STREL - Naturalistic Dataset and Methods for Studying Mental Stress and Relaxation Patterns in Critical Leading Roles

Corinna Rott, Fettah Kiran, Mien Segers, Piet Van den Bossche, and Ioannis Pavlidis, *Fellow, IEEE*

Abstract—We investigate mental stress and relaxation patterns in professionals occupying leadership roles in emergency care and special police units. A key finding is that on days that involve critical missions, these individuals experience negative mental stress (i.e., distress) that escalates as the day unfolds. In contrast, on non-leadership workdays, mental distress remains relatively stable, while on non-workdays, participants exhibit positive mental stress (i.e., eustress) that subsides over time, facilitating relaxation. These findings stem from a four-day naturalistic study of $n = 24$ professionals, during which we collected physiological, mobility, and psychometric data. Using participant debriefings and sensor-based validation triggers (e.g., GPS, cadence), we labeled activities in 5-minute intervals and focused our analysis on sedentary periods. We defined stress during these periods as a normalized heart rate that exceeds two standard deviations above a personalized baseline, thus isolating mental stress uncontaminated by physical exertion. A logistic regression model based on this stress labeling method yielded results largely consistent with those obtained from Kubios' SNS Index, reinforcing its validity. In cases of disagreement, our method aligned better with participant reports and established literature, highlighting advantages in interpretability and specificity. Overall, our work makes three contributions: (a) to affective science, by quantifying the mentally stressful nature of leadership in high-stakes environments; (b) to affective computing, by proposing a wearable compatible method for estimating mental stress during sedentary activity in the wild; (c) to data science, by introducing a well-annotated, multimodal dataset suitable for machine learning benchmarking in stress detection.

Index Terms—Stress, Relaxation, Emergency Care, Special Police Forces, Leadership, Naturalistic Study, Naturalistic Dataset, Affective Computing, Heart Rate, Heart Rate Variability, SNS Index.

1 Introduction

Work in professional roles, compared to non-professional roles, offers greater job satisfaction and security, but is also associated with elevated stress levels [1]. The professional workforce is diverse, comprising subgroups with varying stress profiles. At the extreme end, emergency care and police personnel routinely operate in high-intensity contexts, exacerbating stress and potentially impacting mental health [2].

This occupational stress has attracted substantial research, including surveys and naturalistic monitoring. Xu et al. surveyed physicians and nurses in an emergency unit and identified workload as the primary stressor; maintaining normalcy was reported as a predominant coping mechanism [3]. Over time, stress studies have expanded beyond surveys to include physiological monitoring. Early studies collected physiological signals in laboratory settings—e.g., mental arithmetic tasks—to enable stress labeling for machine learning (ML) models [4]. While these studies benefited from known ground truth, they lacked ecological validity. Nonetheless, they catalyzed growth in non-obtrusive stress sensing [5].

Recently, naturalistic studies have gained traction. These studies mirror real-life conditions, continuously collecting physiological and mobility data through wearable sensors, along with periodic psychometric questionnaires [6]. Without hard ground truth, stress labeling is inferred from physiological features or subject feedback. Ahmadi et al. monitored $n = 23$ hospital nurses for 12 hours using wrist-worn E4 sensors (Empatica, Italy) to collect heart rate (HR), electrodermal activity (EDA), and skin temperature (ST). They used the Baevskii Stress Index - based on heart rate variability (HRV) - to label stress and reported positive correlations with HR and ST [7].

In a similar effort, Hosseini et al. recorded HR, ST, and EDA for $n = 15$ nurses over one week during the COVID-19 pandemic, totaling 1,250 hours of data [8]. A Random Forest model, trained on AffectiveROAD [9], was used to detect stress episodes, which were subsequently validated by participant recall. Key stressors included COVID patient care and difficult interactions with family members. The study's longitudinal scope and stressor identification were notable contributions. Additional naturalistic datasets, such as TILES [10], gathered physiological and behavioral data from medical residents and relied on daily surveys for stress labeling.

This rise in naturalistic datasets offers opportunities for cross-study insights—but also introduces challenges. Inconsistent stress labeling approaches—ranging from surveys to

- *Corinna Rott and Piet Van den Bossche are with the School of Business and Economics, Maastricht, 6211 LK, Netherlands and the Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium. Mien Segers is with the School of Business and Economics, Maastricht, 6211 LK, Netherlands.*
- *Fettah Kiran and Ioannis Pavlidis are with the Affective and Data Computing Laboratory, University of Houston, Houston, Texas, 77204 E-mail: ipavlidis@uh.edu*

Manuscript received March 1, 2025. Paper accepted on September 25, 2025.

physiological metrics—complicate generalization. Cahoon and Garcia attempted to develop a domain-independent ML model and found that inconsistencies across datasets limited performance [11].

Outside healthcare, police officers also face intense stress. Queirós et al. surveyed $n = 2{,}057$ Portuguese police and found high rates of operational stress (85%), burnout (11%), and distress (28%) [12]. Garbarino et al. linked occupational stress in special police units to mental health concerns and showed that Five-Factor personality traits (notably neuroticism and agreeableness) modulate stress vulnerability [13], [14].

Of particular interest is the stress experienced by leaders in these critical professions. Leadership is inherently stressful across domains [15], and in crisis contexts, the burden is amplified [16]. Leadership stress can impair decision-making and performance during critical events. Understanding its dynamics is essential to inform training and mitigation strategies. Accordingly, our study investigates stress and relaxation patterns in emergency care and special police leaders across critical, routine, and leisure contexts.

To refine stress estimation, we restrict analysis to sedentary periods, minimizing confounding from physical exertion. This isolates cardiovascular changes attributable to mental (i.e., cognitive/emotional) stress. We further distinguish between distress (negative mental stress, e.g., overload [17]) and eustress (positive mental stress, e.g., play or parenting [18]). This aligns with stress appraisal theory, which emphasizes that individual interpretations shape physiological and psychological outcomes [19], [20].

Detecting and differentiating these stress forms requires both arousal and valence. Arousal is derived from physiological data; valence from psychometric and observational sources [6]. In our study, valence determination relied on structured daily debriefings. Together, these modalities enabled a nuanced, moment-by-moment picture of mental stress dynamics in high-stakes leadership roles.

1.1 Contributions and Research Question

Naturalistic studies of stress, such as those discussed in the Introduction [7], [8], [10], are growing in prominence and offer deeper insight into real-world stress responses. However, as a relatively new research paradigm, they face several methodological limitations. We identify three recurring issues in the existing literature:

1. **Lack of activity-aware stress labeling:** Reported studies often assign stress labels to physiological data without considering participants' concurrent activities. For instance, a participant walking from the office to a coffee shop may exhibit elevated heart rate, which is then misinterpreted as hyperarousal due to mental stress. In reality, this is physical arousal unrelated to cognitive or emotional load—the conditions most relevant to stress research. This blurring of definitions contributes to what Sander recently described as the arousal concept becoming "worse than useless" [21].
2. **Limited contextual coverage:** Prior studies typically focus on work periods alone, omitting non-work and sleep intervals. This results in the loss of key predictors and removes intra-subject baselines essential for anchoring stress assessments.
3. **Narrow professional scope:** Most studies target a single profession (e.g., hospital staff), missing potential generalizations or shared stress dynamics across high-stakes roles. The resulting methodological fragmentation makes cross-study conclusions difficult to draw [11].

Our study directly addresses these limitations:

1. We classify participant activities across the full observation window and compute stress exclusively during sedentary intervals. This reduces confounding by physical exertion and enhances confidence that the observed stress is of mental (cognitive/emotional) origin.
2. We include full-day coverage encompassing work, non-work, and sleep periods. This provides a holistic, 360° view of participant life rhythms, enabling detection of both distress (e.g., during critical work) and eustress (e.g., during leisure).
3. We examine crisis leaders from both emergency care and special police units—two distinct but comparably intense professional domains. This dual-focus design increases the ecological breadth of our findings and highlights cross-professional stress patterns.

These methodological enhancements allow us to pose the following research question:

> **RQ:** *What are the mental stress signatures of crisis leaders on critical workdays, non-critical workdays, and non-workdays?*

This question has both scientific and practical relevance. While much has been written about the potentially harmful effects of occupational stress on frontline responders [22], and especially their leaders [23], [24], the fact remains that many such professionals sustain long and successful careers. What coping mechanisms make this resilience possible?

We hypothesize that effective leaders may engage in self-regulatory strategies—such as embedding periods of low arousal (relaxation) into their routines—that buffer the impact of distress. Detecting these patterns is central to our analysis.

The rest of the paper is organized as follows: Section 2 presents the study design. Section 3 describes our methods for activity and stress labeling. Section 4 reports the analytical findings. Section 5 interprets the results and discusses implications.

2 Study Design

Based on an institutionally approved study protocol by the Ethics Review Committee of the University of Maastricht (No. ERCIC_535_24_02_2024), we recruited 25 volunteers to participate in a stress and relaxation (STREL) observational study; from these recruits, $n = 24$ completed the study. In this cohort, $n = 14$ were the leaders of emergency care units in the Netherlands and $n = 10$ were the leaders of special police forces in Switzerland. All participants signed an informed consent. Although the study participants had a leading role in their organization, the full weight of this role was manifest only on certain days when the said participants were handling critical events. Consequently, the observation period encompassed the following four days: one critical workday (CD), two standard workdays (WD1 and WD2), and one non-workday (ND). For emergency care leaders, the critical workday was a day they were managing critical incidents in the emergency department, while standard workdays were days they were doing regular clinical rounds. For special police leaders, the critical day was the day they were in charge of the security of a very important person (VIP) or a high-value detainee (HVD), while standard workdays were days they were planning future security operations or engaged in debriefing or led training classes. For all participants, the non-workday was typically a weekend day they spent at home or elsewhere, freely participating in recreational activities, family time, and house chores. At the end of each day, the experimenter debriefed the participants, extracting an account of what transpired during the day, noting stressful events, happy moments, and relaxation periods. The transcripts of these debriefings significantly helped in activity labeling and interpretation of analytical results. Participants also had to complete trait and state questionnaires that were implemented in Qualtrics$^{\text{XM}}$ (Qualtrics, Seattle, WA). The trait questionnaires were sent to their smartphones before the beginning of the study and the state questionnaires towards the end of each day. Participants also received a Polar Vantage V2 smartwatch + H10 chest strap combo (Polar, Kempele, Finland) and were instructed on its use. They had to wear these devices for the four days of observation, including in their sleep. The devices were recording geographic, motion, physiological, and sleep information at high resolution (1000 Hz).

2.1 Study Data

Daily Debriefs: The experimenter debriefed participants at the end of each day. Brief summaries of these debriefs are available in the supplementary materials. Here we provide as examples excerpts from the debriefs of participant T025 for the CD and ND days. These samples give activity information and vivid impressions of the participant's distress in CD and eustress in ND, revealing their usefulness in activity labeling and the identification of arousal polarity.

