Original Paper

Alignment Between Heart Rate Variability From Fitness Trackers and Perceived Stress: Perspectives From a Large-Scale In Situ Longitudinal Study of Information Workers

Gonzalo J Martinez[1], PhD; Ted Grover[2], PhD; Stephen M Mattingly[1], PhD; Gloria Mark[2], PhD; Sidney D'Mello[3], PhD; Talayeh Aledavood[4], PhD; Fatema Akbar[2], PhD; Pablo Robles-Granda[5], PhD; Aaron Striegel[1], PhD

[1]Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States

[2]Informatics Department, University of California, Irvine, CA, United States

[3]Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO, United States

[4]Department of Computer Science, Aalto University, Espoo, Finland

[5]Thomas M Siebel Center for Computer Science, University of Illinois Urbana-Champaign, Urbana, IL, United States

Corresponding Author:
Gonzalo J Martinez, PhD
Computer Science and Engineering
University of Notre Dame
400 Main Building
Notre Dame, IN, 46556
United States
Phone: 1 (574) 631 8320
Email: gonzalo.martinez@ieee.org

Abstract

Background: Stress can have adverse effects on health and well-being. Informed by laboratory findings that heart rate variability (HRV) decreases in response to an induced stress response, recent efforts to monitor perceived stress in the wild have focused on HRV measured using wearable devices. However, it is not clear that the well-established association between perceived stress and HRV replicates in naturalistic settings without explicit stress inductions and research-grade sensors.

Objective: This study aims to quantify the strength of the associations between HRV and perceived daily stress using wearable devices in real-world settings.

Methods: In the main study, 657 participants wore a fitness tracker and completed 14,695 ecological momentary assessments (EMAs) assessing perceived stress, anxiety, positive affect, and negative affect across 8 weeks. In the follow-up study, approximately a year later, 49.8% (327/657) of the same participants wore the same fitness tracker and completed 1373 EMAs assessing perceived stress at the most stressful time of the day over a 1-week period. We used mixed-effects generalized linear models to predict EMA responses from HRV features calculated over varying time windows from 5 minutes to 24 hours.

Results: Across all time windows, the models explained an average of 1% (SD 0.5%; marginal R^2) of the variance. Models using HRV features computed from an 8 AM to 6 PM time window (namely work hours) outperformed other time windows using HRV features calculated closer to the survey response time but still explained a small amount (2.2%) of the variance. HRV features that were associated with perceived stress were the low frequency to high frequency ratio, very low frequency power, triangular index, and SD of the averages of normal-to-normal intervals. In addition, we found that although HRV was also predictive of other related measures, namely, anxiety, negative affect, and positive affect, it was a significant predictor of stress after controlling for these other constructs. In the follow-up study, calculating HRV when participants reported their most stressful time of the day was less predictive and provided a worse fit (R^2=0.022) than the work hours time window (R^2=0.032).

Conclusions: A significant but small relationship between perceived stress and HRV was found. Thus, although HRV is associated with perceived stress in laboratory settings, the strength of that association diminishes in real-life settings. HRV might be more reflective of perceived stress in the presence of specific and isolated stressors and research-grade sensing. Relying on wearable-derived HRV alone might not be sufficient to detect stress in naturalistic settings and should not be considered a proxy for perceived stress but rather a component of a complex phenomenon.

XSL·FO
RenderX

KEYWORDS

Introduction

Motivation and Overview

The World Health Organization classified stress as a 21st-century epidemic [1], as chronic stress can have adverse effects on health and well-being. Stress is the perceived imbalance in demands and resources and is experienced when a situation is appraised as personally significant and taxes or exceeds resources for coping [2]. In the short term, stress is associated with negative feelings, decreased performance and productivity, and muscular problems such as tension and headaches [3,4]. In the long term, stress can lead to significant health problems, including cardiovascular disease, impaired immunity functions, and lower overall quality of life [5,6]. Therefore, the ability to monitor stress through unobtrusive means could help improve health outcomes and well-being.

Stress measurements fall roughly into two broad categories: measuring stress directly through physiological markers such as heart rate (HR) variability (HRV) [7,8], cortisol [9], or electrodermal activity [10] and using physiological data to predict perceived stress using self-reports as ground truth [11-14]. Theories on the role of appraisal on the stress response suggest a positive relationship between perceived stress (through appraising a situation as threatening or demanding) and physiological reactions such as changes in cortisol (ie, the stress hormone), respiration, and HR [2,15-17]. Laboratory studies generally confirm this relationship (see the *Background* section). However, measuring perceived stress in daily life remains an exceedingly challenging task.

