical analysis that she actively applies in laboratory experiments and teaching.



Dr. Hans Petter Hildre is professor and head of department at Department of Ocean Operations and Civil Engineering at the Norwegian University of Science and Technology (NTNU). He is teaching product design and system architecture design. Hans Petter is Centre Director for Centre for Research Driven Innovation (SFI-MOVE) within marine operations. This is cooperation between NTNU, SINTEF, University Sao Paulo and 15 companies at the west coast of Norway.



Dr. Wendy Ju is an Assistant Professor at the Information Science Jacobs Cornell Technion Institute at Cornell Tech. Her research focus lies on interaction with autonomous systems such as automobiles and robots. She graduated with her PhD in Mechanical Engineering, from Stanford University in 2008. Before coming to Cornell Tech, she was the Executive Director for the Center for Design Research at Stanford University.



Dr. Martin Steinert is a Professor of Engineering Design and Innovation at the Department of Mechanical and Industrial Engineering at the Norwegian University of Science and Technology (NTNU). He teaches and researches fuzzy front-end engineering for radical new product/service/system concepts with a special focus onto physiology sensors. He graduated with his Ph.D from the University of Fribourg in 2006. Before coming to NTNU, he was Deputy Director at the Center for Design Research and at the d.research program (Hasso Plattner Design Thinking Research program) and acting Assistant Professor at Stanford University.



Dr. Stephanie Balters is a Postdoctoral Researcher at the Center for Interdisciplinary Brain Sciences Research at the Stanford School of Medicine. Her research focus lies on robust fNIRS systems for interface-embedded neurofeedback and fNIRS hyperscanning applications for understanding human-to-human interactions. She received her PhD in Engineering Design from NTNU in Norway. Before joining the School of Medicine, Stephanie was a visiting researcher at the Center for Design Research and at the Computer Science Department at Stanford University. She is an active member of the NATO Human Factors Specialist Team for Unmanned Aerial Systems.



Dr. Joseph Geeseman is currently an active duty Aerospace Experimental Psychologist in the United States Navy with a rank of Lieutenant Commander. His current duty station is Naval Air Station Patuxent River, MD, where he is the Assistant Director for Science and Technology for the Human Factors Engineering Department. Dr. Geeseman received his PhD in Brain and Cognitive Sciences from Southern Illinois University – Carbondale in 2012. His current research interests include human performance in extreme environments, human systems integration, and human perception with a focus on multi-modal interaction.