> *On day CD, T025 drove to work and started her shift around 7:30/08:00. Her shift should have ended around 17:30; however, she was called to deal with an acute emergency at the last minute. The call came around 16:30. A patient was brought in with heavy bleeding and polytrauma. The bleeding was internal and he was coughing blood. Hence, they had to react quickly. In the event of such heavy bleeding, the airways can be blocked by blood, which is a life-threatening condition. Under her leadership, the team was able to bring the patient out of this critical condition at 18:30. After a long day, she drove home and went for a short walk to unwind. Afterward, she might have fallen asleep on the couch before going to bed.*
>
> *On Thursday the 11th, she enjoyed her ND day by first going for a run and then biking to grab a coffee with a friend. Although she was in a hurry to get there, she enjoyed riding the bike*

Biographic Questionnaire: Participants had to complete the biographic questionnaire once. The questionnaire elicited information about the participant's occupation, age, sex, family status, and body mass index (BMI). The family status of the participants, especially whether they had preteen children or not, was important to gain insight into their activities during non-work periods. The BMI of the participants was a covariate of interest because a high BMI can play a confounding role in cardiac recordings [25], which we use for stress labeling. The research team also had information about the location of the home and workplace of the participants, which helped validate the activities of the participants by cross-referencing the reported locations with the logged coordinates.

Big-Five Inventory: Participants had to complete the Big Five Inventory once. Previous research reported that some personality traits, such as nervousness, amplify stress responses in critical professions [14]. Consequently, personality traits are useful predictors in a naturalistic stress study like ours. The Big Five Inventory measures personality along five dimensions [26]:

- Agreeableness ($B5_A$): The level of friendliness of the participant with the score range $[2, 10]$.
- Conscientiousness ($B5_C$): The level of organized nature of the participant with the score range $[2, 10]$.
- Extraversion ($B5_E$): The level of the outgoing nature of the participant with the score range $[2, 10]$.
- Neuroticism ($B5_N$): The level of nervousness of the participant with the score range $[2, 10]$.
- Openness ($B5_O$): The level of curiosity of the participant with the score range $[2, 10]$.

NASA Task Load Index (NASA-TLX): NASA-TLX has been increasingly used as a subjective predictor in naturalistic stress studies [27] because it can provide clues about the source of observed stress (e.g. time pressure). Conventionally, NASA-TLX measures multidimensional loading with respect to a specific task. For this reason, in experimental studies, it was administered to participants after the completion of each task. This is impossible in naturalistic studies, where participants perform many tasks as part of their daily lives, often simultaneously. Recent research showed that NASA-TLX can be applied not only to single tasks but also to a sequence of tasks without loss of validity [28]. This supported our use of NASA-TLX at the end of each day to provide a measure of the participants' perceived daily loading. Consequently, from each participant, we received daily scores for the NASA-TLX six subscales [29]:

- Mental Demand ($NASA_{MD}$): Perception of daily mental load with the score range $[1-20]$.
- Physical Demand ($NASA_{PD}$): Perception of daily physical exertion with the score range $[1-20]$.
- Temporal Demand ($NASA_{TD}$): Perception of daily time pressure with the score range $[1-20]$.
- Performance ($NASA_P$): Perception of success in executing daily tasks with the score range $[1-20]$.
- Effort ($NASA_E$): Perceived amount of work expended to achieve said daily performance with the score range $[1-20]$.
- Frustration ($NASA_F$): Perceived level of irritation in performing daily tasks with the score range $[1-20]$.

Geographic, Motion, Physiological, and Sleep Data: In terms of geographic information, the Vantage V2 watch recorded the GPS coordinates. Hence, the GPS channel provided the location of the participant over time, allowing association with certain types of activities (e.g., at the workplace during the period $t_2 - t_1$). In terms of motion information, the device recorded the cadence and speed. The combination of geographic and motion information was used to validate the differentiation of physical from non-physical activities on the move (e.g., biking from driving).

In terms of physiological information, the Polar H10 chest strap recorded HRV, HR, and daily HR baselines. We used HR signals and HR baselines to compute stress labels during sedentary activities. In parallel, we fed HRV signals to the Kubios software (Kubios, Kuopio, Finland) to calculate the Sympathetic Nervous System (SNS) Index [30] and the Stress Index (SI) [31] for developing an alternative set of stress labels. We used this alternative set of stress labels as a benchmark against our HR-baseline-based stress labeling method. Sleep times were determined using physiological analytics provided by the Kubios software and participant-reported sleep times.

3 METHODS

3.1 Data Aggregation and Summarization

We bring the quantitative, rank, and categorical data into a single data frame to attain time co-registration. Quantitative data include physiological and motion signals recorded by the Polar Vantage V2 + H10 combo. These are heart rate (HR) in beats per minute or BPM, heart rate variability (RR) in ms, cadence (CDN) in cycles per minute or CPM, and speed (SPD) in km/h. The duration of sleep ($SLEEP$) in hours is a reconciliation of the sleep times reported by participants and Kubios analytics. Quantitative data also include the derived variables Stress Index (SI) and SNS Index (SNS), which are calculated by the Kubios software using HRV data from the H10 chest strap. Both of these indices quantify physiological arousal and, because they are widely used, can serve as a benchmark for any custom stress labeling operation.

The rank data include the scores of the Big Five and NASA-TLX questionnaire subscales. The categorical data include participant ID with levels {T001, ..., T012, T014, ..., T025}; occupation (OCC) with levels {EC, SP}, where EC and SP stand for Emergency Care and Special Police, respectively; day (DAY) with levels {WD1, CD, WD2, OD}; and period of day (PRD) with levels {MRN, AFN}, where MRN and AFN stand for Morning and Afternoon, respectively. The morning and afternoon levels correspond to the daily periods [12 a.m. to 12 p.m.) and [12 a.m. to 12 a.m.], respectively, reflecting the circadian cycle, which is known to modulate human physiology and behavior [32].

The high-frequency signal data, with a sampling rate of up to 1000 Hz, are the ones that determine the temporal resolution of the analysis. HR and HRV—key determinants of stress—are our primary physiological considerations. We summarize the HR and HRV data into 5-min intervals by computing the corresponding means. This is in accordance with the guidelines of the cardiological societies for the study of longitudinal cardiovascular data [33]. To achieve temporal alignment across data streams, we apply the same five-minute summarization to all other signals, including the Kubios Stress Index and the SNS Index, as well as cadence and speed. The last two variables indicate mobility behaviors. Thankfully, five-minute intervals are also meaningful units for behavioral analysis because they can capture even micro-activities, such as bathroom visits, which have a similar timescale [34].

3.2 Activity Labeling

In naturalistic stress studies, activity labeling is often given a short shrift but is a prerequisite for meaningful stress labeling. It is paramount to differentiate physical from non-physical activities. Only during non-physical activities can observed physiological arousal be confidently attributed to mental stress. Activity labeling was performed in three steps:

1. A data scientist from the team applied a deterministic protocol that mapped time-stamped debriefing notes onto sensor verified criteria.
2. The study psychologist independently reviewed all labels. Concordance between the psychologist and the data scientist was complete for the driving and outdoor physical activity segments. Questions were raised only in some indoor walking segments for which there were no debrief notes and no GPS signal (e.g., walk to cafeteria to grab lunch), leaving only the cadence and speed signals as sensor triggers.
3. These few ambiguous epochs were resolved through follow-up questions to the participant, resulting in a set of consensus labels. Because the procedure is rule-driven and disagreements were fully adjudicated, conventional inter-rater reliability metrics were not applicable.

This process falls under activity recognition pipelines, where ground truth labels are obtained from timestamped logs and verifiable sensor data (e.g., GPS), typically validated by a single rater or the subject, rather than through full inter-rater protocols [35].

The data scientist performed the labeling in the following order: ‘Sleeping’, ‘Driving’, ‘Biking’, ‘Running’, ‘Work’, ‘Non-Work’, and ‘Walking’. As an example, we will describe in detail how activity labeling was performed for participant T025, the result of which is depicted in Fig. 1a.

Sleeping: There is information in the debrief notes about when the participants went to sleep and when they got up each day. The data scientist validates this information with the sleep signals captured from the participants’ smartwatches and assigns ‘Sleeping’ labels accordingly.

Driving: To assign ‘Driving’ labels, the data scientist uses the GPS signal in combination with the speed and cadence signals to validate the information from the debriefing. For example, the debrief states that participant T025 started driving toward work around 7 a.m. on day CD. Hence, the data scientist checks the vicinity of 7 a.m. in the timeline to see when the GPS signal indicates moving away from her house location and along a highway on the map. Once he