Gold standard biological measures of stress such as cortisol (a stress hormone) tend to be time consuming, expensive, and intrusive; they do not allow continuous measurement and may not align with self-reports [18,19]. Researchers have considered other physiological measures associated with the stress response such as HRV, electrodermal activity, and respiration, which can be obtained using less intrusive means such as wearable sensors [20-22]. Wearable sensors are some of the least intrusive methods of measuring physiological stress and yield continuous measures with increased frequency and finer temporal granularity than self-reports or cortisol samples. In recent years, the increased quality and battery life and the low cost of wrist-worn wearables have made it possible for studies to focus on the alignment between physiological (HRV) and self-reported measures in daily life [12,23,24], bringing to light some of the limitations of translating laboratory findings to real-world settings.

Although laboratory studies that induced stress supported an association between HRV and perceived stress (eg, using the Stroop Color-Word Interference Test and mental arithmetic problems [25-28]; also see the study by Kim et al [29] for a review that found differences in HRV in response to stress), studies in daily life settings with and without wearables have yielded mixed results. For instance, in a study of 223 male white-collar workers, Kageyama et al [30] found that daily job stressors did not correlate with short-term electrocardiogram (ECG)–derived HRV features. In contrast, in a study of 909 participants, Sin et al [31] found that ECG-derived HRV features negatively correlated with longer-term (as opposed to daily) perceived stress measured over a period of 8 days. Similarly, Hynynen et al [32] found that HRV measured in an orthostatic test (sitting up after a period of sleep) but not during night sleep was related to longer-term self-reported (global) stress over the past month. Specifically, HRV features were lower in the group with high stress than in the group with lower stress, whereas HR was higher in the group with high stress. Furthermore, in a study of 20 surgeons monitored continuously over 24 hours, Rieger et al [33] separated surgeons into groups experiencing high and low stress and found significantly higher HR and lower HRV during sleep in the group with high stress.

In real-world settings involving wearables, few studies have used HRV to predict perceived stress and have also found mixed results. Hernandez [23] collected physiological and behavioral data to predict self-reported momentary stress (high vs low) from 15 participants during 5 regular days of work.

Hernandez [23] used a support vector machine model using HRV features, achieving an average accuracy of 56%, slightly better than the 50% at baseline. Similarly, in a 4-month study of 35 participants, Muaremi et al [12] achieved a classification accuracy of 59% in a 3-level prediction task of perceived stress (low, moderate, and high), with 40% at baseline. In a simpler classification task of high versus low stress, Wu et al [24] found that HRV features yielded a classification accuracy of 78% in a study of 8 participants for 2 weeks in a data set with 59% of the samples corresponding to low stress.

These studies demonstrate that HRV associations with perceived stress obtained in situ and with wearables are less consistent than in laboratory studies. The evidence is inconclusive as to whether HRV in real-life settings could reflect *daily* or *momentary* perceived stress, as is often assumed in popular applications [8,34-37]. The greatest success comes from a few small-scale studies with simplified (eg, binarized from ordinal ratings with the removal of the more difficult middle cases) stress classification tasks. Given the recency of incorporating HRV measurement in consumer-grade wearable devices to track stress in daily life and the lack of large-scale studies addressing this issue, we report on a main study, where we collected HRV data from wrist-worn wearables, as well as self-reports for 657 participants across 9 weeks, and a follow-up with 327 (49.8%) of the same participants over 1 week approximately a year later.

We extend previous studies that predicted stress from wearable HRV data in two ways: (1) we collected HRV data in a large-scale longitudinal study in a naturalistic setting (ie, without

control over what stressors occur and when); and (2) we incorporated retrospective stress evaluations, including measures of the timing of stressful periods, to investigate whether contextual knowledge of when stress occurs could help predict perceived stress. Our studies also aimed to shed light on potential factors that could explain why self-reports of stress often do not correlate with physiological measures. Specifically, we aimed to understand the extent to which HRV predicts perceived stress in naturalistic settings. Furthermore, given that HRV is a measure of arousal, we also examined the extent to which HRV is specific to stress beyond other high-arousal affective states, including anxiety, negative affect, and positive affect.

