The Hidden Load: Parenting Young Children While Leading in Critical Professions

Corinna Rott
Educational Research and Development
University of Maastricht
Maastricht, Limburg, Netherlands
corinna.rott@maastrichtuniversity.nl

Fettah Kiran
Computer Science
University of Houston
Houston, Texas, USA
fkiran@uh.edu

Malgorzata W. Kozusznik
Department of Marketing, Innovation, and Organisation
Ghent University
Ghent, Belgium
gosia.kozusznik@ugent.be

Mien Segers
School of Business and Economics
University of Maastricht
Maastricht, Netherlands
m.segers@maastrichtuniversity.nl

Piet Van den Bossche
University of Antwerp
Antwerp, Belgium
piet.vandenbossche@uantwerpen.be

Ergun Akleman
Visualization
Texas A&M University
College Station, Texas, USA
ergun.akleman@gmail.com

Ioannis Pavlidis
Computer Science
University of Houston
Houston, Texas, USA
ipavlidis@uh.edu

Abstract

Parenting while serving as a frontline leader is uniquely stressful, yet little is known about how family responsibilities shape physiological stress in these roles. We followed emergency physicians and tactical police leaders, comparing parents of young children with non-parents across four days: one critical mission day, two standard workdays, and one non-workday. Using wearable sensing, expert activity labeling, and daily debriefs, we inferred stress only in sedentary epochs via a normalized-heart-rate method, with an HRV-based index as benchmark. Parents showed higher stress on workdays and non-workdays, but not on critical mission days, where attentional narrowing and strict device policies appear to suppress parenting-related differences. We contribute: (i) in-the-wild physiological evidence that parenthood amplifies stress mainly under permeable boundaries, (ii) a pragmatic stress-labeling pipeline for safety-critical settings, (iii) a configuration-based account linking boundaries, attention, and parenting, and (iv) design implications for stress-aware boundary management systems, supported by an open analysis repository.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Occupational stress, Parenting and caregiving, Boundary permeability, Attentional control, In-the-wild stress monitoring, Wearable sensing, Heart rate variability (HRV), Normalized heart rate (NHR), Frontline leadership, Boundary management technologies

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790570

ACM Reference Format:
Corinna Rott, Fettah Kiran, Malgorzata W. Kozusznik, Mien Segers, Piet Van den Bossche, Ergun Akleman, and Ioannis Pavlidis. 2026. The Hidden Load: Parenting Young Children While Leading in Critical Professions. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3772318.3790570

1 Introduction

Professionals in frontline leadership roles, such as emergency physicians and tactical police officers, routinely move between life-or-death situations and ordinary administrative work [9, 36]. Some days, they coordinate resuscitations, mass-casualty responses, or tactical operations under strict protocols and intense time pressure. On other days, they lead clinical rounds, scheduled procedures, training, or report writing. Many of these professionals are also parents of young children, here defined as preteen children in preschool or primary school, and must juggle high-stakes responsibility at work with caregiving, logistics, and emotional labor at home.

A large body of work in occupational health and family studies shows that parenting young children can be profoundly meaningful and chronically taxing [37, 38]. Parenting stress is shaped not only by the child's needs, but also by job demands, schedule constraints, and the availability of support. In parallel, research on public-safety personnel and other high-risk occupations shows that sustained exposure to acute stressors can affect performance and health, with heart rate (HR) and heart rate variability (HRV) emerging as key physiological indicators of allostatic load [3, 8, 18, 26, 43]. Yet we still know relatively little about how these two strands intersect in everyday life: how the responsibilities of parenting young children shape physiological stress for frontline leaders as they move across critical incidents, standard workdays, and non-workdays.

At the same time, personal smartphones and messaging platforms have transformed how people remain reachable across life domains. Throughout this paper, when we refer to "device use" or "device policies", we specifically mean participants' personal smartphones, not the research wearables. Smartphones make it possible to coordinate childcare, respond to school or daycare issues, and handle family logistics from almost anywhere. Research on digital availability and work–family boundaries shows that this constant reachability can both help and harm: it enables flexible coordination, but also blurs the boundaries between work and home, increasing overload and stress when expectations of responsiveness become too strong [6, 34, 52]. Especially for parents of young children, the very channels that enable care can also create a sense of being "always on".

In this paper, we bring these threads together by studying stress patterns among frontline professionals in high-stakes health and security roles, comparing those who are parents of young children to those who are not. We examine how their physiological stress varies across critical days, standard workdays, and non-workdays, and how this variation relates to the ways in which work and family can or cannot permeate one another in everyday practice.

1.1 Stress, attention, and boundary permeability

Theoretically, our work is informed by two complementary perspectives on how stress unfolds across roles: *attentional narrowing* in crises and *boundary permeability* between work and family.

Attentional narrowing. First, accounts of goal-directed behavior and role transitions suggest that high-stakes tasks can trigger strong attentional narrowing or "goal shielding" [4, 13, 15, 16]. In acute and tightly scripted situations, for example, leading a resuscitation or coordinating a tactical intervention, professionals must maintain a focus on protocol, safety, and team coordination. External concerns, including family-related thoughts, may recede from conscious awareness as attention is channeled into a narrow task envelope. From this perspective, critical incidents may produce intense but comparatively "clean" stress: physiological load is high, yet cognitively it is tied to a single, dominant role.

Boundary permeability. Second, research on boundary theory and digital availability emphasizes that stress is not only a psychological phenomenon, but also socially and technologically mediated [4, 52]. People differ in how permeable they prefer the boundaries between work and non-work to be, and in how much control they feel over these boundaries. Empirically, work-related smartphone use outside work hours is associated with higher strain and work–family conflict when it undermines detachment and recovery [11, 12, 34, 39, 53]. HCI research has extended this lens by examining how notification systems and communication tools shape perceived availability and boundary control. Schüss and Gross' CHI 2019 work on *ad-hoc availability* shows how people dynamically negotiate being reachable across domains, and how this hybrid state – neither fully "on" nor "off" – relates to stress [40]. Jung et al.'s notion of *pragmatic boundaries* argues that notification systems should respect users' boundary preferences and contexts, rather than assuming universal norms of constant connectivity [25].

Parenting intensifies these boundary tensions. Parents of young children often maintain ad-hoc availability towards family: they may need to be reachable for urgent caregiving issues, school calls, or coordination with partners. Psychological work shows that parenthood can be associated with both elevated daily hassles and deep sources of meaning, depending on resources and context [37, 38]. When work is demanding and family needs are frequent, digital channels can become conduits through which multiple roles collide.

Our context adds a further structural layer: organizational device policies. In our study, frontline professionals in high-stakes health and security roles face markedly different communication regimes across days. On *critical days*, tactical police leaders and emergency physicians are subject to strict, protocol-driven constraints: personal devices are typically switched off, stored away, or otherwise inaccessible during high-stakes operations. On *standard workdays*, these same professionals may keep phones nearby, respond to messages during quieter moments, and juggle routine leadership tasks with family coordination. On *non-workdays*, family demands and leisure activities dominate, often interwoven with digital communication.

Taken together, these perspectives suggest that stress in such roles is shaped by a *configuration* of factors rather than a single mechanism. Acute critical incidents may generate strong attentional narrowing and protocol-driven device detachment. Standard workdays, by contrast, often involve more permeable boundaries and ad-hoc availability, allowing family preoccupations and caregiving logistics to intersect with professional concerns. Non-workdays constitute yet another configuration: formal work demands recede, while family life, childcare, and household organization become central.

1.2 Gaps in knowledge

Despite extensive research on parenting stress, digital availability, and high-risk occupations, several gaps remain:

(1) **Physiological parenting stress in safety-critical roles.** Most work on parenthood and stress relies on self-report measures of well-being or daily hassles in general working populations [37, 38]. We lack in-the-wild physiological evidence on how parenting young children shapes embodied stress for frontline leaders in safety-critical roles, particularly across different types of workdays and non-workdays.
(2) **Day-to-day trajectories across crises, routine work, and time off.** Research on public-safety personnel and other high-risk professions has begun to examine HR and HRV as indicators of occupational load [26], but typically focuses on either isolated critical incidents or aggregated workload. There is limited work following the same individuals across alternating periods of acute crises, routine leadership, and time away from work, while simultaneously considering family responsibilities.
(3) **Linking digital availability and boundaries to physiological stress.** The HCI literature on digital availability and boundary management has largely centered on knowledge

workers or everyday users of smartphones and messaging systems [25, 34, 40, 52]. These studies richly describe how people experience being reachable and how technologies mediate boundaries, but rarely connect these dynamics to physiological stress or to occupations with strict device protocols.

Consequently, we know little about how parenting young children modulates physiological stress for frontline professionals in high-stakes health and security roles under different boundary regimes: when personal devices are strictly prohibited, when availability is informally negotiated during routine leadership, and on non-workdays when formal work demands recede and everyday life — including family relationships, domestic tasks, socializing, and leisure — moves to the foreground.

1.3 Our approach

To address these gaps, we conducted an in-the-wild, mixed-methods observational study of **46 frontline professionals in high-stakes health and security roles** in the Netherlands and Switzerland. Our sample comprised emergency physicians and tactical police leaders who held leadership responsibilities in critical operations, including both professionals who were parents of at least one preteen child (preschool or primary school age) and colleagues who were not parents. We followed each participant across three types of days:

- **Critical days (CD):** days on which participants were responsible for coordinating a team in high-stakes operations, such as emergency physicians leading resuscitations or trauma responses, and tactical police leaders directing large-scale tactical exercises or operations.
- **Standard workdays (WD1, WD2):** routine leadership days dominated by scheduled clinical work, supervision, training, or bureaucratic tasks (e.g., report writing, planning, debriefing), with intermittent on-call responsibilities.
- **Non-workdays (ND):** days without professional duty, typically involving family time, domestic work, socializing, and leisure activities; for many participants, these leisure periods included physically active pursuits such as biking or running.

Participants wore a research-grade heart-monitoring device continuously, allowing us to derive HR- and HRV-based indicators of physiological stress while controlling for physical activity. We complemented these signals with structured debrief interviews and daily logs capturing sleep, work periods, commute, and contextual information about caregiving and household responsibilities. Our analysis focuses on *day-level* stress signatures rather than moment-by-moment trajectories. This reflects both the temporal resolution that we can robustly support with our physiological measures and the granularity of contextual information available from debriefings, which often characterize activities over extended periods (e.g., a day dominated by report writing and briefings) rather than precise minute-by-minute labels.

Methodologically, we use logistic mixed-effects models to estimate the probability that a given 5-minute epoch reflects elevated stress, combining HR- and HRV-based markers and including microcadence as a covariate to adjust for movement. We then aggregate these estimates to compare day-level stress patterns across critical days, standard workdays, and non-workdays, with parenthood and other individual factors as predictors.

1.4 Key findings

Our analysis reveals a nuanced picture of how parenting young children interacts with professional role and day type:

- On **critical days**, both parents and non-parents show elevated physiological stress, consistent with the high stakes and time pressure of acute operations. In this regime, strict protocols and enforced device detachment appear to concentrate attention on the professional role, and we do not observe strong additional amplification of stress for parents beyond this already high baseline.
- On **standard workdays**, parents of young children exhibit higher day-level stress than their non-parent peers, even after controlling for physical activity and individual differences. These days combine routine leadership responsibilities with more permeable boundaries: participants can be reached by family, and caregiving logistics (e.g., pickups, schedule changes) frequently intersect with work demands.
- On **non-workdays**, our models still indicate frequent periods of elevated physiological arousal for both parents and non-parents. Parents show a higher modeled probability of such episodes than their non-parent colleagues, but debrief interviews suggest that these peaks are often linked to positively valenced experiences. In our sample, many non-parents described non-workdays as combining purposeful physical activity with later periods of rest and relaxation, whereas many parents described being positively busy with children and family activities across much of the day. Participants typically characterized these episodes as energizing or enjoyable rather than as distress.

Taken together, these results point to a distinctive pattern of parenting-related differences in physiological stress. The clearest differences between parents and non-parents do not arise during the acute crises themselves, but in everyday configurations: in the *hidden load* of managing ongoing caregiving and family logistics under more permeable boundaries on standard workdays, and in positively charged, high-activity family periods on non-workdays. Because our physiological markers capture intensity rather than valence, debriefs are essential for interpreting whether heightened arousal reflects distress or positively valenced engagement. In line with the terminology we use later in the paper, we treat these as instances of *distress* versus *eustress*: in our sample, much of the elevated arousal on non-workdays is described as eustress for both parents and non-parents, whereas many workday episodes are characterized as distressful or draining.