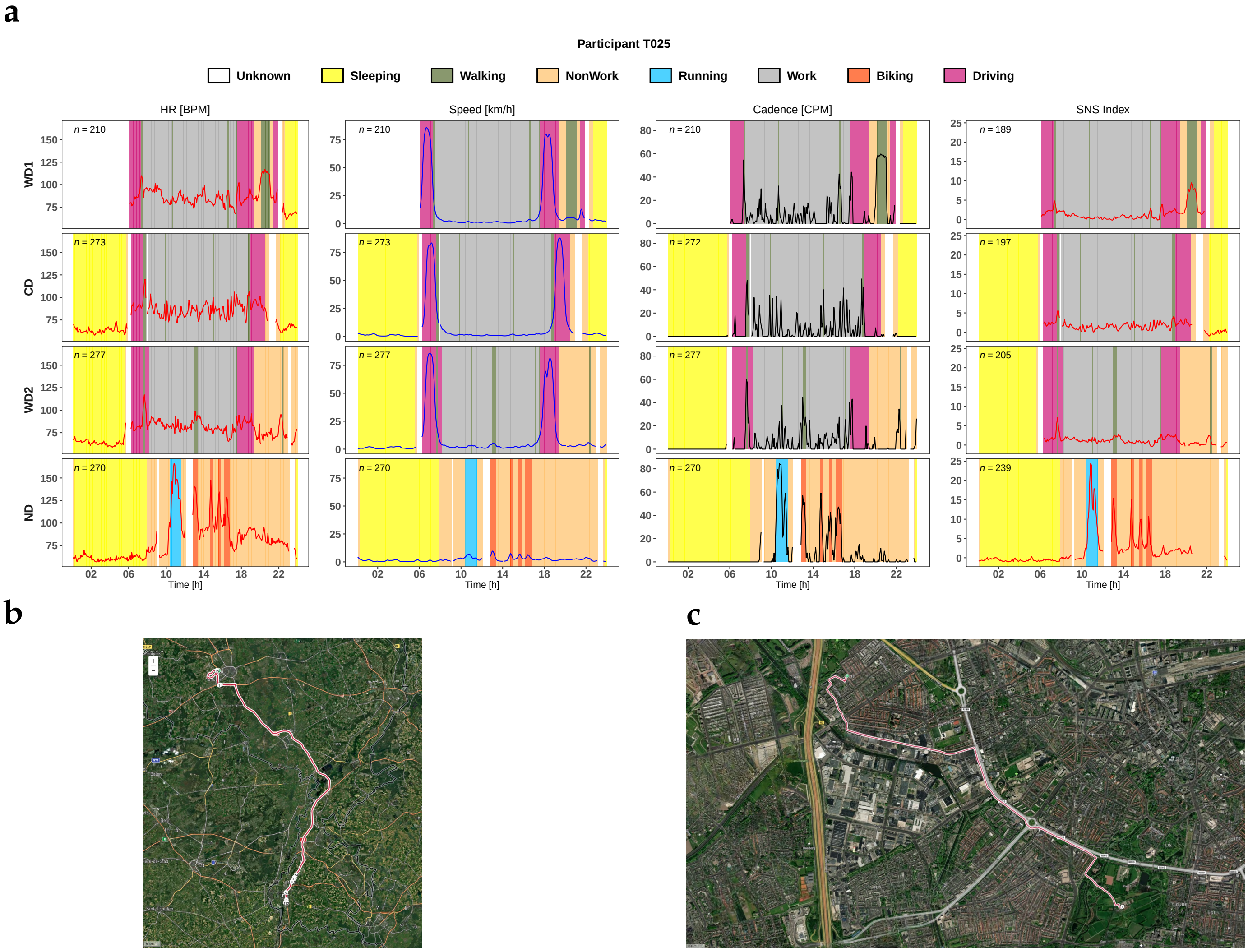


Fig. 1: Activity labeling of participant T025. a. Visualization of activity labels on days WD1, CD, WD2, and ND. The labels are conveyed as background colors in four columns of signal panels. The speed and cadence signals in the middle columns are instrumental in activity validation. The HR and SNS Index signals in the outer columns track physiological arousal and can be used as extra validity checks. For instance, when cadence is high, HR and SNS signals must also be high, conveying sympathetic activation due to physical activity. b. Morning commute of participant T025 on workday WD1. The itinerary is outlined in pink and captured through the GPS channel. c. Recreational biking itinerary of participant T025 on non-workday ND. The itinerary is outlined in pink and captured through the GPS channel.

spots the beginning of such a move, he follows its evolution on the map (Fig. 1b) while simultaneously checking the speed signal versus the cadence signal for extra validation. Since this is a driving activity, he expects to see high speed and nearly zero cadence because drivers are sedentary (magenta strips in Fig. 1a). As long as all three signals (GPS, speed, cadence) show evidence of vehicular movement, he keeps labeling intervals as 'Driving'. When the GPS signal stops evolving and shows that the participant reached her workplace (which is known to the research team), then this is a sign that the driving activity has ended.

Biking: All biking episodes in our dataset occurred outdoors. To assign 'Biking' labels, the data scientist uses the GPS signal in combination with the speed and cadence signals to validate the information from the debriefing. For example, the debrief states that participant T025 used her bike in the afternoon of day ND to go and meet a friend at a coffee shop and then run some errands. Hence, when the data scientist observes movement of the GPS signal during that period (Fig. 1c), he understands that it is a bike activity and labels it as such, something he also validates from the cadence and speed values, by observing cadence values in the range [50, 70] CPM and speed values in the range [12, 22] km/h, which are typical of city biking; see dark orange strips in Fig. 1a. Furthermore, because this is a physical activity, it elevates HR, which offers an additional validation clue - see in Fig. 1a the spikes in the HR signal in the afternoon of ND.

Running: All running in our dataset, except for one exception we know of, takes place outdoors. To assign 'Running' labels, the data scientist uses the GPS signal in combination with the speed and cadence signals to validate the information from the debriefing. For example, the debrief states that participant T025 ran outdoors on the morning of ND. Thus, when the data scientist observes a movement of the GPS signal away from T025's house and back that morning, he understands that it is likely a running activity. He labels it as such, after he also validates it from the cadence and speed values, by observing cadence values in the range [75, 85] CPM and speed values in the range [7, 12] km/h, which are indicative of recreational running [36]; see the cyan strip in Fig. 1a. Furthermore, because this is an intense physical activity, it greatly elevates HR, which offers an additional validation clue - see the major HR spike in the morning of ND in Fig. 1a.

Non-Work and Work: After labeling the driving, biking, and running activities that can be easily validated from their evolving GPS signals, the data scientist performs an initial labeling of sedentary 'Work' and 'Non-Work' activities. The biographic notes for participants have the GPS locations of their workplaces and homes, which enables the data scientist to assign the labels 'Work' and 'Non-Work' to intervals where the GPS signal appears to be anchored in the corresponding locations. Some of these labels will change as in a second pass the data scientist identifies short periods of walking.

Walking: Lastly, the data scientist labels 'Walking' activity, which is the only somewhat challenging activity to validate. The validation difficulty with walking is that it does not always happen outside, but sometimes inside buildings. For sustained walks that happen outside, there is information in the debriefings, and it is easy to validate them using the GPS and other signals, as was done for running. For example, the debrief states that participant T025 on day WD1, walked to the house of her parents-in-law to have dinner after work. Hence, following her moving GPS signal during that time, the data scientist labels the corresponding interval as walking (green strip on WD1 panels of Fig. 1a). However, many times sustained walking takes place inside or in the immediate vicinity of buildings, where GPS resolution is limited or non-informative. Furthermore, some of these micro-activities were not captured in the debriefings. For example, from the cadence signal it is understood that participant T025 likely had a walking activity around noon on day WD2. After discussion with the psychologist, the team had to go back to the participant and inquire about this, who provided additional information that she walked to grab her lunch. This was the validation that was needed to reclassify this initially labeled sedentary ('gray') interval as a walking ('green') interval. In Fig. 1a, see the green strip on day WD2, which is inside the workplace environment (gray zone). The green strip is characterized by elevated cadence and HR relative to other intervals within the gray zone that have much lower cadence, indicative of sedentariness.

Other cases during work, where the intervals changed from initially gray (sedentary) to green (walking), are after arrival by car in the morning and before departure by car in the evening. See, for example, in Fig. 1a the thin green strips between the magenta and gray regions on day CD. During these periods, the cadence is elevated, as is the HR, indicating physical activity. This is because participant T025 had to walk for several minutes from the parking ramp to her office and vice versa, something the team verified with her.

3.3 Stress Labeling

In stress studies, it is constitutional to have information on the evolving stress levels of the participants. This is achieved through stress labeling. The conventional stress labeling method is based on self-reporting. Self-reporting fits well with laboratory experiments, where after the execution of each preplanned task, participants receive a questionnaire to elicit their perceived stress [37]. Self-reporting becomes unwieldy in naturalistic studies, where monitoring of short preplanned tasks is replaced by continuous monitoring of real life for days or weeks. In these cases, stress labeling is increasingly derived from physiological measurements alone. Several physiologically driven stress labeling methods have been reported in the literature, and new ones are appearing all the time. The Baevskii Stress Index (SI) [31], which is computed from HRV parameters, is one such method. Kubios' SNS Index, whose calculation is based on HRV-derived features, including the Stress Index (SI) and the normalized Poincaré plot index SD2 [30], is another measure used for stress labeling.

None of the existing stress labeling methods is dominant. For this reason, it is suggested to use more than one labeling method to ensure robustness and capture different aspects of stress responses [38]. In this spirit, we use both a new and an established stress labeling method. The new stress labeling method is based on within-participant daily statistics of normalized HR (NHR) - a signal that is more widely available in smartwatches than HRV.

3.3.1 NHR Stress Labeling

Here we provide a dissection and justification of our NHR stress labeling method:

Defining the NULL hypothesis (No Stress): For each participant p_i and day d_j, we calculate the signal NHR by subtracting the individual baseline $HR_{BL}(p_i, d_j)$ of that day from the continuous signal $HR(p_i, d_j, t_k)$, where every moment t_k corresponds to a 5-minute interval. Then we focus on the sedentary periods, forming the sub-signal $NHR_{\mathcal{D}}(p_i, d_j, t_l) \subset NHR(p_i, d_j, t_k)$, where in moments t_l the participant performs office work (Work), stays home / non-work establishment (Non-Work), or drives (Driving). Equation (1) expresses the formulation of the $NHR_{\mathcal{D}}$ signal.

$$NHR_{\mathcal{D}}(p_i, d_j, t_l) = \overline{HR}_{\mathcal{D}}(p_i, d_j, t_l) - HR_{BL}(p_i, d_j) \\ \Rightarrow \quad NS \equiv NHR_{\mathcal{D}} = 0 \qquad (1)$$

The value $NHR_{\mathcal{D}} = 0$ indicates that the participant is at his/her daily baseline ('No Stress') and we treat it as the fixed NULL hypothesis, that is, the physiological level of no stress (NS). Note that the NS point is usually different from the mean of the sedentary $NHR_{\mathcal{D}}$ distribution, $\overline{NHR}_{\mathcal{D}}(p_i, d_j) \neq 0$ due to asymmetry from stress episodes. Also note that by stress, we refer to mental stress, that is, a combination of cognitive and emotional stress.

Reduction relative to the No Stress reference: Rather than centering the data and applying conventional z scores, we

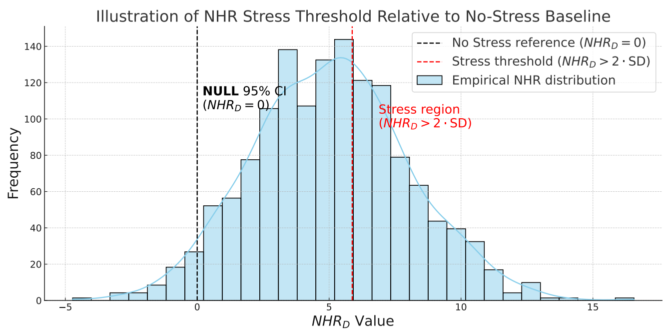


Fig. 2: Illustration of NHR_D stress threshold relative to No Stress reference.

retain the original empirical distribution and assess the deviation from $NHR_{\mathcal{D}} = 0$, using the observed standard deviation $\mathrm{SD}_{NHR_{\mathcal{D}}}(p_i, d_j)$ of the participant p_i, during sedentary periods of the day d_j. We define the stress threshold as follows:

$$\mathrm{Threshold}_{\mathrm{Stress}}(p_i, d_j) = 0 + 2 \cdot \mathrm{SD}_{NHR_{\mathcal{D}}}(p_i, d_j) \tag{2}$$

Under the assumption of approximate normality (supported empirically, see QQ plots in Figs. 3a2 and 4a2), this threshold captures the upper 2.5% of expected variation under NULL. Any value of $NHR_{\mathcal{D}}$ that exceeds this threshold is considered to reflect significant autonomic activation beyond routine physiological fluctuations, that is, a stress response (S). Since there is no physical activity involved, this stress response is likely of cognitive and emotional origin.

Empirical and visual support: Figure 2 illustrates how the $+2 \cdot \mathrm{SD}$ threshold is applied relative to the fixed reference point of $NHR_{\mathcal{D}} = 0$, in the general case when the empirical distribution is not centered at zero.