The contributions of this study are as follows:

1. We quantified the degree of association between HRV and perceived stress in a longitudinal large-scale in situ study with information workers.
2. HRV can be calculated in many ways over many time scales (eg, 5 minutes to 24 hours). We identified low frequency (LF)/high frequency (HF) ratio, very LF (VLF), triangular index, and SD of the averages of normal-to-normal intervals (SDANN) calculated between 8 AM to 6 PM as the HRV features most strongly associated with perceived stress. Using these optimal features, we found that HRV is a predictor of perceived stress; however, the relationship is not as strong as in the laboratory, indicating that HRV is limited as a sole indicator of perceived stress, as is often used in modern applications.
3. We found that the same features that indicate stress also predict anxiety, negative affect, and positive affect. However, HRV still uniquely predicts stress after accounting for the shared variance of these related constructs with stress.
4. We describe the limitations of using HRV to measure perceived stress in situ and offer suggestions to improve perceived stress measurement.

Background

Stress is defined as the physiological response to maintain homeostasis in unexpected situations or when perceiving a threat [38-41]. The stress response is manifested in 2 systems, the autonomic nervous system (ANS)—through the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS)—and the hypothalamic-pituitary-adrenal (HPA) axis [42]. The SNS outputs epinephrine, which promotes rapid and widespread physiological changes such as increased HR [43,44], whereas the PNS generally does the opposite [40,45-47]. The HPA axis outputs cortisol, a stress hormone, which supports the SNS system by increasing available glucose by suppressing other body systems such as immune function and growth [5,48,49]. In general, SNS activity ends when a stressor ends, whereas HPA axis activity may persist for up to 90 minutes after the stressor ends [50-52]. Thus, especially over time and with chronic stressors (eg, caregivers of patients with dementia), there may be a sustained cortisol response in the absence of specific SNS activity [53-55]. Many of the chronic detrimental effects of stress, such as the increased risk of heart disease,

diabetes, and mortality, are associated with increased cortisol [5,56-58].

HRV is a measure of ANS activity and has been associated with health and physical and mental stress [25,29,59-65]. HRV measurement relies on the detection of RR intervals; that is, the time between upward deflections in an ECG. Effective clinical ECG measurements require the assistance of a trained clinician to ensure correct electrode placement. A more user-friendly version for (fitness conscious) consumers is chest straps (eg, Zephyr Bioharness [66,67]) that capture waveforms in the same manner as an ECG and do not require a clinician while still being vulnerable to improper positioning.

At the other end of the spectrum, photoplethysmography sensors approximate the measurement of RR intervals by detecting beat-to-beat intervals (BBI) evidenced by volumetric changes in the microvascular bed of tissue [68,69]. Traditionally used in wearable equipment such as fitness trackers, smartwatches, and armbands, they are easy to fit and have extended battery life, therefore allowing for continuous measurement of BBI and, in consequence, HRV. This has enabled a myriad of applications that use these sensors to measure HRV and provide a measurement of "stress" [8,34-37]. However, although HRV is associated with stress in laboratory studies, as discussed previously, HRV only measures one component of the stress response: ANS activity. Although the short duration and acute stressors may evoke a strong SNS response, chronic stressors that are characterized by increased cortisol in the absence of an SNS response may not be detected by HRV alone but could still influence self-reports of perceived stress.

The differences between SNS and HPA axis activity, their measurement, and the time courses of responses may play a role in when (or whether) a relationship is found between physiological responses and self-reported stress (eg, cortisol assessed via blood shows faster responses than cortisol measured by saliva). For instance, one study [51] induced stress and found that self-reported stress was associated with physiological stress (increased HR and cortisol) only if assessed during the stressor task. Self-reported stress before or after the stressor did not correlate with physiological stress during the same period. Other studies suggest there may be a lag between perceived and physiological stress where subjective stress responses precede cortisol (endocrine) responses [70]. Gaab et al [71] found that anticipatory but not retrospective cognitive appraisal of stress (self-report) is an important determinant of the cortisol stress response, indicating that the timing of the self-report in relation to the stressor affects whether a relationship is found between perceived and physiological stress. In contrast, Oldehinkel et al [72] found that perceived stress *before* a social stressor in the laboratory did not predict physiological responses, although changes in perceived arousal and unpleasantness were associated with changes in HR, respiratory sinus arrhythmia, and cortisol during the stressor. Furthermore, perceived stress measured *after* the stressor was inversely associated with HR during the stressor.