1.5 Contributions

Together, our work makes contributions across empirical, methodological, theoretical, design, and reproducibility dimensions:

C1 – Empirical: We provide in-the-wild physiological evidence on how parenting young children interacts with day type to shape stress among frontline professionals in high-stakes health and security roles. By following the same individuals across critical days, standard workdays, and non-workdays, we show that parenthood amplifies physiological stress markers mainly in routine and family-dominated contexts, where boundaries are more permeable, rather than simply adding to the already intense load of acute crises, where access to personal devices and family contact is tightly constrained.

C2 – Methodological: We demonstrate a robust approach to studying stress in naturalistic, safety-critical settings that combines continuous HR- and HRV-based markers with logistic mixed-effects modeling and explicit control for physical activity. Our day-level aggregation strategy acknowledges the limits of moment-by-moment interpretation in such contexts while still capturing meaningful differences across roles and days.

C3 – Theoretical: We integrate boundary theory, digital availability research, and work on parenting stress to articulate how *configurations* of attentional demands, device policies, and family responsibilities shape stress across days. Rather than attributing effects to a single mechanism (e.g., "device detachment"), we conceptualize critical, standard, and non-workdays as distinct boundary regimes in which parenting exerts different influences on both distress and eustress.

C4 – Design implications: For the CHI community, we extend existing work on ad-hoc availability and pragmatic boundaries [25, 40] into safety-critical and parenting-intensive contexts. Our findings suggest opportunities for technologies that better align communication practices with boundary preferences and caregiving obligations – in line with broader HCI work on calm technologies and stress-aware systems that seek to monitor and mitigate load unobtrusively [27, 42, 55]. For example, by supporting structured family coordination before and after critical operations (while respecting strict no-communication protocols during the operations themselves), and by helping parents negotiate and reflect on their availability on standard workdays without feeling "always on".

C5 – Reproducibility: To support transparency and reuse, we share our data and analysis code for our stress-inference pipeline on GitHub: UH-ACDC/Parenting-CHI-2026. This includes model specifications, feature extraction scripts, and descriptions of our data-cleaning procedures.

2 Study Design

Based on an institutionally approved protocol by the Ethics Review Committee of the University of Maastricht (No. ERCIC 535 24 02 2024)—the institution's IRB-equivalent body—we recruited 48 volunteers for an observational field study. Two participants did not yield sufficient physiological data for the stress-labeling analyses: in both cases, heart rate (HR) was recorded, but the Polar H10 chest strap providing RR-intervals for heart rate variability (HRV) was not consistently paired with the watch across the four protocol days, despite detailed instructions. Because a central aim of the study was to compare stress labeling based on HR alone with stress indices that additionally draw on HRV, we required HRV data across all four protocol days for each participant. For these two participants we could not compute HRV-based indices in a consistent way, so they were excluded from model-based analyses, leaving $n = 46$ in the final analyzed sample ($n = 14$ emergency physicians with leadership responsibilities in the Netherlands, $n = 32$ tactical police officers in supervisory or team-lead roles in Switzerland). This corresponds to an attrition rate of about 4% at the participant level. All participants provided written informed consent.

We designed the observation period to capture within-person contrasts across high-stakes work, routine work, and time away from work. For each participant, we scheduled four target days in advance, based on institutional duty rosters and the participant's availability:

- **Critical workday (CD).** For emergency physicians, the CD was a shift in which the participant was designated as responsible for coordinating resuscitations, triage, or other acute emergency care (e.g., trauma bays, rapid response). For tactical police officers, the CD involved operational protection of VIPs or high-value detainees, or participation in large-scale tactical exercises (e.g., hostage negotiation and intervention scenarios) with clearly defined operational protocols and time pressure.
- **Standard workdays (WD1 and WD2).** Standard days were scheduled on typical weekdays without major planned incidents. For physicians, these days generally involved scheduled clinical rounds, outpatient clinics, and administrative work. For tactical police officers, they typically involved planning, report writing, or lower-stakes training activities.
- **Non-workday (ND).** The ND was usually a weekend or other day with no professional duties. Participants' descriptions indicated that these days predominantly involved physical activities (e.g., cycling, running, hiking), family time (e.g., playground visits, parties), socializing, and household chores.

Day types were thus determined prospectively from duty schedules and institutional plans, not relabeled retrospectively based on how stressful a day turned out to be. On critical days in both professions, organizational protocols constrained personal smartphone use more tightly than on standard workdays or non-workdays. For tactical police officers, personal smartphones were often required to be switched off and kept off their person (e.g., stored in lockers or vehicles) during live or simulated operations. For emergency physicians, personal smartphone use was explicitly forbidden in operating rooms and other sterile procedural areas, and strongly discouraged during resuscitations and other acute emergencies. In contrast, on routine workdays and non-workdays, participants generally had more discretion over if and when to check personal smartphones. Our analyses draw on this naturally occurring variability in boundary permeability, rather than on experimentally manipulating device access.

Taken together, this protocol yields a conceptual 2×2 structure that we probe in later analyses. Parenthood (having preteen children vs. not) varies *between* participants, while boundary regime varies *within* participants: critical workdays combine high-stakes tasks, strict operational protocols, and sharply reduced access to personal communication channels, whereas standard workdays and non-workdays represent two forms of more permeable boundaries with greater discretion over device use. In terms of our theoretical lens (Section 1.1), critical days are those in which attentional narrowing and protocol-driven device detachment concentrate cognition on a single dominant role, while standard and non-work days are those in which attention and communication are more diffusely spread across professional and family concerns. Our models therefore test how parenthood and boundary regime, and their interaction, shape stress responses in this 2×2 configuration, while analytically distinguishing routine work from non-work within the more permeable regime (Section 4).

2.1 Study Data

Daily Debriefs: At the end of each target day, participants completed a brief structured debrief conducted by the study recruiter, a trained psychologist on our team. Most debriefs were carried out by phone. The debriefs captured (i) main activities and locations across the day (e.g., "morning resuscitations, afternoon clinic," "family bike ride, birthday party"), (ii) salient stressors and positive events, and (iii) contextual details such as commuting, sleep timing, and caregiving responsibilities. These debriefs provided qualitative insights that aided in labeling and interpreting activity periods identified from sensors.

Below are excerpts from participant T030 (tactical police officer, father of young children), illustrating the contrasts between CD and ND. To illustrate the richness of the debriefing data, we constructed a "data story" in comic-strip form (Fig. 1). This approach builds on prior CHI work that combines empirical observation with artistic synthesis to convey lived experience while preserving anonymity [1]. By translating daily self-reports into an anonymized narrative, the comic helps readers emotionally connect with participants' experiences, complementing the quantitative analyses that follow.

> *On day CD he was leading a tactical assessment group with noticeable increases in sympathetic arousal. Tactical evaluations were ongoing throughout the morning and afternoon that day.*
>
> *On day ND, participant T030 went biking with his kids in the morning. Biking was at a slow pace to allow his toddler to catch up. In the afternoon, the family celebrated a friend's birthday. The birthday party was filled with fun activities for the children.*

Beyond supporting activity labeling, the debriefs also yielded contextual information relevant to personal smartphone use and communication boundaries. Participants described where they typically kept their personal smartphones on different workdays (e.g., in a locker, in a pocket, on a desk) and when organizational protocols explicitly constrained use (e.g., during live operations or surgical procedures). For the tactical police cohort, these retrospective accounts were complemented by non-participant observation in the situation room on large-scale tactical exercise days: a psychologist from our research team was present as an observer, taking field notes but not interacting with officers or intervening in operations. This mode of non-participant observation is well established in organizational and healthcare research [14, 31, 32, 45], and prior work suggests that carefully briefed observers have limited impact on routine behavior in high-reliability settings [47].

Importantly, the research protocol itself introduced only the wearable devices described below; these Polar watches and chest straps have no cellular connectivity and cannot be used for calls, messaging, or other forms of online communication. Personal smartphones and any other work-issued devices remained under existing institutional policies. Participants were informed that the study did not alter those policies. Our references to "device use" in the paper specifically concern personal smartphone availability and use in these established organizational contexts.

Biographic and Personality Measures: At enrollment, participants completed a short biographic questionnaire (occupation, age, sex, family status, height, weight, and home/work location). Family status, particularly whether participants had preteen children, informed analyses of family-related stress. Body Mass Index (BMI), computed from self-reported height and weight, was included as a covariate due to its influence on cardiac signals [2]. Personality was assessed with the Big Five Inventory [35], providing trait-level predictors (e.g., neuroticism has been shown to amplify stress responses in critical professions [17]).

Daily Workload Measures: At the end of each target day, participants completed the NASA Task Load Index (NASA-TLX), rating mental demand, physical demand, temporal demand, performance, effort, and frustration [20]. We used the standard unweighted (raw) TLX ratings, which are commonly applied in naturalistic and in-the-wild studies where repeated pairwise weighting is impractical [21, 23]. Beyond quantifying subjective workload, NASA-TLX offers a structured account of how participants experienced their daily demands. This provides an explanatory path to observed physiological stress responses: while physiology reflects autonomic activation, NASA-TLX captures perceived demands across multiple dimensions of work and life. In naturalistic settings, where no single indicator is definitive, this complementarity improves construct validity and strengthens interpretation of stress-labeling results [41].

Physiological and Activity Data: Participants wore Polar Vantage V2 smartwatches with H10 chest straps (Polar, Kempele, Finland) continuously for four days, including sleep. These devices were approved by the participating institutions as non - communication fitness devices under local protocols. The watches and chest straps recorded motion and cardiac signals at high resolution; GPS and cadence data supported activity classification (e.g., sedentary, walking/running, cycling, driving). Validation work indicates that Polar devices provide sufficiently accurate HRV estimates for stress-related field studies when paired with a chest strap [24]. Heart rate (HR) and heart rate variability (HRV) were used to compute stress labels via two complementary methods: (i) HR-baseline–adjusted stress labeling during sedentary periods, and (ii) HRV-derived Sympathetic Nervous System Index (SNS) and Stress Index (SI) via Kubios software [5, 48]. Sleep periods were self-reported in the daily debriefs and validated against changes in cardiac and movement patterns.

We chose to center the protocol on cardiac measures, foregoing additional biosensing channels such as respiration or electrodermal activity (EDA). While EDA and respiration can provide valuable information about sympathetic activation and cognitive load, they typically require extra devices and skin contact sites, which would have increased participant burden and posed greater risks of non-compliance in these demanding professions. Concentrating on HR and HRV from a single wearable configuration provided a pragmatic compromise between ecological validity, data quality, and feasibility in a four-day in-the-wild study of emergency physicians and tactical police officers.

Figure 1: A "data story" derived from participant T030's daily debriefing data. The comic-strip rendering synthesizes key signals and self-reported experiences into a narrative that conveys the subject's lived experience while preserving anonymity. Background elements (e.g., buildings and landscapes) were rendered with the aid of AI-based tools under the direction of the research team, whereas the human figures were hand-drawn by the project's illustrator. Such artistic renderings, grounded in empirical data, foster empathy and engagement in naturalistic studies where video footage is unavailable.

3 Methods

3.1 Data Aggregation and Summarization

To achieve time co-registration, all quantitative, ordinal, and categorical data were integrated into a single data frame. Each row corresponds to one participant p_i, one day d_j, and one 5-minute epoch t_k.

Quantitative signals. Physiological and motion data were recorded using the Polar Vantage V2 smartwatch paired with an H10 chest strap, including heart rate (*HR*, beats per minute), inter-beat intervals (*RR*, ms), cadence (*CDN*, cycles/min), speed (*SPD*, km/h), and duration of sleep (*SLEEP*, hours). Derived measures included the Kubios Stress Index (*SI*) and Sympathetic Nervous System Index (*SNS*), both widely used indicators of physiological arousal and benchmarks for stress labeling.

Ordinal measures. Personality traits (Big Five subscales) and daily workload (NASA-TLX subscales) provided trait- and state-level predictors of stress responses.

Categorical variables. Each record was annotated with participant ID (unique code), occupation (*OCC*; emergency physician vs. tactical police leader), day type (*DAY*; critical workday [CD], standard workday 1 [WD1], standard workday 2 [WD2], non-workday [ND]), and period of day (*PRD*; morning [0–12h] vs. afternoon [12–24h]), the latter reflecting circadian cycles known to modulate physiology and behavior [7].