Theoretical and applied justification: This approach is consistent with classical statistical inference, where deviation from a fixed NULL (rather than the empirical mean) is tested using the observed variance [39], [40]. It also aligns with recent wearable computing literature that advocates personalized context-aware stress thresholds [41], [42].

3.3.2 SNS Stress Labeling

To benchmark our new stress labeling method, we implemented a comparative labeling approach using an established physiological stress metric. To this end, we focus on the SI and SNS Index provided by the Kubios software and are based on computations that involve HRV-derived measures [30], [31]. Because we found that the SI and SNS Index are highly correlated in the STREL dataset (Pearson correlation, $r = 0.91, p < 0.001$), we chose only the SNS Index (SNS) to serve as a benchmark. We adopted the literature-recommended threshold of $SNS = 1$ to distinguish between Stress and No Stress states [30]. Consequently, we apply this threshold across the $SNS(p_i, d_j)$ distributions of participant p_i on day d_j to label stressful (red sub-histogram in Fig. 3b1) vs. non-stressful moments (green sub-histogram in Fig. 3b1). If the stress labels computed with our method are on par with the stress labels computed with the SNS method, then logistic regression models with corresponding response variables and the same predictors should yield similar results. In this case, the benefit of our method is that it is simpler and more broadly applicable because it uses only HR, which is available on all smartwatches.

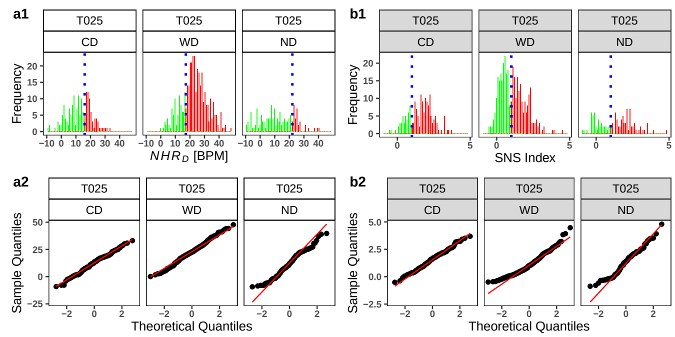


Fig. 3: Physiologically-driven stress labeling for participant T025. a1/b1. Distribution of sedentary NHR / SNS on days CD, WD = WD1 ∪ WD2, and ND. The blue dotted lines mark $2 \cdot$ SD from $NHR = 0$ (No Stress) and $SNS = 1$ for the NHR and SNS methods, respectively. a2/b2. QQ plots of sedentary NHR / SNS on days CD, WD, and ND. The plots suggest no severe deviation from normality, and thus 95% CI = $2 \cdot$ SD can be used as reliable statistics for defining significant distances from the No Stress hypothesis in NHR.

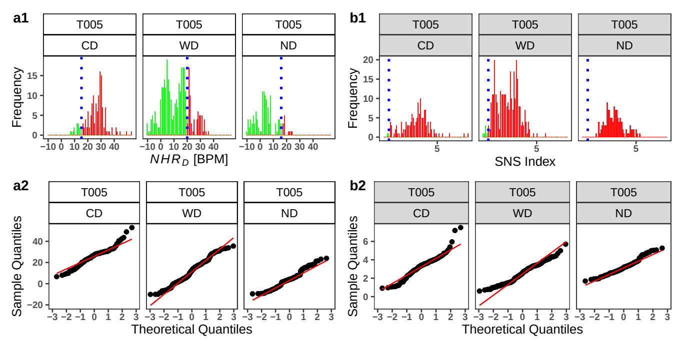


Fig. 4: Physiologically-driven stress labeling for participant T005. a1/b1. Distribution of sedentary NHR / SNS on days CD, WD = WD1 ∪ WD2, and ND. The blue dotted lines mark $2 \cdot$ SD from $NHR = 0$ (No Stress) and $SNS = 1$ for the NHR and SNS methods, respectively. The SNS labeling method shows the participant to be stressed all the time, including throughout ND, which we know from the debriefings not to be true. By contrast, the NHR labeling method shows results that are in agreement with the debriefings, including a high-stress CD day vs. a low stress ND day. a2/b2. QQ plots of sedentary NHR / SNS on days CD, WD, and ND. The plots suggest no severe deviation from normality, and thus 95% CI = $2 \cdot$ SD can be used as reliable statistics for defining significant distances from the No Stress hypothesis in NHR.

3.3.3 Comparative Analysis of NHR vs. SNS Methods

Let us examine individual participants and days more closely, focusing on an example of strong agreement and one of strong disagreement between the NHR and SNS methods. In the histograms of Fig. 4a1 / b1, we observe that NHR and SNS agree that participant T005 was stressed most of CD, with about 90% of the labels marked S (red) by both methods. Note that the baseline HR of T005 in CD was 69 bpm while his mean HR was 94.9 bpm, resulting in a large $\overline{NHR} = 25.9$ bpm. In contrast, the two methods are largely in disagreement about the stress condition of participant T005 in ND. The NHR method indicates that the participant was mostly non-stressed (S labels only 10% of the time), while the SNS method indicates that the participant was always stressed that day (S labels 100% of the time). Note that the baseline HR of T005 in ND was 83 bpm, while his mean HR was 87.3 bpm, resulting in a small $\overline{NHR} = 4.3$ bpm.

According to the debriefings, participant T005 not only was assigned a difficult mission in CD (that is, monitoring a person of interest in a public event), but was also emotionally charged from an argument he had with his ex-partner before starting his shift. Hence, the NHR and SNS methods are correct in their stress estimations for that day. The situation is different for ND. According to the debriefings, participant T005 spent most of ND relaxing and watching television. The NHR labels align with the self-reported experience of the participant, while the SNS labels contradict this. This suggests that the SNS method is more prone to false positives in this context. Two key reasons explain this.

1. **Lack of baseline correction in the SNS.** The NHR method subtracts a daily participant-specific baseline, while the SNS method applies fixed thresholds without baseline normalization. On day ND, T005's baseline HR was 83 bpm and his mean HR was 87.9 bpm, producing a small $\overline{NHR} = 4.9$ bpm. The elevated HR baseline likely reflects behavioral factors (e.g., extended TV watching, snacking, caffeine intake), which increase HR without necessarily inducing psychological stress [43]. The small NHR values throughout the day fall well within the $2 \cdot$ SD No Stress range. Although the SNS Index does not directly use HR, elevated HR influences the RR interval statistics (e.g., SDNN, LF / HF), inflating the SNS Index. Since the SNS Index uses absolute cutoff points ($SNS > 1 \Rightarrow$ S label), it misclassifies this elevation as stress.
2. **Sensitivity to transient vagal shifts.** HR-only metrics like NHR are less reactive to short-term parasympathetic fluctuations. In contrast, SNS incorporates HRV features (e.g., LF/HF ratio) that are susceptible to vagal shifts triggered by posture changes or yawns — common during sedentary activities like TV watching [44]. These can spuriously elevate SNS values.

In summary, the informative disagreement between the two methods on day ND for T005 highlights the advantages of the personalized, HR-based and activity-aware design of NHR, particularly in low-arousal, free-living contexts.

Prolonged HRV suppression after physical activity. There is one more condition that gives an advantage to the NHR method in naturalistic studies over the SNS method. HR returns to baseline within minutes after physical exertion, whereas HRV can remain suppressed for much longer [45]. Consequently, the SNS, which is HRV-based, is more susceptible to contamination from preceding activity. This effect persists even within sedentary intervals, making the SNS less temporally precise than the NHR to detect recovery. This factor likely plays a role in many cases, leading the SNS to overestimate stress in sedentary intervals, when these intervals are neighboring physical activity intervals. An example of this is day ND for participant T025, which was a day that she engaged in running and biking several times during the day (Fig. 1a). As is evident in Fig. 3, the SNS stress distribution of T025 in the sedentary intervals of ND has a heavier right tail than that of NHR. Specifically, the NHR method assigns S labels 30% of the sedentary time in ND while the SNS method assigns twice as much (61%). We know from the debriefing that participant T025 had a relaxing day with periods of physical activity in the morning and afternoon and a restful evening, which is in line with the NHR estimate and in contradiction to the SNS estimate. The latter has probably been contaminated by dispersed physical activity during the day.

4 Results

4.1 Collinearity Check

Before modeling, we construct the cross-correlation table for all predictor variables to check for collinearities - Fig. 5. We find high correlations only among a handful of variables. Specifically, we find $NASA_{MD}$ to be highly correlated with $NASA_{TD}$ ($r = 0.64$); thus, we eliminate $NASA_{TD}$ from further consideration. We also find $NASA_{MD}$ to be highly correlated with $NASA_E$ ($r = 0.61$); therefore, we eliminate $NASA_E$ from further consideration. Finally, we find $OCC[SP]$ to be highly correlated with $NASA_F$ ($r = 0.57$); thus, we drop $NASA_F$ from further consideration. As we mentioned in a previous section, $OCC[SP]$ stands for categorical variable occupation, which has two levels: Special Police (SP) and Emergency Care (EC), with EC serving as a reference. This filtering of correlative predictors leaves only three of the original six NASA-TLX subscales to be included in our models. As is evident from the matrix in Fig. 5, all other predictor pairs have weak correlations ($|r| < 0.4$), and therefore no further predictor exclusion is necessary.

4.2 Descriptive Statistics

The sample included 24 adult participants (15 male/9 female; age = 40.96 ± 7.58). Of these 24 participants, $n = 14$ were leaders in emergency care units and $n = 10$ were leaders in special police forces. The participants had a healthy non-obese physique (BMI = 26.27 ± 4.7), and therefore there were no apparent confounders of their cardiovascular function.

Key predictor variables include cadence, sleep, and psychometrics. Cadence (CDN) serves as a differentiator between sedentary and physical activities. Figure 6a shows the CDN distributions in sedentary versus physical activities for the STREL dataset. The sedentary distribution is exponential in shape, tucked into the lower value range of the x-axis ($n = 12708$ samples, 4.15 ± 6.98 CPM), while the physical activity distribution is fat-tailed, covering the higher value range of the x-axis ($n = 1655$ samples, 36.78 ± 16.32 CPM). There is minimal overlap between the two distributions. In sedentary activities, motion patterns reflect occasional micro-movements (e.g., going from one room to

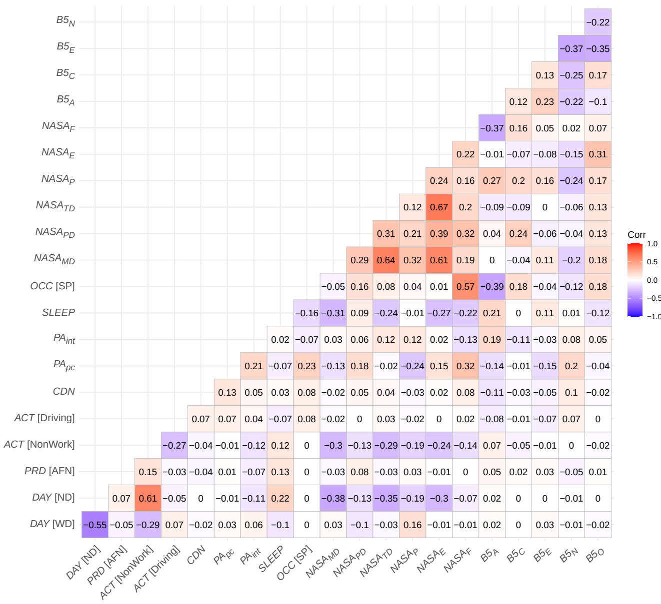


Fig. 5: Cross-correlation matrix for all model predictors. There are a few significant correlations between rank and categorical variables shown as intense red cells. The vast majority of the predictor pairs, however, are non-correlated.