Regarding field studies, in a literature review on the association between salivary cortisol and self-reported stress, Hjortskov et al [18] reported a lack of sufficient evidence of an association

between self-reported mental stress and the cortisol response in field studies. The review suggested that the large diversity in study designs and stress measurements possibly obscured any potential relationship. However, these findings from previous studies on the association between perceived and physiological stress indicate a relationship that may be dependent on the temporal resolution of both measurements.

Taken together, the data suggest that HRV is a reliable measure of perceived stress during stressful tasks in the laboratory. However, reliability can be eroded in naturalistic studies for several reasons. First, ecological momentary assessments (EMAs) for stress may not occur (or be answered) during a stressor, which may reduce the accuracy of physiological signals for predicting self-reported stress. Second, HRV-based measures of stress would require a stressor that evokes an HR or HRV response rather than a chronic stressor that may influence self-reports but not HR (eg, a chronic illness). Third, self-reported stress may be reflecting memory biases or coping responses (eg, see the studies by Redelmeier and Kahneman [73] and Scheier et al [74]). Fourth, there are contradictory results for the best time to measure the physiological response of a self-reported stressor (albeit possibly because of methodological differences), coupled with the lack of precise and complete information on stressors that influence the perceived stress level themselves. Finally, HRV measured from wearable sensors might not be sufficiently reliable and might be too sensitive to noise (eg, motion artifacts), thereby obfuscating any potential relationship [70]. Given these challenges, this study sought to investigate the relationship between HRV measured through wearable sensors and perceived stress in a large sample across an extended period and in situ.

Methods

Data Collection

This data were collected as part of the larger Tesserae Project [75]. Most participants came from 4 distinct organizations (denoted by O1, O2, O3, and O4), and others from various organizations (denoted by U). Participants were enrolled both on site and remotely. The characteristics of the participants, sensing streams, and study details of the Tesserae study are described in the study by Mattingly et al [75].

Participants were enrolled between January and July of 2018 for the main study, where psychological and physiological measurements of 657 participants were collected during the first 56 days of study participation. This data were used to analyze associations between HRV and self-reported perceived stress. On the basis of the results from this study, we conducted a 1-week follow-up study with 49.8% (327/657) of the same participants in April 2019 to ascertain whether the link between HRV and perceived stress could be improved by refining the self-reporting procedure.

Demographics

Demographics were collected from a survey administered at the onset of participation (Table 1).

Table 1. Demographics summary for each study (N=657).

Variable	Main study	Follow-up study (n=327)
Gender, n (%)		
Male	391 (59.5)	211 (65.5)
Female	266 (40.5)	116 (35.5)
Organization, n (%)		
O1[a]	165 (25.1)	109 (33.3)
O2[a]	237 (36.1)	78 (23.9)
O3[a]	85 (12.9)	52 (15.9)
O4[a]	25 (3.8)	5 (1.5)
U[b]	145 (22.1)	83 (12.6)
Supervisor status , n (%)		
Nonsupervisors	370 (56.3)	206 (63)
Supervisors	285 (43.4)	121 (37)
Unknown	2 (0.3)	0 (0)
Age (years)		
Values, minimum	20	20
Values, maximum	68	68
Values, mean (SD)	35.2 (9.9)	35.9 (10.3)

[a]Distinct organization.

[b]Other organizations.

Psychological Measures

Main Study

Stress was measured using the question, "Overall, how would you rate your current level of stress?" on a 5-point Likert scale ranging from 1 (no stress at all) to 5 (a great deal of stress); The responses were distributed as follows: 5303 responses were 1s (no stress at all); 5108 responses were 2s (very little stress); 3593 responses were 3s (some stress), 573 responses were 4s (a lot of stress); and 118 were 5s (a great deal of stress). This item was validated in an unpublished study [76] (available upon request) with 991 Mechanical Turk participants (Table S10 in Multimedia Appendix 1 provides correlations with other measures). Affect was measured using the 10-item Positive and Negative Affect Short inventory [77,78]. The distribution of the responses is available in Figure 1. Anxiety was measured using a validated single-item omnibus measure of anxiety, "Please select the response that shows how anxious you feel at the moment," on a 5-point Likert scale ranging from 1 (not at all anxious) to 5 (extremely anxious) [79]. EMAs were administered once a day through Qualtrics Surveys at 8 AM, 12 PM, or 4 PM over 8 weeks. Participants were prompted to answer the EMAs through SMS text messages. The responses were distributed as follows: 7501 responses were 1s (not at all anxious); 5081 responses were 2s (a little anxious); 1659 were 3s (moderately anxious); 354 were 4s (very anxious); and 100 were 5s (extremely anxious).