Temporal aggregation. High-frequency signal data (up to 1000 Hz) determined the temporal resolution of the analysis. Following cardiological guidelines [50], HR and HRV were summarized in non-overlapping 5-minute intervals. The same epoching was applied to all other sensor streams (e.g., *SI*, *SNS*, cadence, speed) for temporal alignment. Five minutes is a standard window for short-term HRV analysis and a naturalistic behavioral unit that can capture short activities and transitions (e.g., walking to a meeting) without pretending to resolve second-by-second events [19]. All subsequent activity and stress labels are defined at this 5-minute resolution.

3.2 Signal Preprocessing and Data Cleaning

All HRV analyses were performed on R–R interval series exported to the Kubios HRV software [48, 49]. Preprocessing comprised four sequential steps:

1. *Beat detection.* R peaks were detected using the Polar H10 internal detector, with R–R intervals sampled at high temporal resolution and imported into Kubios.
2. *Noise detection.* Kubios' automatic noise detection was used to identify and exclude segments with excessive noise or data loss (e.g., poor chest-strap contact, loss of signal).
3. *Artifact and ectopic-beat correction.* Remaining artifacts and ectopic beats were corrected using Kubios' threshold-based method with cubic spline interpolation, as recommended for short-term HRV analysis. Epochs that remained severely corrupted after this step were removed from further analysis.
4. *Detrending.* Slow trends in the R–R series were removed using the Kubios smoothness priors method with a cutoff below 0.04 Hz, as recommended for HRV and stress analysis [49].

HR-based measures (including *HR* and our normalized heart rate, *NHR*) are more robust to brief motion artifacts than HRV features, and were therefore available for a larger fraction of epochs. As a result, some 5-minute intervals contribute to HR- and NHR-based analyses but not to HRV-based SNS analyses. We return in Section 3.4 to the conceptual reasons for using NHR as our primary stress channel in the in-the-wild setting, while treating SNS as an HRV-based benchmark.

3.3 Activity Labeling

In naturalistic stress studies, activity labeling is often underemphasized, yet it is a prerequisite for meaningful stress categorization. Differentiating physical from non-physical activities is paramount: only during non-physical activities can observed physiological arousal be confidently attributed to mental stress rather than exertion. Each 5-minute interval is therefore assigned to one of the following mutually exclusive activity states:

- **Sleep** – intervals between self-reported bed- and wake-times, validated with smartwatch sleep analytics.
- **Physical Activity** – locomotor activity with elevated cadence and/or speed (e.g., running, biking, brisk walking). Running and cycling segments were identified with high sustained values; walking segments, often indoors (e.g., within the hospital), were characterized by intermediate cadence and were harder to distinguish from sedentary states on sensor evidence alone.
- **Driving** – GPS displacement along road networks with high speed and near-zero cadence.
- **Sedentary Work** – GPS anchored at the workplace with low cadence and no physical-activity flags.
- **Sedentary Non-Work** – GPS anchored at home or other non-work locations with low cadence.
- **Unknown** – rare intervals with absent signal or inconsistent sensor evidence; these were excluded from analyses.

Only *sedentary work*, *sedentary non-work*, and *driving* constitute non-physical activities in which mental (cognitive and emotional) stress can be disentangled from physical exertion. These three categories form the eligible set for stress labeling; sleep, physical activity, and unknown intervals are excluded from stress labeling. Importantly, the labels are defined at the 5-minute level and do not aim to capture moment-by-moment micro-activities within those intervals (e.g., exactly when a meeting started or ended).

Expert labeling procedure. Our labeling was performed in three steps:

Step 1: A data scientist on the team applied a deterministic protocol that mapped time-stamped debriefing notes onto sensor-verified criteria (GPS trajectory, speed, cadence, location).

Step 2: A study psychologist independently reviewed the resulting label sequences. Concordance was complete for driving and outdoor physical-activity segments. Ambiguities arose only in some indoor walking segments without debrief notes and with missing GPS signal (e.g., walking to the cafeteria), where cadence and speed signals were the only available indicators.

Step 3: These few ambiguous epochs were resolved through follow-up questions with participants, resulting in consensus labels. Because the procedure was rule-driven and all disagreements were fully adjudicated, conventional inter-rater reliability statistics (e.g., Cohen's κ) were not informative; instead, the process produced consensus ground truth.

Driving as a sedentary activity. We treat driving as part of the non-physical activity set because participants are seated with minimal gross motor output; steering and pedal use impose negligible metabolic load compared with walking or running. Under this operationalization, elevated *NHR* or *SNS* during driving reflects mental (cognitive and emotional) stress expressed in a non-exertional posture, with contributions both from the driving task itself and from whatever else participants are thinking about (e.g., work or family concerns). Our analyses therefore include driving in the sedentary pool for stress labeling and interpret any stress episodes during driving as mental stress in a seated context.

Figure 2 visualizes the activity labels for participant T030 against smartwatch signals over the four observation days. Prior work on in-the-wild labeling has often relied on continuous participant self-reports fused with sensor streams to obtain ground truth labels [51]. Our approach differs in that labeling was carried out by an expert dyad (data scientist + psychologist) applying deterministic rules and sensor criteria, with participants only consulted in rare ambiguous cases. This reduced participant burden while still ensuring label validity and consensus ground truth. As such, our protocol advances activity labeling for naturalistic stress research by offering a reproducible, low-burden alternative to continuous self-reporting.

3.4 Stress Labeling

In stress studies, it is fundamental to track participants' evolving stress levels. This is achieved through stress labeling. The conventional approach is based on self-reporting, which works well in laboratory experiments where participants complete preplanned tasks and then report their perceived stress [33]. However, self-reporting becomes impractical in naturalistic studies, where continuous monitoring spans days or weeks rather than short, scripted tasks. In such settings, stress labeling is increasingly derived from physiological measurements, which capture autonomic responses that are often inaccessible to conscious awareness and less prone to recall bias [41]. These advantages make physiology a viable basis for stress labeling in the wild.

Several physiologically driven methods have been proposed, and new ones continue to emerge. The Baevskii Stress Index (SI), computed from HRV parameters [5], is one such method. Another is the Kubios Sympathetic Nervous System (SNS) Index, which integrates HRV-derived features including SI and the normalized Poincaré plot index SD2 [28, 29, 48].

No single stress labeling method has become dominant. To ensure robustness and capture complementary aspects of stress responses [10], we adopt two approaches: (i) a new HR-based labeling method that leverages normalized heart rate (NHR), a signal broadly available on consumer smartwatches, and (ii) the Kubios SNS Index as an established HRV-based benchmark.

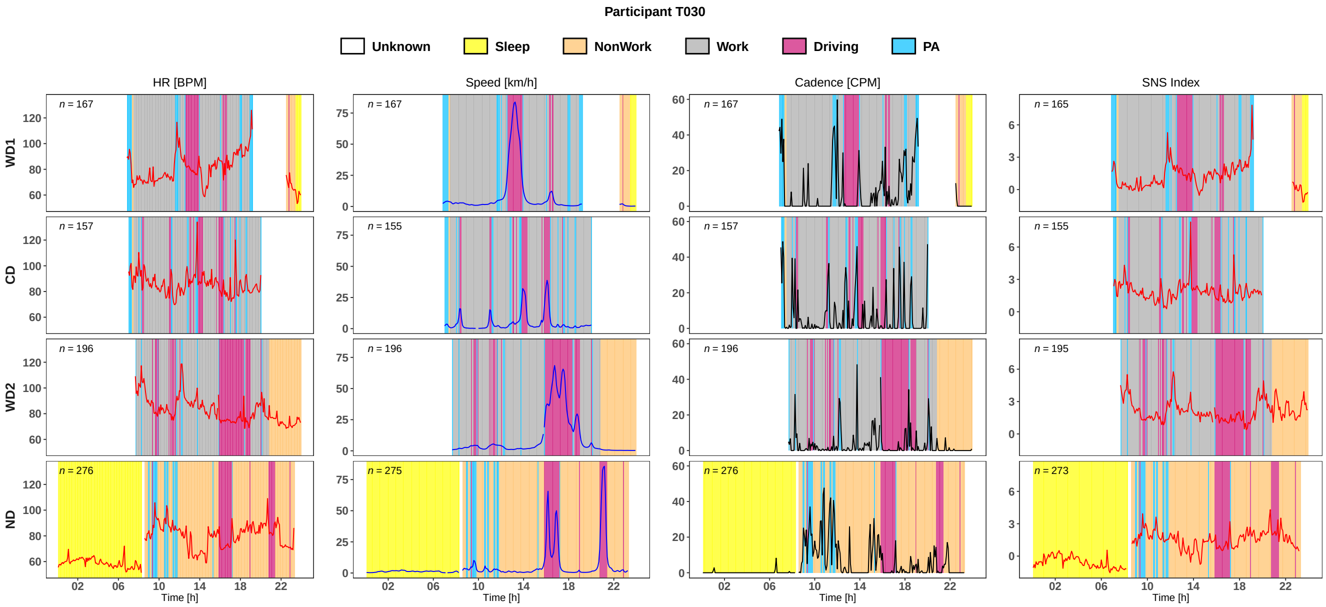

Figure 2: Activity labeling of participant T030. Visualization of activity labels on days WD1, CD, WD2, and ND. Labels are conveyed as background colors in four columns of signal panels. The speed and cadence signals in the middle columns are instrumental for activity validation, while HR and SNS Index signals in the outer columns track physiological arousal and provide an additional plausibility check. For instance, when cadence is high, HR and SNS typically rise, reflecting sympathetic activation due to physical activity. Driving was classified as sedentary and therefore included in subsequent stress labeling analyses. This triangulation across signals ensured that labels aligned with physiological plausibility.

3.4.1 NHR Stress Labeling

We propose a new stress labeling method based on normalized heart rate (*NHR*). By defining stress as deviation from an individual's daily baseline, this method offers a simple and reproducible way to capture mental stress during daily life.

Defining the Null hypothesis (No Stress). For each participant p_i and day d_j, we calculate NHR by subtracting the device-provided daily baseline $HR_{BL}(p_i, d_j)$ from the continuous HR signal $HR(p_i, d_j, t_k)$, where each moment t_k corresponds to a 5-minute interval. The baseline $HR_{BL}(p_i, d_j)$ is given by the Polar device as that day's baseline HR, estimated from low-activity periods, and provides a day-specific reference that accommodates diurnal and context-related shifts. We then restrict analysis to sedentary periods, forming the sub-signal $NHR_{\mathcal{D}}(p_i, d_j, t_l) \subset NHR(p_i, d_j, t_k)$, where t_l corresponds to sedentary work, sedentary non-work, or driving. Equation (1) expresses this formulation:

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

Here, $NHR_{\mathcal{D}} = 0$ represents the daily baseline of the participant, which we treat as the fixed null hypothesis of "No Stress" (NS). Because stress episodes create asymmetry, the empirical mean of the sedentary $NHR_{\mathcal{D}}$ distribution, $\overline{NHR}_{\mathcal{D}}(p_i, d_j)$, is often not equal to zero. This operationalization is *valence-agnostic*: deviations from baseline can reflect both positively and negatively valenced high-arousal states (eustress and distress). Because we do not have exhaustive information about events that would allow us to reliably assign valence at fine temporal resolution, we focus on day-level patterns rather than attempting to label each 5-minute interval as "good" or "bad" stress.

Deviation relative to the No Stress reference. Rather than centering data and applying z-scores, we retain the empirical distribution and assess deviations from $NHR_{\mathcal{D}} = 0$ using the observed standard deviation $\mathrm{SD}_{NHR_{\mathcal{D}}}(p_i, d_j)$ during sedentary periods. The stress threshold is defined as:

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

Assuming approximate normality (empirically supported; see QQ plots in Figs. 3a2/b2 and 3a4/b4), this threshold captures the upper 2.5% of expected variation under the Null. Any value of $NHR_{\mathcal{D}}$ exceeding this threshold is considered a stress response (S), reflecting significant autonomic activation beyond routine variation. Because analysis is limited to sedentary intervals, these responses can be attributed to cognitive and emotional stress rather than gross motor activity.

Empirical and visual support. Figure 3 (a1/b1 and a3/b3) shows the sedentary $NHR_{\mathcal{D}}$ and SNS distributions across CD, WD, and ND for two illustrative participants, with dotted lines marking the fixed reference point and the $+2 \cdot SD$ threshold. The corresponding QQ plots (a2/b2 and a4/b4) suggest approximate normality within sedentary intervals. Figure 4 illustrates how the $+2 \cdot \mathrm{SD}$ threshold is applied relative to the fixed No Stress reference point of $NHR_{\mathcal{D}} = 0$, even when the empirical distribution is shifted.