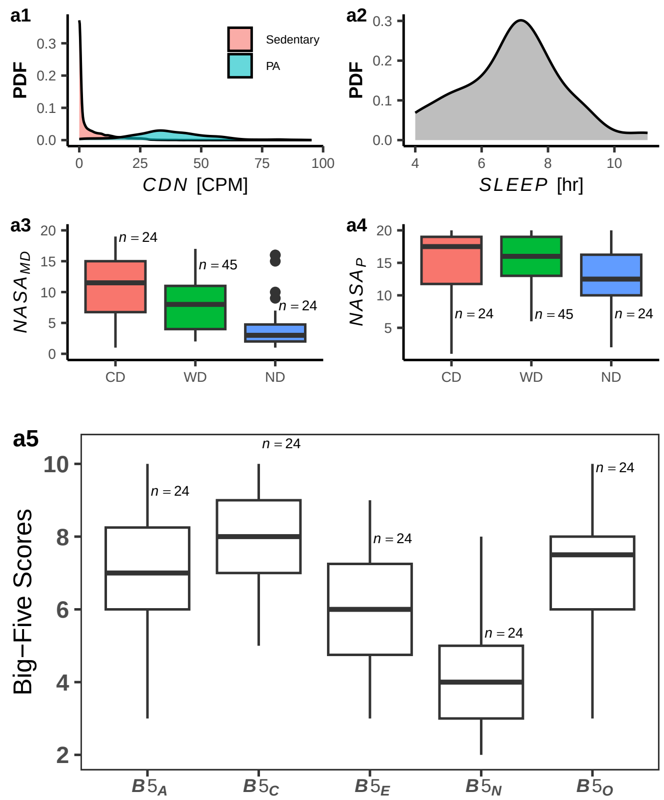


Fig. 6: Descriptive plots of modeling variables. a1. Probability density functions of Cadence (CDN) for sedentary and physical activities (PA). a2. Probability density function of sleep duration. a3. Box-plots of NASA TLX Mental Demand scores ($NASA_{MD}$), stratified per day. a4. Box-plots of NASA TLX Performance scores ($NASA_P$), stratified per day. a5. Box-plots of Big-Five subscales $B5_E$, $B5_A$, $B5_C$, $B5_N$, $B5_O$ for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness, respectively.

the next), which we describe by the term micro-cadence ($mCDN$). Physiological arousal during sedentary periods reflects mental (i.e., cognitive and emotional) stress. In physical activities, motion patterns reflect purposeful exercise (for example, running in the park). Physiological arousal during physically active periods reflects physical stress. Physical activities can improve cognitive and emotional stress, but such a reduction can be reliably measured only during sedentary activities. In our models, we take into account the sedentary and physical activity distributions of CDN by including as predictors micro-cadence ($mCDN$) and physical activity indicators, respectively. The physical activity indicators capture both the daily extent (expressed as a percentage PA_{pc}) and intensity (PA_{int}) of physical activities.

Sleep is constitutional to daily physiological functioning and, for this reason, it is also an important study variable ($SLEEP$). Figure 6a2 shows the $SLEEP$ distribution for the STREL data set ($n = 88$ samples, 6.92 ± 1.52 hr). The distribution is fat-tailed, with the left tail more pronounced than the right tail. This suggests that in most cases the participants enjoyed a seven-hour sleep, which is close to normal for adults. In some cases, the participants slept much less and in fewer cases much more than seven hours. The former pattern happened on workdays, while the latter pattern was on non-workdays, as can be seen in the detailed data (Supplementary Materials).

Daily psychometrics, in the form of NASA TLX scores, provide subjective support to physiological indicators of stress, and thus are worth including as predictors in the models. In fact, NASA Mental Demand ($NASA_{MD}$) and NASA Performance ($NASA_P$) emerged as significant model predictors (Section 4.3). Figures 6a3 and 6a4 show the boxplots of the daily $NASA_{MD}$ and $NASA_P$ scores, respectively. The box plots are stratified by the type of day, with CD denoting the critical day, WD = WD1 ∪ WD2 standard workday, and ND non-workday. Perceived mental load and performance are indicative of individual challenges and accomplishments. The boxplots in Fig. 6a3 suggest that mental load is most elevated on critical workdays ($n = 24$ samples, 10.67 ± 5.74) and least elevated on non-workdays ($n = 24$ samples, 4.88 ± 4.77); standard workdays fall in-between ($n = 45$ samples, 8.00 ± 4.40). These marginal statistics provide credence to the designation of critical days as such, while also demonstrating the decompressing nature of non-workdays.

In contrast to the clear marginal differences in $NASA_{MD}$, there are no clear marginal differences among critical workdays ($n = 24$ samples, 14.50 ± 6.44), standard workdays ($n = 24$ samples, 15.58 ± 3.68), and non-workdays ($n = 45$ samples, 12.83 ± 4.91) for $NASA_P$. In particular, the $NASA_P$ scores are consistently high, indicating that the participants were largely successful in what they were doing.

Trait psychometrics provide insights into the participants' characters, addressing questions about human sample homogeneity. Figure 6a5 shows the boxplots of the Big Five scores for the STREL dataset. It is clear that the participants are highly conscientious ($n = 24$ samples, 8.00 ± 1.62) and non-neurotic ($n = 24$ samples, 4.25 ± 1.59), manifesting a select group of individuals with similar attributes. These characteristics are fitting to the criticality of the participants' professions and their leadership roles within these professions. Agreeableness is also quite high ($n = 24$ samples, 7.13 ± 1.80), a characteristic that coincides with the team nature of the participants' professions. Extraversion and Openness cover a healthy range, with descriptive statistics ($n = 24$ samples, 6.04 ± 1.88) and ($n = 24$ samples, 7.08 ± 1.74), respectively.

4.3 Models

We report on the predictors of instantaneous stress of participants when they are at work, not at work, or driving. The unit of analysis is the mental stress state of the participant p_i (S vs. NS) in a 5-minute interval during the said activities. The response variable is the odds $P(S_v(p_i, d_j, t_l))$ that participant p_i is stressed in the sedentary 5-minute interval t_l on day d_j. Stress labels are extracted through channels NHR ($S_{v=NHR_D}$) or SNS ($S_{v=SNS}$). The relevant logistic regression model is shown in Eq. (3). The Stress intervals t_l at work / not at work / in the car on day d_j for the participant p_i fulfill one of the following conditions: $NHR_D > 2SD$ or $SNS > 1$. No Stress intervals correspond to the complement of that. Hyperarousal associated with said stressful states has emotional and cognitive origins. Physical stress is not the likely source of the observed hyperarousal because work, non-work, and driving activities are largely sedentary.

The type of day and the period of day are associated with the research question of this study, that is, whether critical workdays (i.e., CD) have significantly different mental stress signatures than typical workdays (i.e., WD) and non-workdays (i.e., ND). Accordingly, the first row of Eq. (3) lists the key predictors $DAY[\text{WD}](p_i)$, $DAY[\text{ND}](p_i)$, and $PRD[\text{AFN}](p_i, d_j)$. The dummy variables $DAY[\text{WD}](p_i)$ and $DAY[\text{ND}](p_i)$ represent the typical workday and the non-workday, respectively; the critical day CD serves as the reference level. The dummy variable $PRD[\text{AFN}](p_i, d_j)$ represents the afternoon period of a day; the morning period MRN serves as the reference level. The second row of Eq. (3) lists the interaction terms between the predictors DAY and PRD.

The remainder of Eq. (3) holds the control variables. Specifically, the third row of Eq. (3) lists the percentage of time that participant p_i devoted to physical activity $PA_{pc}(p_i, d_j)$ and the intensity of that activity $PA_{int}(p_i, d_j)$ on day d_j. The next three rows list all the interaction combinations among DAY, PA_{pc}, and PA_{int}. In the seventh row, we encounter the variable $ACT(p_i, d_j, t_l)$, which indicates the type of sedentary activity in the 5-min interval t_l on day d_j for the participant p_i; the levels are Non-Work and Driving, with the Work level serving as a reference. In the eighth row, we find the variables $mCDN$, $SLEEP$, and OCC. The variable $mCDN(p_i, d_j, t_k)$ quantifies the small bursts of cadence (aka micro-cadence) in the said sedentary periods; for instance, going from one room to another during office work. The variable $SLEEP(p_i, d_j, t_l)$ lists the number of hours the participant slept on that day. The variable $OCC(p_i)$ holds the participant's occupation; it has two levels: Emergency Care $\equiv$ EC (reference level) and Special Police $\equiv$ SP.

The last two rows of Eq. (3) feature the study's psychometric variables. One row holds the mental demand $NASA_{MD}(p_i, d_j)$, physical demand $NASA_{PD}(p_i, d_j)$, and performance $NASA_P(p_i, d_j)$ subscales of the NASA-TLX subjective workload assessment tool with respect to the participant p_i on day d_j. The other row contains the Big Five personality scores of agreeableness $B5_A$, conscientiousness $B5_C$, extraversion $B5_E$, neuroticism $B5_N$, and openness $B5_O$ for the participant p_i.

By performing a backward-elimination process based on the Akaike Information Criterion (AIC), we optimize the full models expressed in Eq. (3). For the model where the response variable is the odds of stress, with the stress labels extracted from NHR, only the conscientiousness is removed from the model, improving the AIC from 12575 to 12573. For the model where the response variable is the odds of stress, with the stress labels extracted from SNS, only the agreeableness is removed from the model, improving the AIC from 8721 to 8719. The parameter estimates for the two optimal models are listed in Table 1, while key results are visualized in Fig. 7. Several insightful conclusions arise out of these results, which are discussed next.

5 Discussion of Results

Naturalistic dataset: The naturalistic dataset we release with this paper, known as STREL, has three powerful characteristics:

1. It includes not just one, but two critical professions (Emergency Care and Special Police), facilitating generalizations in the critical profession space.
2. It encompasses workdays with and without a critical leadership component, as well as non-workdays. Such a comprehensive variety of daily patterns enables within-participant stress comparisons to address the study's research question.
3. It features multimodal data collection that includes 24/7 physiological, motion, and location signals along with psychometric scores and biographic notes. This cornucopia of data enables the labeling of activities and the estimation of valence and arousal of a cognitive or emotional nature. The combination of arousal and valence information affords differentiation between distress on workdays and eustress on non-workdays. With such nuanced activity and stress labels at our disposal, we were able to gain powerful insights into the lives of the participants through statistical models. With the said labels in place, the STREL dataset is also primed for training and testing machine learning (ML) algorithms for stress detection, something that would extend its usefulness well into the future.

Stress labeling of naturalistic affective computing data: Most naturalistic studies in affective computing are based on HR and HRV data, which are collected via wearable devices. In our approach, we performed activity

$$
\begin{aligned}
\text{logit}(P(S_v(p_i,d_j,t_l))) \sim & b_0 + b_1 DAY[\text{WD}](p_i) + b_2 DAY[\text{ND}](p_i) + b_3 PRD[\text{AFN}](p_i,d_j) + \\
& b_4 DAY[\text{WD}] \times PRD[\text{AFN}](p_i,d_j) + b_5 DAY[\text{ND}] \times PRD[\text{AFN}](p_i,d_j) + \\
& b_6 PA_{pc}(p_i,d_j) + b_7 PA_{int}(p_i,d_j) + b_8 DAY[\text{ND}] \times PA_{pc} \times PA_{int}(p_i,d_j) + \\
& b_9 DAY[\text{WD}] \times PA_{pc}(p_i,d_j) + b_{10} DAY[\text{ND}] \times PA_{pc}(p_i,d_j) + \\
& b_{11} DAY[\text{WD}] \times PA_{int}(p_i,d_j) + b_{12} DAY[\text{ND}] \times PA_{int}(p_i,d_j) + \\
& b_{13} DAY[\text{WD}] \times PA_{pc} \times PA_{int}(p_i,d_j) + b_{14} DAY[\text{ND}] \times PA_{pc} \times PA_{int}(p_i,d_j) + \\
& b_{15} PA_{pc} \times PA_{int}(p_i,d_j) + b_{16} ACT[\text{Non-Work}](p_i,d_j,t_k) + b_{17} ACT[\text{Driving}](p_i,d_j,t_k) + \\
& b_{18} mCDN(p_i,d_j,t_k) + b_{19} SLEEP(p_i,d_j) + b_{20} OCC[\text{SP}](p_i) + \\
& b_{21} NASA_{MD}(p_i,d_j) + b_{22} NASA_{PD}(p_i,d_j) + b_{23} NASA_{P}(p_i,d_j) + \\
& b_{24} B5_A(p_i) + b_{25} B5_C(p_i) + b_{26} B5_E(p_i) + b_{27} B5_N(p_i) + b_{28} B5_O(p_i)
\end{aligned}
\tag{3}
$$

TABLE 1: Results of the optimal logistic regression models emanating from the full model of Eq. (3). Columns with a white background correspond to the version of the model with stress labels from NHR_D. Columns with a gray background correspond to the version of the model with stress labels from SNS. Predictors highlighted in yellow indicate variables where the two model versions are in complete agreement. Significance levels have been set as follows: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

Predictor	Coefficient b.		Standard Error		95% CI		z value		Pr($>\|z\|$)	
Intercept	-0.776	0.400	0.317	0.483	$[-1.397, -0.156]$	$[-0.546, 1.346]$	-2.450	0.830	0.014^{*}	0.408
DAY[WD]	-0.217	0.648	0.099	0.147	$[-0.411, -0.023]$	$[0.361, 0.936]$	-2.190	4.418	0.028^{*}	$< 0.001^{***}$
DAY[ND] ✓	1.380	0.844	0.140	0.188	$[1.104, 1.655]$	$[0.476, 1.212]$	9.820	4.491	$< 0.001^{***}$	$< 0.001^{***}$
PRD[AFN] ✓	0.548	0.600	0.095	0.131	$[0.363, 0.733]$	$[0.343, 0.857]$	5.800	4.573	$< 0.001^{***}$	$< 0.001^{***}$
DAY[WD] $\times$ PRD[AFN] ✓	-0.369	-0.607	0.116	0.153	$[-0.596, -0.141]$	$[-0.907, -0.307]$	-3.180	-3.963	0.002^{**}	$< 0.001^{***}$
DAY[ND] $\times$ PRD[AFN] ✓	-0.670	-0.572	0.144	0.179	$[-0.953, -0.387]$	$[-0.923, -0.222]$	-4.650	-3.198	$< 0.001^{***}$	0.001^{**}
PA_{pc}	0.289	-0.167	0.064	0.085	$[0.163, 0.415]$	$[-0.334, 0.000]$	4.500	-1.958	$< 0.001^{***}$	0.050
PA_{int}	-0.018	0.442	0.066	0.086	$[-0.146, 0.111]$	$[0.274, 0.610]$	-0.270	5.164	0.788	$< 0.001^{***}$
$PA_{pc} \times PA_{int}$	-0.973	0.303	0.068	0.109	$[-1.107, -0.840]$	$[0.090, 0.515]$	-14.270	2.788	$< 0.001^{***}$	0.005^{**}
DAY[WD] $\times PA_{pc}$	-0.430	1.318	0.087	0.137	$[-0.600, -0.260]$	$[1.051, 1.585]$	-4.960	9.659	$< 0.001^{***}$	$< 0.001^{***}$
DAY[ND] $\times PA_{pc}$	0.151	0.653	0.088	0.119	$[-0.021, -0.324]$	$[0.420, 0.885]$	1.720	5.492	0.090	$< 0.001^{***}$
DAY[WD] $\times PA_{int}$	-0.023	-0.845	0.087	0.131	$[-0.192, 0.147]$	$[-1.103, -0.588]$	-0.260	-6.441	0.795	$< 0.001^{***}$
DAY[ND] $\times PA_{int}$ ✓	-0.393	-0.240	0.090	0.124	$[-0.570, -0.217]$	$[-0.482, 0.003]$	-4.360	-1.937	$< 0.001^{***}$	$0.053^{\cdot}$
DAY[WD] $\times PA_{pc} \times PA_{int}$	0.555	-0.734	0.081	0.157	$[0.396, 0.714]$	$[-10.43, -0.426]$	6.830	-4.668	$< 0.001^{***}$	$< 0.001^{***}$
DAY[ND] $\times PA_{pc} \times PA_{int}$	0.844	-0.505	0.077	0.116	$[0.694, 0.994]$	$[-0.732, -0.278]$	11.030	-4.356	$< 0.001^{***}$	$< 0.001^{***}$
ACT[Non-Work] ✓	-0.782	-0.352	0.070	0.084	$[-0.918, -0.645]$	$[-0.516, -0.188]$	-11.230	-4.210	$< 0.001^{***}$	$< 0.001^{***}$
ACT[Driving]	-0.068	0.684	0.079	0.102	$[-0.224, 0.087]$	$[0.484, 0.884]$	-0.860	6.706	0.388	$< 0.001^{***}$
$mCDN$ ✓	0.942	0.552	0.028	0.030	$[0.887, 0.997]$	$[0.493, 0.611]$	33.580	18.420	$< 0.001^{***}$	$< 0.001^{***}$
$SLEEP$	-0.092	0.086	0.038	0.048	$[-0.168, -0.018]$	$[-0.007, 0.179]$	-2.430	1.802	0.015^{*}	0.071^{**}
OCC[SP]	0.775	-1.477	0.499	0.740	$[-0.204, 1.753]$	$[-2.927, -0.027]$	1.550	-1.997	0.121	0.046^{*}
$NASA_{MD}$ ✓	0.472	0.534	0.047	0.072	$[0.380, 0.564]$	$[0.393, 0.675]$	10.040	7.435	$< 0.001^{***}$	$< 0.001^{***}$
$NASA_{PD}$	0.349	0.040	0.036	0.059	$[0.278, 0.420]$	$[-0.077, 0.156]$	9.610	0.673	$< 0.001^{***}$	0.501
$NASA_{P}$ ✓	0.070	0.244	0.032	0.053	$[0.007, 0.132]$	$[0.140, 0.347]$	2.190	4.621	0.029^{*}	$< 0.001^{***}$
$B5_{E}$	-0.559	1.324	0.277	0.436	$[-1.103, -0.016]$	$[0.469, 2.179]$	-2.020	3.036	0.044^{*}	0.002^{**}

labeling guided by participant debriefings and using GPS/cadence/speed signals as trigger validators. Then, using normalized HR measurements (NHR) and the statistical definition of significant differences, we were able to assign stress labels strictly within sedentary intervals of the STREL dataset. Such labels were reflective of participants' cognitive and emotional stress, without any undue influences from physical stress.

Other studies use HR/HRV data to estimate arousal and associated stress levels without considering the cardiovascular effect of physical activity [7], [8], [10]. Estimating stress agnostic to the occasional presence of physical activity leads unavoidably to overestimation of cognitive/emotional stress and potentially erroneous conclusions. This is evident in the SNS model results. For example, comparing Fig. 7b3 vs. Fig. 7a3, we observe that the micro-cadence curve in the SNS model starts at the probability of stress $\sim$ 50% vs. the probability of stress $\sim$ 25% in the NHR model. This suggests that SNS Index computations are contaminated by nonsedentary values, something that makes the micro-cadence effect of sedentary intervals look more like the full cadence effect of physical activity intervals.