Theoretical and applied justification. This approach follows classical statistical inference, in which deviation from a fixed Null is tested using the observed variance [30, 54]. It also aligns with

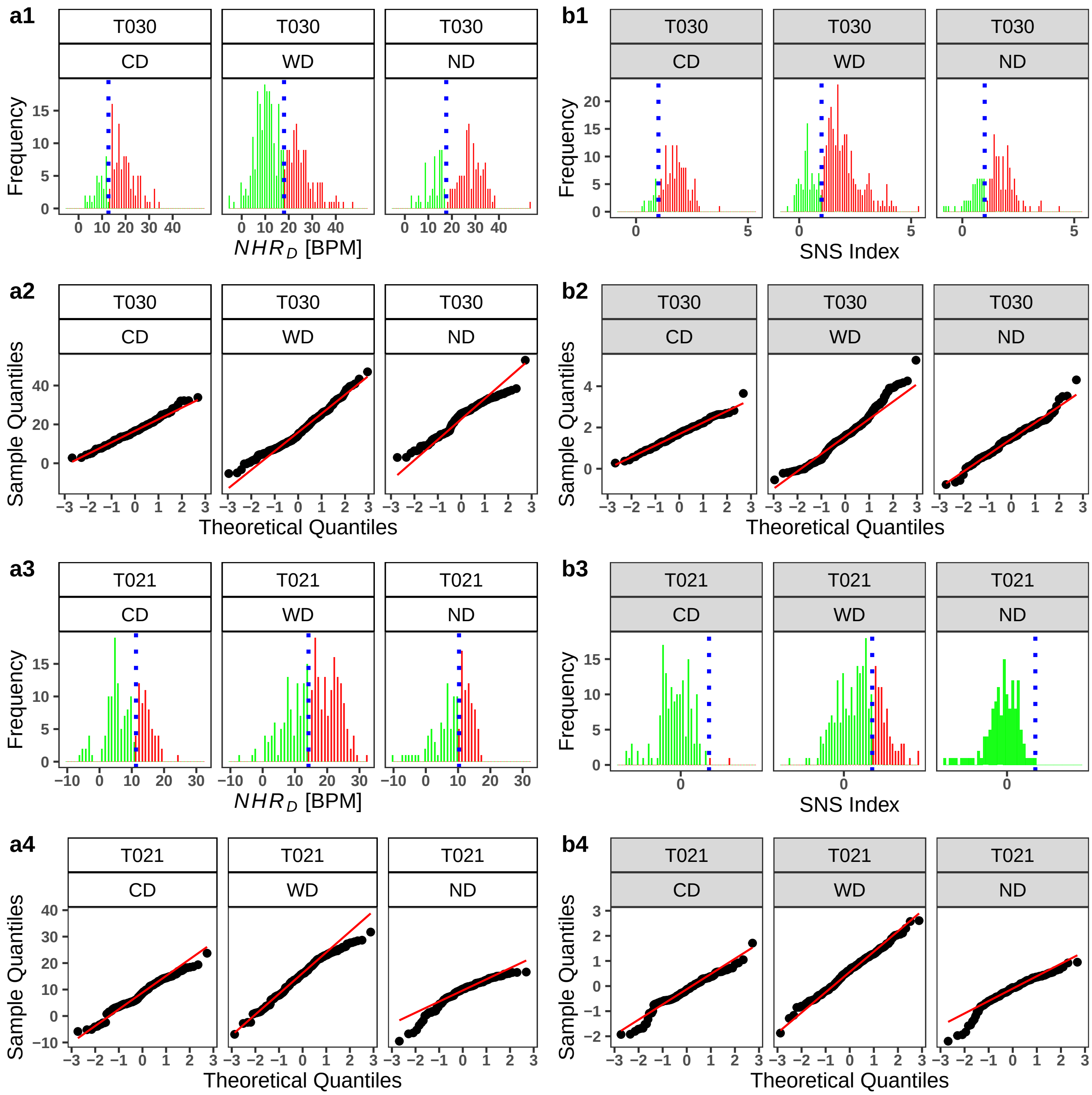

Figure 3: Stress labeling for participant T030 (parent with small children) and T021 (non-parent). a1/b1 and a3/b3: Distributions of sedentary NHR / SNS **on CD, WD = WD1 ∪ WD2, and ND. Blue dotted lines indicate thresholds (**$2 \cdot$ **SD from** $NHR = 0$ **and** $SNS = 1$**). a2/b2 and a4/b4: QQ plots of sedentary** NHR / SNS**, suggesting approximate normality. These plots illustrate how thresholds are applied within each method; they are not intended for direct comparison of individual labels.**

recent wearable computing literature that advocates for personalized, context-aware stress thresholds in the wild [44]. Importantly, HR-based deviations relative to a daily baseline are less affected by motion artifacts and non-stationary R–R dynamics than HRV features, particularly in the field. HRV indices are highly sensitive to artifacts and non-stationary behavior even during apparent sedentary periods, and can remain suppressed for some time after physical activity due to prolonged parasympathetic withdrawal and accumulated metabolites [22, 46]. In contrast, HR recovers more quickly and can be reliably sampled on consumer devices at 5-minute resolution. These considerations make NHR a practical and reproducible stress labeling protocol for large-scale studies with consumer wearables.

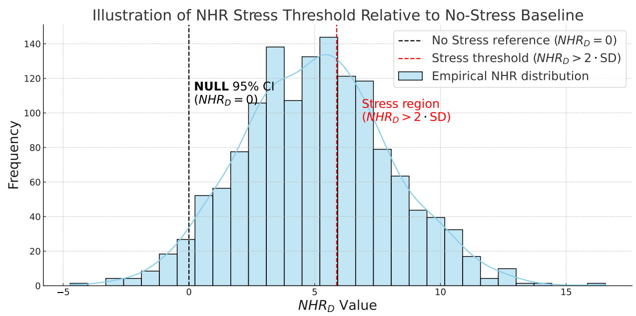

Figure 4: Illustration of the $NHR_{\mathcal{D}}$ stress threshold relative to the No Stress reference ($NHR_{\mathcal{D}} = 0$).

3.4.2 *SNS Stress Labeling*

For comparison, we derived an alternative set of stress labels using an established HRV-based measure. Specifically, we focus on the Stress Index (SI) and the Sympathetic Nervous System Index (SNS) provided by the Kubios software [5, 48]. In our dataset, SI and SNS are highly correlated (Pearson $r = 0.89, p < 0.001$); therefore, we only used the SNS Index (SNS). Following Kubios guidelines, we adopted the threshold $SNS = 1$ to distinguish between Stress and No Stress states. We applied this threshold across $SNS(p_i, d_j)$ distributions of participant p_i on day d_j to produce an alternative labeling of stress (see Fig. 3b1/b3).

Rather than comparing NHR and SNS labels one by one, we assess whether regression models fitted on the two sets of labels produce similar patterns of predictor effects. Convergent results strengthen confidence in the modeling conclusions and suggest that the HR-based NHR method approximates the more established HRV-based SNS method, while being simpler and more robust in naturalistic settings.

4 Results

4.1 Collinearity Check

Before modeling, we examined correlations among all predictor variables (Fig. 5). Only a small number of pairs showed high correlations. Specifically, $NASA_{MD}$ was strongly correlated with $NASA_{TD}$ ($r = 0.73$) and with $NASA_E$ ($r = 0.69$), while $OCC[SP]$ was strongly correlated with $NASA_F$ ($r = 0.63$). Thus, we excluded $NASA_{TD}$, $NASA_E$, and $NASA_F$ from further modeling. This left three NASA-TLX subscales in the final predictor set. All other predictor pairs showed weak correlations ($|r| < 0.5$), and no further exclusions were necessary. As noted earlier, $OCC[SP]$ denotes the occupation category (Tactical Police vs. Emergency Care, with EC as the reference).

4.2 Descriptive Statistics

The sample included 46 adult participants (37 male, 9 female; age = 39.30 ± 6.29). Fourteen were leaders in emergency care units and 32 were leaders in tactical police forces. Participants had a healthy, non-obese physique (BMI = 26.25 ± 3.6), so no obvious cardiovascular confounders were present. Key predictors included cadence, sleep, and psychometrics.

Baseline HR across day types. Daily baseline HR values provided by the Polar device were very similar across day types (ND: 62.1 ± 11.0 bpm, WD1: 61.5 ± 11.3 bpm, WD2: 63.8 ± 11.5 bpm, CD: 62.6 ± 10.0 bpm). In a linear mixed model with DayType (ND, WD1, WD2, CD) as a fixed effect and a random intercept for participant, DayType did not have a significant effect on baseline HR, $F(3, 129) = 1.14$, $p = .33$ (all Tukey-corrected pairwise comparisons non-significant). Thus, there is no evidence for systematic anticipatory tonic stress on critical days at the level of daily baselines, and the NHR transformation (deviation from a day-specific baseline) does not build in a bias favoring any particular day type.

Age and gender. The study was intentionally targeted at mid-career professionals in two male-dominated occupations, resulting in a relatively narrow age range and a strongly skewed gender distribution (9 women out of 46 participants). In exploratory models that added age and gender as covariates, the main day-type and parenthood effects reported below were unchanged. However, given the restricted age range and small number of women, we treat age and gender as boundary conditions of the present sample rather than as primary explanatory variables, and we avoid over-interpreting any age or gender coefficients.

Cadence (CDN). Figure 6a1 shows the distributions of CDN in sedentary versus physically active periods. Sedentary values (n = 23,705, 3.7 ± 6.50 CPM) follow an exponential shape concentrated near zero, reflecting occasional micro-movements (termed micro-cadence, $mCDN$). Physically active values (n = 3,292, 38.91 ± 17.74 CPM) are fat-tailed and span higher ranges, with minimal overlap with sedentary values. Physiological arousal during sedentary periods is therefore interpretable as mental stress, whereas during physical activity it reflects exertion. In our models, we include both $mCDN$ and indicators of physical activity extent (PA_{pc}) and intensity (PA_{int}).

Sleep ($SLEEP$). Figure 6a2 shows the sleep distribution (n = 167, 6.72 ± 1.46 hr). The distribution is approximately normal with mean around seven hours; shorter sleep was more common on workdays, while longer sleep appeared on non-workdays.

Daily psychometrics. NASA TLX scores provide subjective complements to physiological indicators. Among the subscales, NASA Mental Demand ($NASA_{MD}$) and NASA Performance ($NASA_P$) later emerged as significant model predictors (Section 4.3). Figure 6a3 shows $NASA_{MD}$ stratified by day type: highest on critical days (n = 46, 12.11 ± 5.39), lowest on non-workdays (n = 45, 4.78 ± 4.44), and intermediate on standard workdays (n = 80, 9.24 ± 5.11). These marginal differences validate the designation of day types. In contrast, $NASA_P$ scores (Fig. 6a4) remained consistently high across critical days (15.04 ± 5.33), standard workdays (16.25 ± 3.35), and non-workdays (13.02 ± 5.83), reflecting participants' overall effectiveness.

Trait psychometrics. Figure 6a5 shows Big Five scores. The participants were very conscientious (8.07 ± 1.50) and low in neuroticism (4.48 ± 1.38), consistent with the demands of their professions. Agreeableness was also high (6.67 ± 1.74), supporting the collaborative nature of their roles. Extraversion (6.30 ± 1.81) and openness (6.87 ± 2.01) spanned broader ranges. Together, these patterns confirm a selective, high-performing professional cohort.

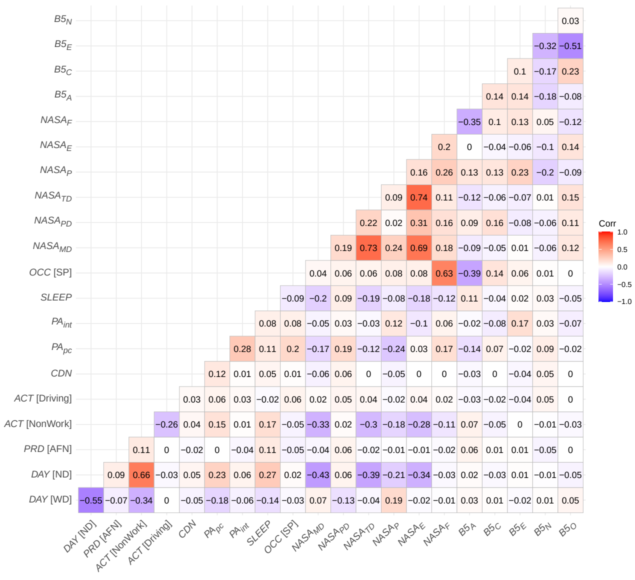

Figure 5: Cross-correlation matrix for all model predictors. Only a few predictor pairs exceeded the threshold for collinearity, while the majority showed weak or negligible correlations.

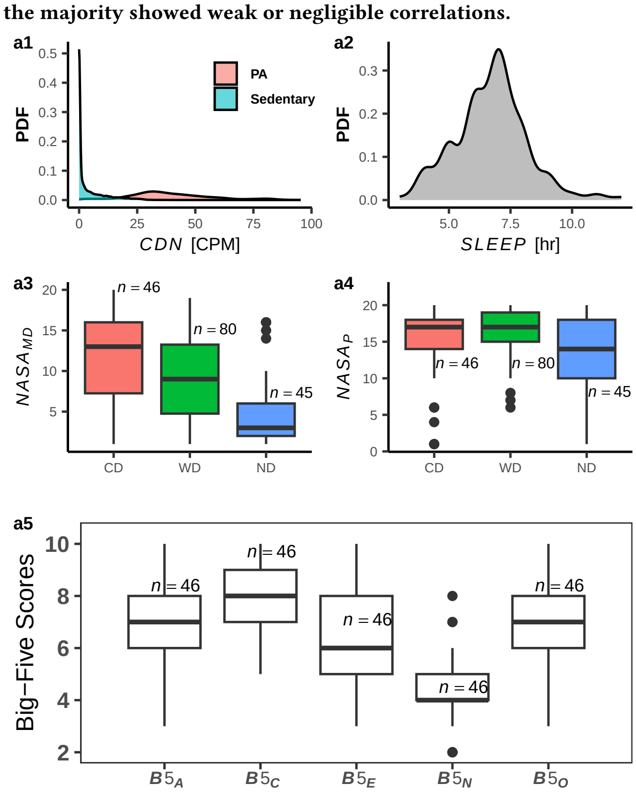

Figure 6: Descriptive plots of modeling variables. a1. Probability density functions of Cadence (CDN) for sedentary and physical activities (PA). a2. Probability density function of sleep duration. a3. Box-plots of NASA TLX Mental Demand ($NASA_{MD}$) per day type. a4. Box-plots of NASA TLX Performance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

4.3 Models

We first provide a descriptive overview of the sedentary activity intervals in our dataset before turning to model-based analyses of stress predictors. This approach allows us to ground the regression results in the empirical distribution of activities across participants and days.

In total, the dataset contained $n_W = 13{,}027$ Work, $n_{NW} = 8{,}011$ Non-Work, and $n_D = 2{,}664$ Driving sedentary intervals, for a combined $n = 23{,}702$ intervals across all participants and days. These descriptive frequencies confirm that Work was the predominant sedentary activity, with substantial though smaller contributions from Non-Work and Driving contexts. The subsequent regression models (Table 1, Fig. 7) build directly on these data, estimating how stress probabilities varied across these activity types, day and period factors, and individual differences. This connection between raw activity distributions and model-based stress predictions grounds the inferential results in the empirical structure of the dataset.

We examine predictors of participants' instantaneous stress while they are at work, not at work, or driving. The unit of analysis is the 5-minute sedentary interval t_l for participant p_i on day d_j, during which the participant is classified as either stressed ($\mathcal{S}$) or not stressed ($\mathcal{NS}$). The outcome variable is the probability $P(S_v(p_i, d_j, t_l))$ of stress, where the labeling channel v is either $NHR_{\mathcal{D}}$ or SNS. Stress intervals are defined as those where $NHR_{\mathcal{D}} > 2\text{SD}$ or $SNS > 1$; all other intervals are labeled as no stress. These thresholds are designed to capture hyperarousal of emotional or cognitive origin. Because the activities under study are sedentary (work, non-work, driving), physical exertion is unlikely to account for the observed stress responses.

$$
\begin{aligned}
&\operatorname{logit}\!\big(P(S_v(p_i, d_j, t_l))\big) \sim \\
&b_0 + b_1\, DAY[\text{WD}](p_i) + b_2\, DAY[\text{ND}](p_i) \\
&\quad + b_3\, PRD[\text{AFN}](p_i, d_j) + b_4\, DAY[\text{WD}] \times PRD[\text{AFN}](p_i, d_j) \\
&\quad + b_5\, DAY[\text{ND}] \times PRD[\text{AFN}](p_i, d_j) + b_6\, KID[\text{YES}](p_i) \\
&\quad + b_7\, DAY[\text{WD}] \times KID[\text{YES}](p_i) + b_8\, DAY[\text{ND}] \times KID[\text{YES}](p_i) \\
&\quad + b_9\, PA_{pc}(p_i, d_j) + b_{10}\, PA_{int}(p_i, d_j) + b_{11}\, PA_{pc} \times PA_{int}(p_i, d_j) \\
&\quad + b_{12}\, DAY[\text{WD}] \times PA_{pc}(p_i, d_j) + b_{13}\, DAY[\text{ND}] \times PA_{pc}(p_i, d_j) \\
&\quad + b_{14}\, DAY[\text{WD}] \times PA_{int}(p_i, d_j) + b_{15}\, DAY[\text{ND}] \times PA_{int}(p_i, d_j) \\
&\quad + b_{16}\, DAY[\text{WD}] \times PA_{pc} \times PA_{int}(p_i, d_j) \\
&\quad + b_{17}\, DAY[\text{ND}] \times PA_{pc} \times PA_{int}(p_i, d_j) \\
&\quad + b_{18}\, ACT[\text{Non-Work}](p_i, d_j, t_l) + b_{19}\, ACT[\text{Driving}](p_i, d_j, t_l) \\
&\quad + b_{20}\, mCDN(p_i, d_j, t_l) + b_{21}\, SLEEP(p_i, d_j) + b_{22}\, OCC[\text{SP}](p_i) \\
&\quad + b_{23}\, NASA_{MD}(p_i, d_j) + b_{24}\, NASA_{PD}(p_i, d_j) + b_{25}\, NASA_P(p_i, d_j) \\
&\quad + b_{26}\, B5_A(p_i) + b_{27}\, B5_C(p_i) + b_{28}\, B5_E(p_i) \\
&\quad + b_{29}\, B5_N(p_i) + b_{30}\, B5_O(p_i) + (1 \mid \text{Participant}).
\end{aligned}
\tag{3}
$$

Eq. (3) is the full logistic regression model where (1|Participant) introduces random intercepts to account for repeated measures per participant. The model is structured as follows:

Timing predictors (Rows 1–3). The first three rows capture day type (DAY[WD], DAY[ND]; reference = critical day CD), period of day (PRD[AFN]; reference = morning), and their interactions.

Parenthood predictor (Rows 3-4). KID[YES] indicates whether the participant has small children (reference = no children). Interactions with day type test whether stress dynamics differ for parents.

Physical activity predictors (Rows 5–9). PA_{pc} is the percentage of daily time spent in physical activity; PA_{int} is activity intensity. Their main effects and interactions with day type are included, as well as the three-way interaction.

Sedentary activity predictors (Row 10). *ACT* specifies the type of sedentary activity during each 5-minute interval: Non-Work or Driving, with Work as the reference.

Behavioral and contextual controls (Row 11). *mCDN* measures micro-cadence (small bursts of movement during sedentary periods), *SLEEP* is daily sleep duration, and *OCC* represents occupation (Emergency Care = reference; Tactical (Special) Police = SP).

Psychometric predictors (Rows 12–14). NASA-TLX subscales (mental demand $NASA_{MD}$, physical demand $NASA_{PD}$, performance $NASA_P$) are included at the daily level. Big Five personality traits ($B5_A$, $B5_C$, $B5_E$, $B5_N$, $B5_O$) are participant-level predictors.

The parameter estimates for the model specified in Eq. (3) are summarized in Table 1, while key effects are visualized in Fig. 7. For clarity, we report only predictors that reached statistical significance in at least one version of the model. Non-significant predictors (e.g., some of the Big Five personality traits) are omitted, as they were not theoretically central and did not contribute meaningfully to stress prediction. This presentation highlights the variables most relevant to our research questions while maintaining a parsimonious interpretation of the results.

Overall, the results indicate several consistent patterns across both stress labeling methods. Stress was significantly lower on typical workdays compared to the critical day baseline, while non-workdays showed mixed effects depending on the labeling channel. Having small children strongly amplified stress responses on both workdays and non-workdays, underscoring the role of parenthood in shaping daily stress dynamics. Physical activity showed protective effects, with both greater duration (PA_{pc}) and higher intensity (PA_{int}) associated with reduced stress, though interaction terms revealed that these benefits depended on day type. Among activity contexts, Non-Work intervals were reliably associated with lower stress compared to Work, whereas Driving showed divergent effects across the two models. Beyond contextual factors, stress probability was positively related to micro-cadence (*mCDN*) and NASA-TLX workload ratings, particularly mental and physical demand. In contrast, most Big Five personality traits were not significant predictors. Taken together, these results highlight the interplay between situational context (day, period, activity type), role demands (parenthood, occupation), and individual differences (workload appraisals, micro-movement) in determining momentary stress responses.

4.3.1 Robustness Analysis: Simplified Design-Only Model

We also fitted a simplified logistic mixed-effects model using the NHR-based stress labels, in order to assess the robustness of our findings to model complexity. This reduced model retained only day type, time of day, parenthood, physical activity patterns, activity context, and micro-cadence, while omitting sleep, occupation, NASA-TLX subscales, and Big Five traits:

$$\begin{aligned}\mathrm{logit}\big(P(S_v(p_i, d_j, t_l))\big) \sim\, & mCDN_{ijl} + DAY_{ij}\times PRD_{ij} + ACT_{ijl} \\ & + DAY_{ij}\times PA_{pc_{ij}}\times PA_{int_{ij}} \\ & + DAY_{ij}\times KID_i + (1 \mid Participant_i). \end{aligned} \tag{4}$$

where the outcome variable is the probability $P(S_v(p_i, d_j, t_l))$ of stress, with labeling channel $v = NHR_{\mathcal{D}}$ at each sedentary 5-minute interval.

The simplified model closely reproduces the core contextual effects obtained in the full model. Micro-cadence remains a strong positive predictor of stress (e.g., $\beta = 0.75$, SE = 0.02, $p < .001$, odds ratio ≈ 2.1 per SD). The odds of stress are substantially lower on typical workdays than on critical days at baseline ($\beta_{\mathrm{WD}} = -1.06$, $p < .001$). Non-work and driving contexts show reduced stress relative to work (e.g., Non-Work vs. Work: $\beta = -0.61$, $p < .001$). The Day × Kids interaction also remains large and highly significant (DAY[WD]$\times KID$[YES]: $\beta = 0.81, p < .001$; DAY[ND]$\times KID$[YES]: $\beta = 0.57$, $p < .001$), indicating that parents experience markedly higher stress probabilities than non-parents on both typical workdays and non-workdays, but not on critical mission days. The pattern of Day-by-physical-activity interactions is also qualitatively unchanged, with significant Day-specific effects of PA extent and intensity and their three-way interaction.

Together, the simplified and full models therefore tell a consistent story: our main conclusions about day types, parenthood, and stress are not artifacts of model complexity, but reflect stable patterns that emerge even in a more economical specification.

5 Discussion

Our study followed frontline leaders in emergency medicine and tactical policing as they moved across critical mission days, standard workdays, and non-workdays, comparing those who were parenting preteen children with colleagues who were not. Using in-the-wild physiological sensing, expert activity labeling, daily workload ratings, and mixed-effects models, we showed that parenting-related differences in stress are highly contextual: they are most pronounced on routine workdays and non-workdays under permeable boundaries, and largely absent on critical days when attention and communication are tightly constrained. Below we discuss how these findings refine current understandings of parenting stress, boundary permeability, and attentional control, and what they imply for design, organizational practice, and future research.

5.1 Parenthood and Contextual Stress

Across both NHR- and SNS-based stress labels, parenthood did not exert a uniform main effect. Instead, parenthood interacted strongly with day type. In the full and simplified models alike, parents of young children showed substantially higher stress probabilities than non-parents on standard workdays and non-workdays, but not on critical mission days. The robustness of these interactions across stress channels and model specifications suggests that they are not artifacts of modeling choices, but reflect stable patterns in how caregiving responsibilities intersect with daily role configurations.