Other factors also play a role in the apparent overestimation of stress by the SNS Index. As we discussed in Section 3.3.3, this is evident in some specific cases, such as participant T005. The stress labels derived from the SNS Index suggest that participant T005 was stressed almost all the time, including throughout the day ND (Fig. 4). This is in error because we know from the debriefings that participant T005 spent ND watching television for much of the day, an activity that is unlikely to induce high stress [46]. In this respect, our stress labeling method based on NHR

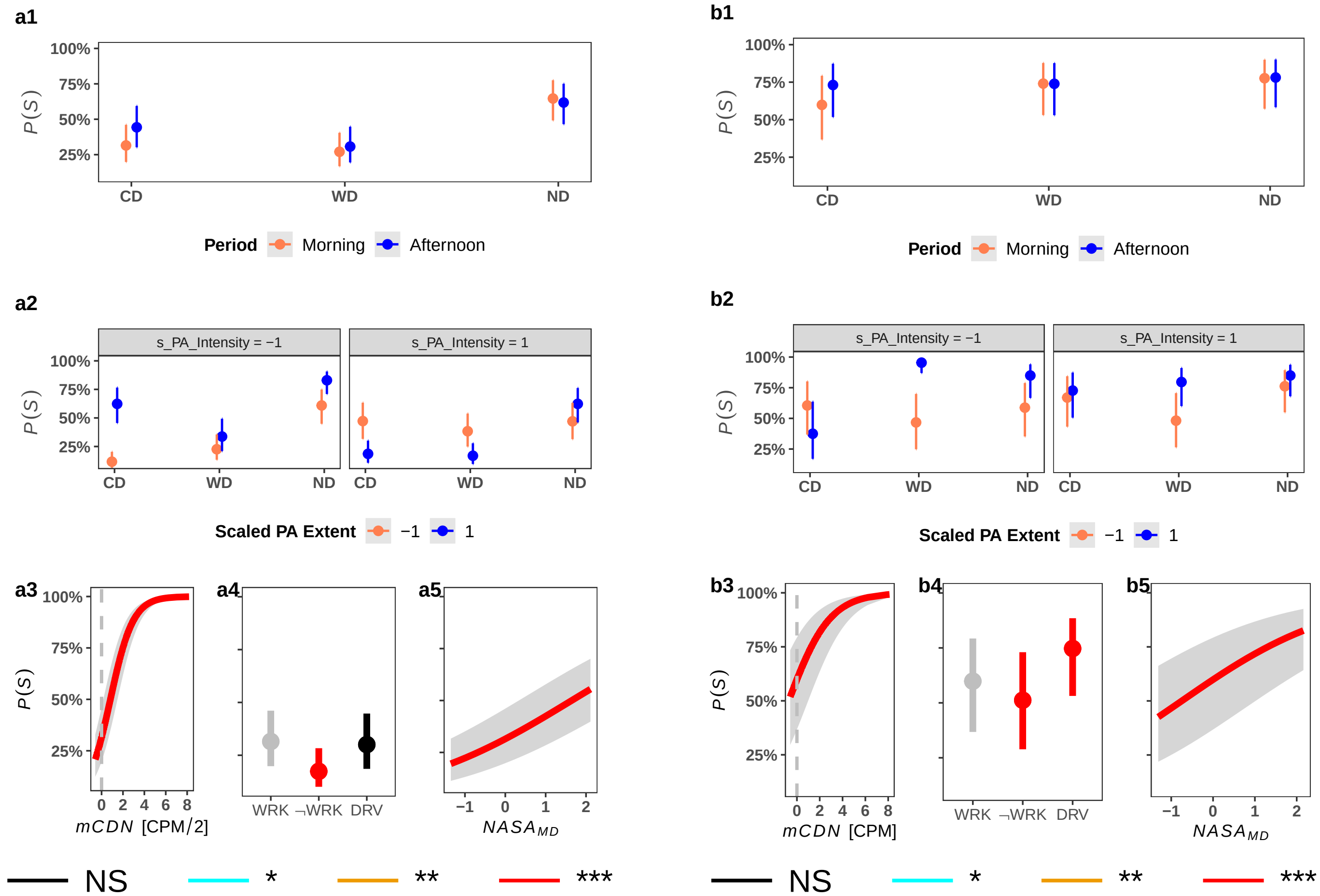


Fig. 7: Interaction and main effects of the optimized logistic regression models on stress derived from NHR_D (panels a1-a5) and SNS (panels b1-b5) data. a1-b1. Interaction effects between type of day (DAY) and period of day (PRD). a2-b2. Interaction effects between type of day (DAY), extent of physical activity said day (PA_{pc}), and intensity of physical activity said day (PA_{int}). a3-b3. Quantitative association of micro-cadence $mCDN$ with the probability of stressful responses for $n = 12709/11176$ instances of Work + Non-Work + Driving activities for NHR_D/SNS derived stress models, respectively. Each instance is a 5-min interval. a4-b4. Categorical associations of sedentary activities with the probability of stressful responses. WRK stands for Work, ¬WRK stands for Non-Work, and DRV stands for Driving. The frequencies of sedentary activities are as follows: $n_W = 6594/5856$ Work, $n_{NW} = 4784/4193$ Non-Work, and $n_D = 1331/1127$ Driving instances for NHR_D/SNS derived stress models, respectively. Each instance is a 5-min interval. a5-b5. Quantitative association of daily $NASA_{MD}$ scores with the probability of stressful responses for $n = 93/90$ days for NHR_D/SNS derived stress models, respectively.

suggests that T005 was mainly not stressed on day ND, mostly stressed on day CD, and somewhere in between on days WD. This stress account is in complete agreement with the participant's debriefings ((Supplemental Materials) and thus much more credible than the stress account presented by the SNS method. The likely reasons for these SNS failures include the lack of baseline correction and the sensitivity of HRV measures to transient vagal changes that abound during television viewing.

Despite their methodological and occasional labeling differences, the NHR and SNS stress methods give similar results to the main stress predictors in the logistic regression models (Table 1). However, examination of the cases where the two methods disagree suggests that the NHR results are in better agreement with the participants' debriefing accounts during low arousal activities (e.g., TV watching) and when physical activity borders sedentary periods. For this reason, NHR stress labels stand as a more reliable alternative to training and testing affective computing ML in naturalistic data, where the said behavioral patterns abound. Furthermore, because the NHR method uses only heart rate, it is more widely applicable than the SNS method that uses HRV measures, since not all smartwatches provide HRV or provide it frequently enough for continuous monitoring.

Stress and relaxation patterns of people in critical leadership positions: By building logistic regression models driven by NHR and SNS, where response variables were the odds of being cognitively/emotionally stressed in 5-minute sedentary intervals, we arrive at the following primary result. On CD days, where participants were in charge of a critical mission, the probability of becoming stressed increased as the day progressed. This was not the case on regular work (WD1 and WD2) and non-work (ND) days (Fig. 7a1/b1). The said result brings to the fore the psychophysiological effect of leadership, which has been missing from the literature.

Secondary but important results suggest that higher mental load and micro-cadence are positively associated with the probability of being stressed (Fig. 7a3/a5/b3/b5). The correlation with mental load points to the cognitive and emotional nature of the stress we are capturing, thanks to the accounting of physical activity that we introduced. The correlation with micro-cadence is likely specific to the two professions under examination. Leaders in emergency care and special police are not typical office-bound professionals who sit in front of a computer all day and barely move. They have to interrupt their office work with micro-trips constitutional to their mission. For example, a leading emergency care physician may have to check on a patient in the next booth, while a special police officer may need to accompany the VIP under his/her care to the next conference room. The frequency of such micro-trips presumably commensurates with the load and intensity of a participant's mission, which is reflected in the increased probability of stress.

Tertiary results show that the probability of being stressed is significantly higher when people work vs. when they do not work (e.g., relaxing at home) (Fig. 7a4/b4) - a common sense outcome. At the same time, the models show that driving is at least as stressful as working, which is in agreement with previous reports in the literature [27]. These two control results act as validity tests for the general goodness of our models.

In the primary, secondary, and tertiary results that we have discussed thus far, the NHR and SNS stress models agree, which is a testament to the reliability of these results. However, the models disagree on the three-way interaction between the type of day, the extent of physical activity, and the intensity of physical activity.

NHR and Physical Activity: The NHR model shows that the duration and intensity of physical activity are associated with lower mental stress on workdays but higher on non-workdays (Fig. 7a2). The beneficial role of physical exercise in the management of mental stress has been well documented in the literature [47], [48], and the workday result agrees with the consensus. The non-workday result can be interpreted in the context of positive excitement (eustress), which we know from the debriefings was prevalent in participants during non-workdays (Supplementary Materials). Hence, substantial physical exercise is likely related to suppression of distress and the promotion of eustress.

SNS and Physical Activity: The SNS model shows that substantial physical exercise is associated with a higher probability of mental stress on both workdays and non-workdays (Fig. 7b2). This result disagrees with not only the scientific literature but is also difficult to interpret. As we noted earlier, such stress overestimation likely stems from the occasional partaking in the SNS computation of heart function values corresponding to nonsedentary intervals.

Concluding Thoughts: The participants in the STREL data set probably have become leaders in their profession for a reason. From the analysis, we conclude that they have a natural way to manage mental stress. Specifically, they make the most out of their non-workday (ND), starting with excitement (eustress) that gives way to low arousal (relaxation) as the day winds down. This likely counterbalances the opposite pattern they experience on days they lead critical missions (CD), where negative stress (distress) takes hold and increases as the day progresses. The said result can serve as a behavioral paradigm for the management of mental stress in maladaptive cases. However, a legitimate question is how robust this result is. Although the participant size is moderate ($n = 24$), the models do not run at the participant level but at the 5-minute interval level with $n = 12709/11176$ observations. Hence, the models have considerable depth. It is also worth noting that the sample includes select individuals with certain characteristics due to the elite nature of their profession and their leading role in it. In other longitudinal naturalistic stress studies with similar participant characteristics and observational sizes (e.g., see the surgeon stress study by Pavlidis et al. [49]), the results stood the test of time.

Limitations: We emphasize that this naturalistic study included participants from emergency care personnel and special police forces, two critical and stressful professions. Therefore, although the methods we developed to address the questions surrounding this study can have broad applicability, the results we arrive at may not scale up for the general population.

Additional Information

The STREL dataset related to this article can be accessed in the Open Science Framework (OSF) data bank: https://osf.io/qshv7/.

The R code that operated on the STREL dataset, producing the results and figures presented in this paper, can be accessed on GitHub: https://github.com/UH-ACDC/STREL/.

Supplementary Materials accompany this paper. They include activity labels, NHR_D/SNS distributions, and brief summaries of debriefs for all 24 participants.

Acknowledgments

This research was funded through a grant from TKI Click NL.

References

[1] I. Bourgeault, J. Park, D. Kohen, J. Atanackovic, and Y. James, "A gendered analysis of work, stress and mental health, among professional and non-professional workers," *Professions and Professionalism*, vol. 11, no. 3, 2021.

[2] J. Mitchell, "Stress in emergency personnel," in *Stress: Concepts, Cognition, Emotion, and Behavior*, G. Fink, Ed. Elsevier, 2016, ch. 48, pp. 385–391.

[3] H. G. Xu, A. N. Johnston, J. H. Greenslade, M. Wallis, E. Elder, L. Abraham, O. Thom, E. Carlström, and J. Crilly, "Stressors and coping strategies of emergency department nurses and doctors: a cross-sectional study," *Australasian Emergency Care*, vol. 22, no. 3, pp. 180–186, 2019.

[4] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott *et al.*, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*. IEEE, 2011, pp. 97–108.

[5] J. A. Levine, I. T. Pavlidis, L. MacBride, Z. Zhu, and P. Tsiamyrtzis, "Description and clinical studies of a device for the instantaneous detection of office-place stress," *Work*, vol. 34, no. 3, pp. 359–364, 2009.

[6] P. Buddharaju, Y. Fujiki, I. Pavlidis, and E. Akleman, "A novel way to conduct human studies and do some good," in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 2010, pp. 4699–4702.

[7] N. Ahmadi, F. Sasangohar, T. Nisar, V. Danesh, E. Larsen, I. Sultana, and R. Bosetti, "Quantifying occupational stress in intensive care unit nurses: An applied naturalistic study of correlations among stress, heart rate, electrodermal activity, and skin temperature," *Human Factors*, vol. 64, no. 1, pp. 159–172, 2022.