This pattern resonates with *allostatic load theory*, in which acute episodes of high demand may be absorbed similarly across groups, while repeated hassles and role conflicts in everyday life generate

Table 1: Results of the logistic regression models specified in 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 with stress labels from SNS. Predictors highlighted in yellow indicate variables where the two model versions are in agreement. For clarity, only predictors that reached statistical significance in at least one model are reported. Significance levels: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

Predictor	Coefficient b.		Standard Error		95% CI		z value		Pr(> |z|)	
Intercept	−0.459	1.363	0.365	0.623	[−1.174, 0.255]	[0.142, 2.583]	−1.260	2.118	0.208	0.029^{*}
DAY[WD] ✓	−0.861	−0.792	0.087	0.110	[−1.031, −0.691]	[−1.008, −0.575]	−9.919	−7.168	$< 0.001^{***}$	$< 0.001^{***}$
DAY[ND]	0.341	−0.673	0.117	0.138	[0.112, 0.571]	[−0.943, −0.403]	2.919	−4.880	0.004^{**}	$< 0.001^{***}$
PRD[AFN]	0.101	−0.186	0.064	0.079	[−0.024, 0.226]	[−0.340, −0.031]	1.581	−2.359	0.114	0.018^{*}
DAY[WD] × PRD[AFN]	0.117	0.397	0.080	0.095	[−0.038, 0.273]	[0.211, 0.583]	1.476	4.187	0.140	$< 0.001^{***}$
DAY[ND] × PRD[AFN]	−0.132	0.289	0.098	0.116	[−0.325, 0.060]	[0.061, 0.517]	−1.346	2.481	0.178	0.013^{*}
KID[YES] ✓	−0.207	−0.506	0.369	0.600	[−0.930, 0.516]	[−1.682, 0.671]	−0.562	−0.842	0.574	0.400
DAY[WD] × KID[YES] ✓	0.705	0.832	0.085	0.113	[0.539, 0.872]	[0.611, 1.053]	8.285	7.385	$< 0.001^{***}$	$< 0.001^{***}$
DAY[ND] × KID[YES] ✓	0.563	0.818	0.099	0.117	[0.368, 0.757]	[0.588, 1.048]	5.671	6.965	$< 0.001^{***}$	$< 0.001^{***}$
PA_{pc}	−0.126	0.409	0.044	0.049	[−0.212, −0.040]	[0.312, 0.505]	−2.865	8.275	0.004^{**}	$< 0.001^{***}$
PA_{int}	−0.181	0.030	0.043	0.056	[−0.266, −0.096]	[−0.079, 0.140]	−4.175	0.539	$< 0.001^{***}$	0.590
$PA_{pc} \times PA_{int}$ ✓	−0.399	−0.474	0.046	0.068	[−0.489, −0.308]	[−0.608, −0.340]	−8.639	−6.945	$< 0.001^{***}$	$< 0.001^{***}$
DAY[WD] × PA_{pc}	0.019	0.303	0.057	0.075	[−0.094, 0.131]	[0.156, 0.450]	0.327	4.034	0.744	$< 0.001^{***}$
DAY[ND] × PA_{pc}	0.510	−0.171	0.052	0.056	[0.409, 0.611]	[−0.281, −0.060]	9.890	−3.029	$< 0.001^{***}$	0.003^{*}
DAY[WD] × PA_{int}	0.446	−0.293	0.064	0.086	[0.321, 0.571]	[−0.461, −0.125]	6.944	−3.410	$< 0.001^{***}$	$< 0.001^{***}$
DAY[ND] × PA_{int}	−0.147	0.246	0.051	0.067	[−0.248, −0.046]	[0.114, 0.378]	−2.855	3.652	0.004^{**}	$< 0.001^{***}$
DAY[WD] × $PA_{pc} \times PA_{int}$ ✓	0.070	0.038	0.067	0.113	[−0.060, 0.201]	[−0.183, 0.260]	1.056	0.340	0.291	0.734
DAY[ND] × $PA_{pc} \times PA_{int}$ ✓	0.302	0.471	0.051	0.072	[0.203, 0.401]	[0.330, 0.613]	5.980	6.533	$< 0.001^{***}$	$< 0.001^{***}$
ACT[Non-Work] ✓	−0.558	−0.371	0.052	0.059	[−0.660, −0.456]	[−0.487, −0.255]	−10.765	−6.263	$< 0.001^{***}$	$< 0.001^{***}$
ACT[Driving]	−0.237	0.303	0.055	0.064	[−0.344, −0.130]	[0.178, 0.427]	−4.344	4.757	$< 0.001^{***}$	$< 0.001^{***}$
$mCDN$ ✓	0.767	0.479	0.019	0.020	[0.730, 0.803]	[0.440, 0.518]	40.845	24.124	$< 0.001^{***}$	$< 0.001^{***}$
$SLEEP$ ✓	−0.011	−0.011	0.021	0.025	[−0.053, 0.031]	[−0.060, 0.038]	−0.512	−0.442	0.609	0.659
OCC[SP]	0.349	−1.427	0.406	0.712	[−0.447, 1.145]	[−2.822, −0.032]	0.859	−2.005	0.391	0.045^{*}
$NASA_{MD}$ ✓	0.453	0.117	0.027	0.032	[0.400, 0.507]	[0.054, 0.180]	16.693	3.625	$< 0.001^{***}$	$< 0.001^{***}$
$NASA_{PD}$ ✓	0.062	0.319	0.022	0.028	[0.019, 0.105]	[0.264, 0.373]	2.828	11.392	0.005^{**}	$< 0.001^{***}$
$NASA_{P}$ ✓	0.154	0.129	0.021	0.027	[0.112, 0.196]	[0.076, 0.182]	7.235	4.781	$< 0.001^{***}$	$< 0.001^{***}$
$B5_E$	−0.247	1.036	0.212	0.348	[−0.662, 0.168]	[0.355, 1.717]	−1.167	2.980	0.243	0.003^{**}

cumulative strain [37, 38]. It also aligns with *role conflict accounts*, where dual high-demand systems—professional leadership and caregiving—compete for attention when boundaries are porous. Our debrief data echo this: on standard workdays, parents described juggling leadership with school calls, daycare coordination, and pickup logistics, whereas non-parents more often focused on work and then disconnecting.

By contrast, on critical mission days, we observed convergence between parents and non-parents. Institutional device bans, strict operational protocols, and high task tempo combined to sharply limit contact with family and other life domains. Qualitative accounts and field notes suggest that during these periods participants' attention was tightly focused on mission goals, with little opportunity for phone use or boundary negotiation. In our theoretical terms, critical days are characterized by a "crisis bundle" of attentional narrowing, protocol-driven device detachment, and collective stakes; this configuration appears to suppress parenting-related differences on top of an already elevated physiological baseline.

Taken together, our findings advance a *contextual* account of parenting stress in critical professions. Parenting young children does not simply add a constant increment of stress; rather, its impact depends on boundary regimes. When boundaries are enforced from outside (critical days), parents and non-parents look similar. When boundaries are more permeable and negotiated from within (standard workdays and non-workdays), the hidden load of caregiving becomes visible in physiological signals.

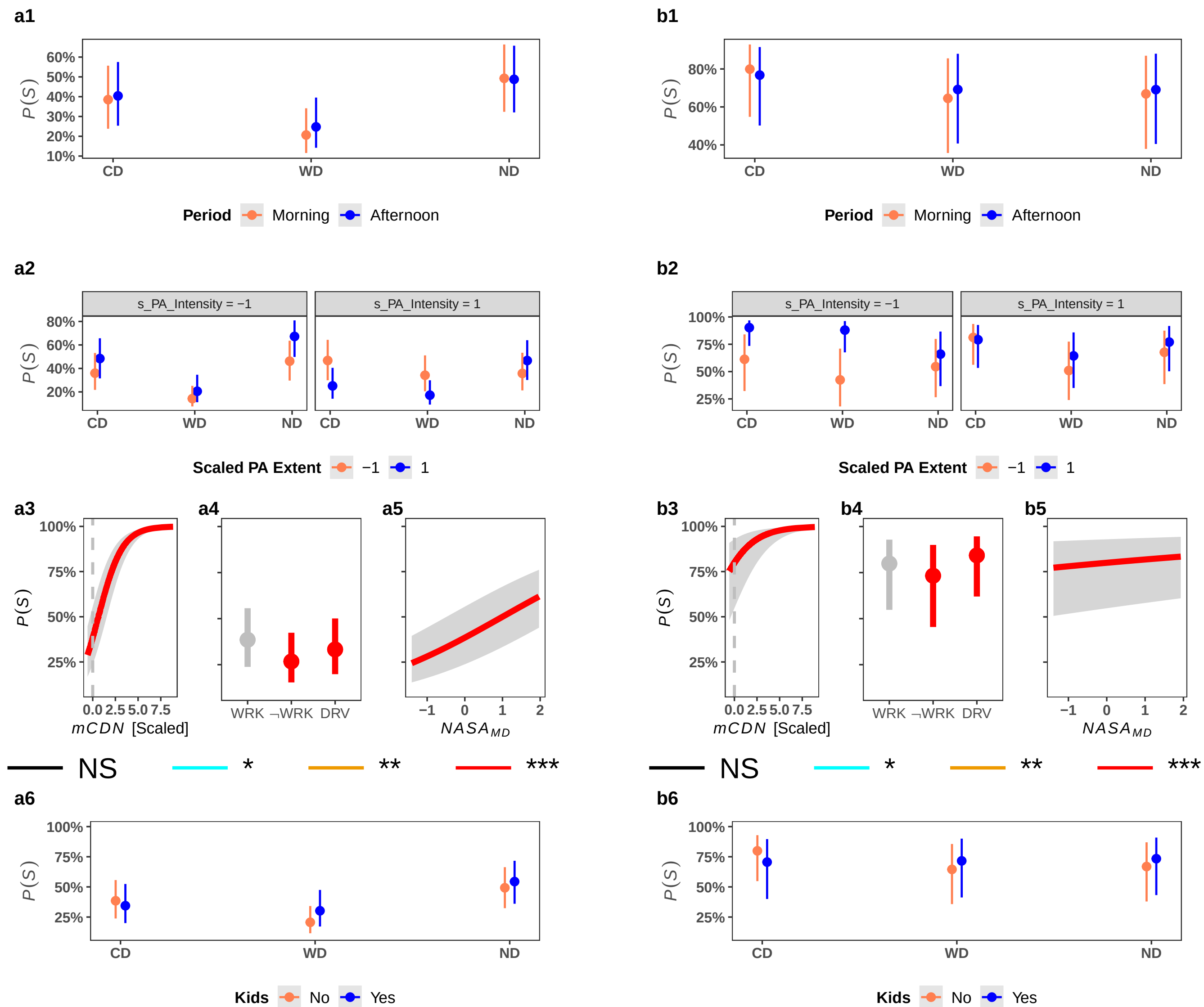

Figure 7: Interaction and main effects of the logistic regression models on stress derived from NHR_D (panels a1–a6) and SNS (panels b1–b5). Panels show only predictors that reached significance in at least one model. Stress probabilities are estimated over the full dataset of sedentary 5-min intervals. Color-coded stars indicate significance levels (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). a1–b1. Interaction effects between type of day (DAY) and period of day (PRD). a2–b2. Interaction effects between type of day (DAY), percentage of physical activity that day (PA_{pc}), and intensity of physical activity that day (PA_{int}). a3–b3. Association of micro-cadence $mCDN$ with stress probability for n = 23,702 Work + Non-Work + Driving intervals. a4–b4. Associations of sedentary activity type with stress probability, based on the following frequencies: n_W = 13,027 Work, n_{NW} = 8,011 Non-Work, and n_D = 2,664 Driving intervals. a5–b5. Association of daily $NASA_{MD}$ scores with stress probability for n = 171 days. a6–b6. Interaction effects between parenthood (KID) and type of day (DAY).

5.2 Day Type, Time of Day, and Stress Channels

Day type and time of day further shaped stress dynamics. Both stress-labeling pipelines converged on substantially lower stress probabilities on typical workdays than on critical days, reflecting the exceptional load of acute operations. Non-workdays showed more mixed patterns: in the NHR-based model, non-workdays sometimes exhibited equal or higher probabilities of high arousal than workdays, whereas in the SNS-based model, non-workdays tended to be less stressful than critical days overall but still contained frequent high-arousal episodes.