[8] S. Hosseini, R. Gottumukkala, S. Katragadda, R. T. Bhupatiraju, Z. Ashkar, C. W. Borst, and K. Cochran, "A multimodal sensor dataset for continuous stress detection of nurses in a hospital," *Scientific Data*, vol. 9, no. 1, p. 255, 2022.

[9] N. E. Haouij, J.-M. Poggi, S. Sevestre-Ghalila, R. Ghozi, and M. Jaïdane, "AffectiveROAD system and database to assess driver's attention," in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, 2018, pp. 800–803.

[10] J. C. Yau, B. Girault, T. Feng, K. Mundnich, A. Nadarajan, B. M. Booth, E. Ferrara, K. Lerman, E. Hsieh, and S. Narayanan, "TILES-2019: A longitudinal physiologic and behavioral data set of medical residents in an intensive care unit," *Scientific Data*, vol. 9, no. 1, p. 536, 2022.

[11] J. L. Cahoon and L. A. Garcia, "Continuous stress monitoring for healthcare workers: Evaluating generalizability across real-world datasets," in *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2023, pp. 1–5.

[12] C. Queirós, F. Passos, A. Bártolo, A. J. Marques, C. F. Da Silva, and A. Pereira, "Burnout and stress measurement in police officers: Literature review and a study with the operational police stress questionnaire," *Frontiers in Psychology*, vol. 11, p. 587, 2020.

[13] S. Garbarino, G. Cuomo, C. Chiorri, and N. Magnavita, "Association of work-related stress with mental health problems in a special police force unit," *BMJ Open*, vol. 3, no. 7, p. e002791, 2013.

[14] S. Garbarino, C. Chiorri, and N. Magnavita, "Personality traits of the five-factor model are associated with work-related stress in special force police officers," *International Archives of Occupational and Environmental Health*, vol. 87, pp. 295–306, 2014.

[15] P. D. Harms, M. Credé, M. Tynan, M. Leon, and W. Jeung, "Leadership and stress: A meta-analytic review," *The Leadership Quarterly*, vol. 28, no. 1, pp. 178–194, 2017.

[16] B. H. Johnsen, H. K. Westli, R. Espevik, T. Wisborg, and G. Brattebø, "High-performing trauma teams: frequency of behavioral markers of a shared mental model displayed by team leaders and quality of medical performance," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 25, pp. 1–6, 2017.

[17] G. Matthews, "Distress," in *Stress: Concepts, Cognition, Emotion, and Behavior*, G. Fink, Ed. Elsevier, 2016, pp. 219–226.

[18] J. Kloidt and L. W. Barsalou, "Establishing a Comprehensive Hierarchical construct of Eustress (CHE)," *Current Psychology*, vol. 43, no. 41, pp. 32 258–32 273, 2024.

[19] Y. W. Goh, S. Sawang, and T. P. Oei, "The Revised Transactional Model (RTM) of occupational stress and coping: An improved process approach," *The Australasian Journal of Organisational Psychology*, vol. 3, pp. 13–20, 2010.

[20] C. Rott, M. Segers, and P. Van den Bossche, "Expected to be calm in any storm: An exploration of the stress experiences of crisis team leaders," *International Journal of Disaster Risk Reduction*, 2024.

[21] D. Sander, "Is "arousal," as a scientific concept, worse than useless?" *Emotion Review*, vol. 17, no. 1, pp. 19–22, 2024.

[22] J. Carson and E. Kuipers, "Stress management," S. Hardy, J. Carson, and B. Thomas, Eds. Stanley Thornes Publishers, 1998, p. 157.

[23] F. Lateef, "Grace under pressure: leadership in emergency medicine," *Journal of Emergencies, Trauma, and Shock*, vol. 11, no. 2, pp. 73–79, 2018.

[24] D. Mutton, "Stress and decision-making in police managers and leaders," in *Police Leadership & Management*. The Federation Press Sydney, 2007, pp. 190–200.

[25] M. A. Alpert, J. Omran, and B. P. Bostick, "Effects of obesity on cardiovascular hemodynamics, cardiac morphology, and ventricular function," *Current Obesity Reports*, vol. 5, pp. 424–434, 2016.

[26] R. R. McCrae and P. T. Costa, "Empirical and theoretical status of the five-factor model of personality traits," in *The SAGE Handbook of Personality Theory and Assessment*, G. J. Boyle, G. Matthews, and D. H. Saklofske, Eds. SAGE Publications Ltd, 2008, vol. 1, pp. 273–294.

[27] M. T. Hasan, H. Alghamdi, S. Taamneh, M. Manser, R. Wunderlich, P. Tsiamyrtzis, and I. Pavlidis, "Investigating cardiovascular activation of young adults in routine driving," *IEEE Transactions on Affective Computing*, 2023.

[28] R. Hernandez, S. C. Roll, H. Jin, S. Schneider, and E. A. Pyatak, "Validation of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) adapted for the whole day repeated measures context," *Ergonomics*, vol. 65, no. 7, pp. 960–975, 2022.

[29] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Advances in Human Psychology: Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Elsevier, 1988, pp. 139–183.

[30] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen, "Kubios HRV–heart rate variability analysis software," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 210–220, 2014.

[31] R. Baevskii, "Analysis of heart rate variability in space medicine," *Human Physiology*, vol. 28, no. 2, pp. 202–213, 2002.

[32] J. L. Broussard, A. C. Reynolds, C. M. Depner, S. A. Ferguson, D. Dawson, and K. P. Wright, "Circadian rhythms versus daily patterns in human physiology and behavior," in *Biological Timekeeping: Clocks, Rhythms and Behaviour*, V. Kumar, Ed. Springer, 2017, pp. 279–295.

[33] Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[34] S. M. Gwynne, A. L. Hunt, J. R. Thomas, A. J. Thompson, and L. Séguin, "The toilet paper: Bathroom dwell time observations at an airport," *Journal of Building Engineering*, vol. 24, p. 100751, 2019.

[35] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 553–560.

[36] C. Tudor-Locke and D. A. Rowe, "Using cadence to study free-living ambulatory behaviour," *Sports Medicine*, vol. 42, pp. 381–398, 2012.

[37] G. Matthews, "Stress states, personality and cognitive functioning: A review of research with the Dundee Stress State Questionnaire," *Personality and Individual Differences*, vol. 169, p. 110083, 2021.

[38] A. D. Crosswell and K. G. Lockwood, "Best practices for stress measurement: How to measure psychological stress in health research," *Health Psychology Open*, vol. 7, no. 2, pp. 1–12, 2020.

[39] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

[40] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. Springer, 2005.

[41] E. Smets, E. Rios Velazquez, G. Schiavone, I. Chakroun, E. D'Hondt, W. De Raedt, J. Cornelis, O. Janssens, S. Van Hoecke, S. Claes, I. Van Diest, and C. Van Hoof, "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *npj Digital Medicine*, vol. 1, no. 1, p. 67, 2018.

[42] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Continuous stress detection using a wrist device: in laboratory and real life," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1185–1193.

[43] G. Park and J. F. Thayer, "From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli," *Frontiers in Psychology*, vol. 5, p. 278, 2014.

[44] M. Malik and A. J. Camm, "Components of heart rate variability—what they really mean and what we really measure," *The American journal of Cardiology*, vol. 72, no. 11, pp. 821–822, 1993.

[45] J. Stanley, J. M. Peake, and M. Buchheit, "Cardiac parasympathetic reactivation following exercise: implications for training prescription," *Sports Medicine*, vol. 43, no. 12, pp. 1259–1277, 2013.

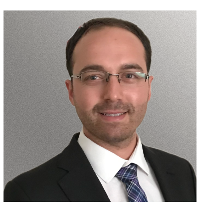
[46] K. J. Koruth, A. Lang, R. F. Potter, and R. L. Bailey, "A comparative analysis of dynamic and static indicators of parasympathetic and sympathetic nervous system activation during tv viewing," *Communication Methods and Measures*, vol. 9, no. 1-2, pp. 78–100, 2015.

[47] U. Rimmele, R. Seiler, B. Marti, P. H. Wirtz, U. Ehlert, and M. Heinrichs, "The level of physical activity affects adrenal and cardiovascular reactivity to psychosocial stress," *Psychoneuroendocrinology*, vol. 34, no. 2, pp. 190–198, 2009.

[48] N. J. Hegberg and E. B. Tone, "Physical activity and stress resilience: Considering those at-risk for developing mental health problems," *Mental Health and Physical Activity*, vol. 8, pp. 1–7, 2015.

[49] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin *et al.*, "Fast by nature-how stress patterns define human experience and performance in dexterous tasks," *Scientific Reports*, vol. 2, no. 1, p. 305, 2012.

Corinna Rott Corinna Rott is a psychologist specializing in stress regulation and team performance. She is currently a joint Ph.D. candidate at Maastricht University and the University of Antwerp, where her research integrates wearable technology with psychophysiological assessments. Through collaborations with the Affective Data and Computing Lab and the Antwerp Social Lab, Corinna has gained expertise in all stages of the research cycle, strengthening her skills in affective computing. She enjoys conducting data-driven debriefings, providing participants with personalized stress insights.



Fettah Kiran M.S. Kiran received his B.S. degree in 2013 from the Department of Computer and Instructional Technologies Education at Ondokuz Mayis University, Samsun, Turkiye. From 2013 to 2015 he worked as a content developer in the Distance Learning Center at the Canakkale Onsekiz Mart University in Turkiye. In 2020, he received his M.S. degree in Computer Science and Engineering from Louisiana State University, Baton Rouge, LA. He is currently a Ph.D. candidate in the Department of Computer Science at the University of Houston. His main research interests are human-computer interaction, affective computing, and naturalistic studies of stress and human performance.

Mien Segers Mien Segers holds the chair of Corporate Learning at the Maastricht University School of Business and Economics (the Netherlands). She is the founding faculty of the Learning and Development in Organizations master program at this school. Her research focuses on the question of how to support and facilitate impactful learning. In her studies, she pays particular attention to team learning. In this respect, she explores new methods to objectively trace online workplace learning processes and to use psychophysiological data to understand team interactions.

Piet Van den Bossche Piet Van den Bossche is a Professor of Learning in Organizations at the University of Antwerp and a Professor of Team Learning at Maastricht University. As codirector of the Antwerp Social Lab, he explores psychophysiological and behavioral methods to study human interactions. His research focuses on how social interactions shape learning and performance, with a particular emphasis on team learning, leadership, and professional development.

Ioannis Pavlidis Dr. Pavlidis is the Eckhard-Pfeiffer Distinguished Professor of Computer Science and Director of the Affective and Data Computing Laboratory at the University of Houston. His research is funded by multiple sources, including the National Science Foundation, transportation agencies, and medical institutions. He has published extensively in the areas of affective computing, data science, and science of science. He was the first to develop contact-free methods for measuring physiological variables, including electrodermal activity, breathing, and heart function, which he used to study stress in the wild.