Debriefs help resolve this apparent tension. Participants often described non-workdays as physically and socially intense: long bike rides, children's birthday parties, or demanding family logistics.

Many of these high-arousal episodes were characterized as meaningful or enjoyable rather than distressing. In contrast, elevated arousal on workdays—especially in parents—was more often linked to deadlines, on-call responsibilities, or overlapping demands from work and family. Because both NHR and SNS are valence-agnostic, they capture arousal but not whether it reflects *distress* or *eustress*. Interpreting day-type differences therefore requires triangulation with qualitative accounts.

Time-of-day effects point to a similar nuance. Across channels, we saw that afternoon periods often failed to show the expected relief relative to mornings, particularly on workdays. Professional obligations and family logistics tend to accumulate toward late afternoon (e.g., coordinating shift changes while arranging pickups), which likely blunts diurnal recovery. SNS, being more sensitive to moment-to-moment HRV fluctuations, captured some of these within-day dynamics more sharply than NHR. For HCI and stress-aware system design, this suggests a division of labor: NHR supports large-scale, device-agnostic deployment and robust day-level comparisons; SNS offers finer-grained insight for more instrumentation-heavy settings.

5.3 Cultural Context and Generalizability

Our sample is situated in a specific cultural and institutional context: mid-career emergency physicians and tactical police leaders in Western European public-sector organizations. These systems are embedded in comparatively strong welfare states, with formalized employment protections, regulated working hours, and widely available childcare and schooling. They also reflect particular norms around professional autonomy, parental leave, and expectations of availability.

These features matter for how boundaries are drawn and experienced. For example, work–non-work boundaries in many Western European settings may be more formally specified than in regions where longer working hours, precarious contracts, or informal labor arrangements are common. Likewise, cultural norms around parenting—such as expectations of parental involvement, extended-family support, and gendered divisions of care—vary widely across North American, Asian, Latin American, and other contexts. In some environments, family obligations might intrude even more heavily into professional life; in others, institutional supports or extended kin networks might buffer the load that individual parents face.

Our findings should therefore be interpreted as evidence about how parenting and boundary regimes interact *within* this particular Western European configuration, not as universal statements about parents in all cultures. We therefore treat our study as a strong case from one context that can be compared with future work in other settings. Cross-cultural extensions—e.g., replicating the protocol in North American trauma centers, Latin American police organizations, or East Asian hospitals—would be especially valuable for examining how institutional policies, family structures, and cultural norms jointly shape the hidden load of parenting in critical professions.

5.4 Methodological and Theoretical Contributions

Methodologically, we make three contributions to in-the-wild stress research.

First, we propose and validate a sedentary-only, HR-based stress label (NHR) anchored in daily baselines and individualized variability. Our analysis of device-provided daily baseline HR showed no systematic differences across critical days, workdays, and non-workdays, and a mixed-effects model with DayType as a predictor did not yield significant effects. Thus, there is no evidence for systematic anticipatory tonic stress on critical days at the level of daily baselines, and the NHR transformation (deviation from a day-specific baseline) does not introduce systematic differences between day types by construction.

Second, we explicitly restrict stress labeling to sedentary intervals and control for micro-cadence. By separating physical activity from sedentary work, sedentary non-work, and driving, and by including micro-movement as a covariate, we reduce the risk of conflating exertion with mental stress. This is particularly important in naturalistic settings where short bursts of movement (e.g., pacing, hallway walking) are frequent. Our expert-driven activity labeling protocol, combining deterministic sensor rules with debrief information and targeted follow-ups, offers a reproducible alternative to continuous self-report.

Third, we assess NHR not by pointwise agreement with HRV-based labels, but by *construct-level convergence*. The full and simplified NHR models and the SNS model converge on the same core predictors—day type, parenthood-by-day interactions, micro-cadence, and mental and physical demand—even when individual episodes differ. This demonstrates that NHR is a scalable proxy with sufficient fidelity to support explanatory modeling and design decisions, while SNS can serve as a higher-fidelity benchmark when full HRV access is feasible.

Theoretically, our results refine how attentional control and boundary permeability are understood in high-stakes leadership. Rather than treating “device detachment” or “attentional narrowing” as standalone mechanisms, we conceptualize critical, standard, and non-work days as distinct *configurations* of attentional focus, institutional device regimes, and family responsibilities. Parenting amplifies stress primarily in configurations where boundaries are negotiated and permeable, but not where they are externally enforced. This configuration-based view helps reconcile why some studies find strong parenthood effects on stress while others do not: the underlying boundary regimes differ.

5.5 Design Implications

Our findings suggest several directions for the design of boundary management and stress-aware technologies in critical work settings.

Adaptive notification and availability management. Existing CHI work on notification management emphasizes user preferences and simple rules (e.g., “do not disturb” modes) [25]. Our data suggest that these rules need to be deeply context-sensitive. Systems that can infer or be informed about critical operations (e.g., from calendars, duty rosters, or manual toggles) could temporarily suppress non-urgent family notifications while preserving channels for truly urgent contact. Conversely, during routine workdays and non-workdays, tools that surface or help plan "good windows" for family coordination may reduce the need for ad-hoc interruptions and negotiation.

Bioadaptive feedback loops. Lightweight bioadaptive feedback could help individuals track when their stress exceeds day-specific baselines. For example, subtle haptic cues or ambient displays could signal prolonged elevation of NHR in sedentary contexts, prompting micro-breaks, breathing exercises, or short walks. Crucially, such feedback must respect attentional constraints in safety-critical settings; cues during acute interventions must not distract. Designs that defer feedback to safe transition points (e.g., after a code or debrief) or that aggregate feedback into reflective summaries may be more appropriate than real-time nudging during crises.

Family-facing boundary scaffolds. Boundary management is a relational task, not just an individual one. Tools that help families articulate and visualize shared expectations—such as simple "traffic light" availability indicators linked to duty schedules, or shared calendars with stress-aware annotations—could support more deliberate decisions about when to call, text, or request help. Our findings that non-work intervals and structured leisure exercise are protective suggest that technologies should not aim for constant disconnection but for planned, negotiated connection that avoids piling extra demands onto already overloaded moments.

Across all of these directions, flexibility and negotiation are key. Families differ in norms and needs, and cultural context shapes what kinds of boundary tools are acceptable. Technologies should augment, rather than replace, local practices of coordination and care.

5.6 Organizational and Policy Relevance

Beyond individual tools, our findings highlight the role of organizational policies as de facto boundary technologies. In our study, institutional rules around smartphone use sharply distinguished critical days from routine workdays. On critical days, enforced disconnection created clear, non-negotiable boundaries; on routine days, informal and uneven practices prevailed, leaving individuals and families to negotiate availability on their own.

Organizations can build on these insights in at least three ways. First, scheduling and rostering practices can be designed to reduce predictable collisions between work and caregiving demands (e.g., aligning shift end times with school schedules when possible, or clustering critical shifts to enable anticipatory planning). Second, training programs can include boundary management as a skill, alongside more traditional resilience content on sleep and physical fitness. Third, policies can recognize that caregiving responsibilities are a structural factor in occupational health, not just a private matter, and provide institutional supports such as flexible on-call arrangements or backup childcare.

Importantly, these interventions must be sensitive to local cultures and legal frameworks. What counts as a reasonable accommodation or acceptable device policy varies across jurisdictions and organizations. Our contribution is to show, physiologically, that boundary regimes are not neutral: they materially shape how the hidden load of parenting expresses itself in the bodies of frontline leaders.

5.7 Physical Activity and Stress Buffering

Physical activity played a nuanced, context-dependent role in stress dynamics. Both labeling methods indicated that movement matters, but in different ways and with different emphases.

In the NHR model, greater daily extent and intensity of physical activity were associated with reduced stress on critical days, with super-additive benefits when both were high. On non-workdays, high-intensity activity was particularly protective, consistent with debrief reports of exercise as a primary recovery strategy (e.g., long bike rides or runs). On standard workdays, the buffering effect of intensity was more modest, likely because professional demands constrained when and how activity could occur.

In the SNS model, daily activity volume alone sometimes appeared as a stressor on critical and workdays, with protective effects emerging primarily when combined with sufficient intensity. This may reflect SNS's sensitivity to HRV fluctuations during incidental movement, such as pacing or multitasking, which can mimic stress-like patterns. When movement is more structured and intense—typical of deliberate exercise—the HRV profile becomes more clearly distinguishable from mental stress.

These findings underscore that *not all movement is equal*. Incidental movement in high-demand contexts may track stress rather than relieve it, whereas purposeful exercise—especially on non-workdays—can provide robust buffering. For design, this suggests that stress-aware systems should distinguish between unstructured fidgeting or pacing and deliberate physical activity when providing feedback or recommendations.

5.8 Other Control Variables and Stress

Other variables in our models provide additional context.

Non-work intervals were reliably associated with lower stress than work intervals across both labeling methods, highlighting the restorative potential of even short pockets of non-work within a day. Driving, however, showed labeling-dependent effects: NHR suggested a buffering role, possibly reflecting transitional downtime, whereas SNS indicated elevated stress, consistent with the cognitive demands and vigilance required for driving.

Micro-cadence—a proxy for small, restless movements during sedentary periods—was a strong positive predictor of stress in all models, aligning with everyday observations of leaders fidgeting, shifting, or pacing during tense situations. Sleep did not emerge as a significant predictor in our models, which likely reflects the relatively narrow range of sleep durations across the four-day protocol and the dominance of situational factors in shaping momentary stress.

Finally, NASA Mental Demand emerged as a robust positive predictor, and NASA Physical Demand and Performance also contributed. These links between subjective workload and physiological arousal reinforce the construct validity of our labeling and modeling approach: participants' own assessments of demanding days align with what their bodies reveal.

5.9 Dataset and Code Release

Our dataset and code can be found in GitHub (UH-ACDC/Parenting-CHI-2026). To our knowledge, there are few open datasets that combine continuous physiological sensing, carefully curated activity labels, and rich contextual information from frontline leaders in critical professions, with explicit comparison of parents and non-parents.

This release serves as (i) a benchmark for stress detection methods in real-world, safety-critical settings, and (ii) a resource for HCI and CSCW researchers interested in boundary management, parenting, and occupational health. Engineers can use the data to validate new feature extraction or modeling pipelines against both NHR- and SNS-based labels, while designers can prototype and simulate stress-aware interventions (e.g., notification filters, biofeedback triggers) using empirically grounded thresholds and day-type distinctions. Because the dataset distinguishes between parents and non-parents and between critical, routine, and non-work contexts, it also supports exploration of personalization strategies and fairness considerations in stress-aware system design.

5.10 Broader Impact

Although our specific focus is on emergency physicians and tactical police leaders in Western Europe, the methodological and conceptual framework we propose extends to other forms of frontline leadership in critical domains, such as surgical teams, air-traffic control, or firefighting. These roles share features of high stakes, tightly regulated device use, and complex family responsibilities.

By demonstrating how parenting, boundaries, and workload jointly shape physiological stress across different day types, and by offering a tractable protocol for stress labeling in the wild, we provide a foundation for future work on stress-aware, boundary-sensitive technologies. At the same time, our cultural discussion emphasizes that such systems must be adapted to local norms, policies, and family structures rather than transplanted wholesale.

Taken together, our contributions—empirical, methodological, theoretical, design-oriented, and open-science—aim to support a richer understanding of the hidden load carried by parents in critical professions and to inform technologies and policies that respect both their professional commitments and their caregiving lives.

Acknowledgments

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

References

[1] Ergun Akleman, MD Tanim Hasan, and Ioannis Pavlidis. 2021. Under the spell of deadlines. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[2] Martin A Alpert, Jad Omran, and Brian P Bostick. 2016. Effects of obesity on cardiovascular hemodynamics, cardiac morphology, and ventricular function. *Current Obesity Reports* 5 (2016), 424–434.

[3] Amy F. T. Arnsten. 2009. Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience* 10, 6 (2009), 410–422. doi:10.1038/nrn2648

[4] Blake E. Ashforth, Glen E. Kreiner, and Mel Fugate. 2000. All in a day's work: Boundaries and micro role transitions. *Academy of Management Review* 25, 3 (2000), 472–491. doi:10.5465/amr.2000.3363315

[5] RM Baevskii. 2002. Analysis of heart rate variability in space medicine. *Human Physiology* 28, 2 (2002), 202–213.

[6] Holly Blake, Juliet Hassard, Jasmine Singh, and Kevin Teoh. 2024. Work-related smartphone use during off-job hours and work-life conflict: A scoping review. *PLOS Digital Health* 3, 7 (2024), e0000554. doi:10.1371/journal.pdig.0000554

[7] Josiane L Broussard, Amy C Reynolds, Christopher M Depner, Sally A Ferguson, Drew Dawson, and Kenneth P Wright. 2017. Circadian rhythms versus daily patterns in human physiology and behavior. In *Biological Timekeeping: Clocks, Rhythms and Behaviour*, Vinod Kumar (Ed.). Springer, 279–295.

[8] Raffaele Castaldo, Paolo Melillo, Marcello Bracale, Marcellino Caserta, Maria Triassi, and Leandro Pecchia. 2015. Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control* 18 (2015), 370–377. doi:10.1016/j.bspc.2015.02.012

[9] John Charles Cheek. 1988. *Plainclothes and Off-Duty Officer Survival: A Guide to Survival for Plainclothes Officers, Undercover Officers, and Off-Duty Police Officers.* Charles C Thomas, Springfield, IL.

[10] Alexandra D Crosswell and Kimberly G Lockwood. 2020. Best practices for stress measurement: How to measure psychological stress in health research. *Health Psychology Open* 7, 2 (2020), 1–12.

[11] Daantje Derks, Arnold B. Bakker, Pascale Peters, and Jessica van Wingerden. 2016. Smartphone use and work–home interference: The moderating role of social norms and employee work engagement. *Human Relations* 69, 5 (2016), 1045–1068. doi:10.1177/0018726715611234

[12] Daantje Derks, Heleen van Mierlo, and Emil P. Schmitz. 2014. A diary study on work-related smartphone use, psychological detachment and exhaustion: Examining the role of the perceived segmentation norm. *Journal of Occupational and Organizational Psychology* 87, 1 (2014), 137–155. doi:10.1111/joop.12076

[13] J. A. Easterbrook. 1959. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review* 66, 3 (1959), 183–201. doi:10.1037/h0047707

[14] Ann Catrine Eldh, Teatske T. J. van der Zijpp, Christel McMullan, Claire Hawkes, and Jo Rycroft-Malone. 2020. Using Non-Participant Observation as a Method to Understand Implementation Context in Evidence-Based Practice. *Worldviews on Evidence-Based Nursing* 17, 3 (2020), 185–192. doi:10.1111/wvn.12449

[15] Michael W. Eysenck and Nazanin Derakshan. 2011. New perspectives in Attentional Control Theory. *Personality and Individual Differences* 50, 7 (2011), 955–960. doi:10.1016/j.paid.2010.08.019

[16] Michael W. Eysenck, Nazanin Derakshan, Rita Santos, and Manuel G. Calvo. 2007. Anxiety and cognitive performance: Attentional control theory. *Emotion* 7, 2 (2007), 336–353. doi:10.1037/1528-3542.7.2.336

[17] Sergio Garbarino, Carlo Chiorri, and NICOLA Magnavita. 2014. 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* 87 (2014), 295–306.

[18] Christopher James Groombridge, Yesul Kim, Amit Maini, De Villiers Smit, and Mark Christopher Fitzgerald. 2019. Stress and decision-making in resuscitation: A systematic review. *Resuscitation* 144 (2019), 115–122. doi:10.1016/j.resuscitation.2019.09.023

[19] Steven MV Gwynne, Aoife LE Hunt, J Russell Thomas, Alexandra JL Thompson, and Lisette Séguin. 2019. The toilet paper: Bathroom dwell time observations at an airport. *Journal of Building Engineering* 24 (2019), 100751.

[20] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Human Psychology: Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Elsevier, 139–183.

[21] MD Tanim Hasan, Huda Alghamdi, Salah Taamneh, Mike Manser, Robert Wunderlich, Panagiotis Tsiamyrtzis, and Ioannis Pavlidis. 2023. Investigating Cardiovascular Activation of Young Adults in Routine Driving. *IEEE Transactions on Affective Computing* (2023).

[22] James A. J. Heathers. 2014. Everything Hertz: Methodological Issues in Short-Term Frequency-Domain HRV. *Frontiers in Physiology* 5 (2014), 177. doi:10.3389/fphys.2014.00177

[23] Raymond Hernandez, Shawn C Roll, Haomiao Jin, Stefan Schneider, and Elizabeth A Pyatak. 2022. Validation of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) adapted for the whole day repeated measures context. *Ergonomics* 65, 7 (2022), 960–975.

[24] David Hernando, Surya Roca, Jorge Sancho, Álvaro Alesanco, and Raquel Bailón. 2018. Validation of the Apple Watch for heart rate variability measurements during relax and mental stress in healthy subjects. *Sensors* 18, 8 (2018), 2619. doi:10.3390/s18082619

[25] Malte F. Jung and et al. 2020. Pragmatic Boundaries: Designing Notification Systems to Respect Boundary Preferences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–13. doi:10.1145/3313831.3376876

[26] Keyeong Kim, R. Stephen, R. Singh, MH Jafari, et al. 2020. Relationships between heart rate variability, occupational performance, and health outcomes in public safety personnel: A scoping review. *Frontiers in Public Health* 8 (2020), 583336. doi:10.3389/fpubh.2020.583336

[27] Stefanie Koldijk, Mark A. Neerincx, and Wessel Kraaij. 2016. Improving Predictive Models of Mental Workload by Incorporating Stress Recovery. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, 883–894.

[28] Kubios Oy. 2024. PNS and SNS indexes in evaluating autonomic function. https://www.kubios.com/blog/hrv-ans-function/. Accessed 20-Aug-2025.

[29] Kubios Oy. 2025. Kubios HRV Scientific User's Guide. https://www.kubios.com/downloads/HRV-Scientific-Users-Guide.pdf. Accessed 20-Aug-2025.

[30] E. L. Lehmann and J. P. Romano. 2005. *Testing Statistical Hypotheses* (3rd ed.). Springer.

[31] Feng Liu and Sally Maitlis. 2010. Non-Participant Observation. In *Encyclopedia of Case Study Research*, Albert J. Mills, Gabrielle Durepos, and Elden Wiebe (Eds.). SAGE, Los Angeles, CA, 610–612. Volume 2.

[32] Feng Liu and Sally Maitlis. 2010. Non-participant observation. In *Encyclopedia of Case Study Research*, Albert J. Mills, Gabrielle Durepos, and Elden Wiebe (Eds.). SAGE, Los Angeles, CA.

[33] Gerald Matthews. 2021. Stress states, personality and cognitive functioning: A review of research with the Dundee Stress State Questionnaire. *Personality and Individual Differences* 169 (2021), 110083.

[34] Melissa Mazmanian, Ingrid Erickson, and Wanda Orlikowski. 2013. The Autonomy Paradox: The implications of mobile email devices for knowledge professionals. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*. ACM, 123–136.

[35] R. R. McCrae and P. T. Jr. Costa. 2008. The five-factor theory of personality. In *Handbook of Personality: Theory and Research*, O. P. John, R. W. Robins, and L. A. Pervin (Eds.). The Guilford Press, New York, NY, 159––181.

[36] Matthew M. Morgan. 2022. Contradictions Between Community-Oriented Police Training and Paramilitary Police Training: Implications for Police Recruit Mental Health Response Training. *Journal of Police and Criminal Psychology* 37, 4 (2022), 876–891. doi:10.1007/s11896-022-09537-3

[37] S. Katherine Nelson, Kostadin Kushlev, and Sonja Lyubomirsky. 2014. The pains and pleasures of parenting: When, why, and how is parenthood associated with more or less well-being? *Psychological Bulletin* 140, 3 (2014), 846–895. doi:10.1037/a0035444

[38] Kei Nomaguchi and Melissa A. Milkie. 2017. Parenting stress and well-being: A decade review. *Journal of Marriage and Family* 79, 3 (2017), 705–725. doi:10.1111/jomf.12452

[39] Sandra Ohly, Alexander Schmitt, Philipp A. Freund, and Dieter Zapf. 2014. Work-related smartphone use and well-being in the evening: The role of autonomous and controlled motivation. *Journal of Personnel Psychology* 13, 4 (2014), 174–183. doi:10.1027/1866-5888/a000114

[40] Lena Schüss and Tom Gross. 2019. Ad-hoc Availability: Pervasive Digital Availability and Its Relation to Stress. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, LBW1215:1–LBW1215:6. doi:10.1145/3290607.3313079

[41] Norbert K. Semmer, Simone Grebner, and Achim Elfering. 2003. Beyond self-report: Using observational, physiological, and situation-based measures in research on occupational stress. In *Emotional and Physiological Processes and Positive Intervention Strategies*, Paul L. Perrewé and Daniel C. Ganster (Eds.). Research in Occupational Stress and Well Being, Vol. 3. Emerald Group Publishing Limited, 205–263. doi:10.1016/S1479-3555(03)03006-3

[42] Cornelia Setz, Bert Arnrich, Johannes Schumm, Robert La Marca, Gerhard Tröster, and Urs Ehlert. 2010. Discriminating Stress from Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 410–417.

[43] Fred Shaffer and Jay P. Ginsberg. 2017. An overview of heart rate variability metrics and norms. *Frontiers in Public Health* 5 (2017), 258. doi:10.3389/fpubh.2017.00258

[44] Elena Smets, Emmanuel Rios Velazquez, Giuseppina Schiavone, Imen Chakroun, Ellie D'Hondt, Walter De Raedt, Jan Cornelis, Olivier Janssens, Sofie Van Hoecke, Stephan Claes, Ilse Van Diest, and Chris Van Hoof. 2018. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *npj Digital Medicine* 1, 1 (2018), 67.

[45] Richard Spano. 2006. Observer Behavior as a Potential Source of Reactivity: Describing and Quantifying Observer Effects in a Large-Scale Observational Study of Police. *Sociological Methods & Research* 34, 4 (2006), 521–553. doi:10.1177/0049124106286331

[46] Jamie Stanley, Jonathan M. Peake, and Martin Buchheit. 2013. Cardiac Parasympathetic Reactivation Following Exercise: Implications for Training Prescription. *Sports Medicine* 43, 12 (2013), 1259–1277. doi:10.1007/s40279-013-0083-4

[47] Karin Svensberg, Bente Grøstad Kalleberg, Lene Mathiesen, Yvonne Andersson, Synne Eidhammer Rognan, and Sofia Kälvemark Sporrong. 2021. The Observer Effect in a Hospital Setting – Experiences from the Observed and the Observers. *Research in Social and Administrative Pharmacy* 17, 12 (2021), 2136–2144. doi:10.1016/j.sapharm.2021.07.011

[48] Mika P. Tarvainen, Juha-Pekka Niskanen, Jukka A. Lipponen, Perttu O. Ranta-aho, and Pasi A. Karjalainen. 2014. Kubios HRV – Heart Rate Variability Analysis Software. *Computer Methods and Programs in Biomedicine* 113, 1 (2014), 210–220. doi:10.1016/j.cmpb.2013.07.024

[49] Mika P. Tarvainen, Perttu O. Ranta-aho, and Pasi A. Karjalainen. 2002. An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering* 49, 2 (2002), 172–175. doi:10.1109/10.979357

[50] Task Force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. 1996. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 5 (1996), 1043–1065.

[51] Yonatan Vaizman, Katherine Ellis, and Gert Lanckriet. 2017. Recognizing Detailed Human Context in the Wild from Smartphones and Smartwatches. *IEEE Pervasive Computing* 16, 4 (2017), 62–74. doi:10.1109/MPRV.2017.3971131

[52] Ward van Zoonen, Anu Sivunen, Kirsimarja Blomqvist, Thomas Olsson, Anu Ropponen, Kaisa Henttonen, and Matti Vartiainen. 2020. On the importance of coping with being always available: Boundary control, communication channel preferences, and the relationship with online overload and stress. *Journal of Applied Communication Research* 48, 4 (2020), 489–509. doi:10.1080/00909882.2020.1755050

[53] Judy Wajcman, Emily Rose, Helen Brown, and Mark Bittman. 2015. Enacting temporal boundaries in work: An exploration of mobile practices. *Time & Society* 19, 2 (2015), 194–210.

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

[55] Mark Weiser and John Seely Brown. 1997. The Coming Age of Calm Technology. In *Beyond Calculation: The Next Fifty Years of Computing*, Peter J. Denning and Robert M. Metcalfe (Eds.). Springer, 75–85.