Given that the variables were measured repeatedly for each participant throughout the study, we used the repeated-measures correlations [80] procedure to correlate the response variables in the main study. The correlations are shown in Table 2.

Figure 1. Distribution of positive and negative affect in the main study.
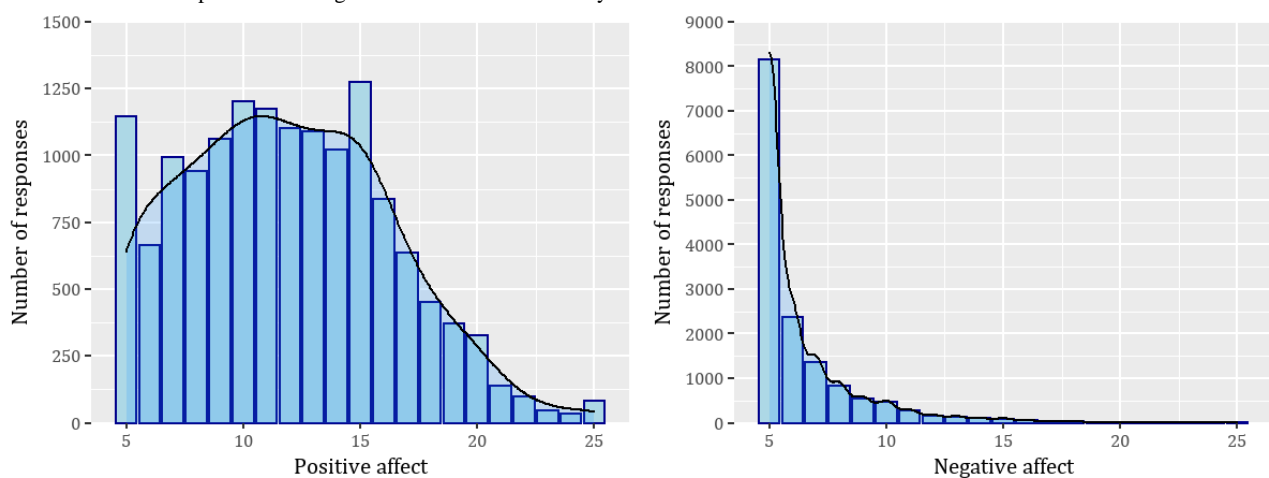
Table 2. Repeated-measures correlation between response measures in the main study and 95% CI.

Variables	Stress, r_{rm} (95% CI)	Anxiety, r_{rm} (95% CI)	Negative affect, r_{rm} (95% CI)	Positive affect, r_{rm} (95% CI)
Stress	1	0.64 (0.63 to 0.65)	0.56 (0.54 to 0.57)	−0.03 (−0.04 to −0.01)
Anxiety	0.64 (0.63to 0.65)	1	0.62 (0.61 to 0.63)	−0.02 (−0.03 to 0.00)
Negative affect	0.56 (0.54 to 0.57)	0.62 (0.61 to 0.63)	1	−0.05 (−0.07 to −0.03)
Positive affect	−0.03 (−0.04 to −0.01)	−0.02 (−0.03 to 0.00)	−0.05 (−0.07 to −0.03)	1

Follow-up Study

In the follow-up study, EMAs were sent at 4 PM every day over a week (Monday to Sunday). We collected stress by asking the same item as in the main study along with the following questions: "When did the most stressful part of your day start?"—answered by entering hours and minutes in free-form fields; "When did the most stressful part of your day end?"—also answered by entering hours and minutes in free-form fields; and "How stressful was that time?"—answered on a 5-point Likert scale ranging from 1 (no stress at all) to 5 (a great deal of stress). The responses to the stress question as stated in the main study were distributed as follows: 205 responses were 1s (no stress at all); 530 responses were 2s (very little stress); 484 responses were 3s (some stress), 22 responses were 4s (a lot of stress); and 132 were 5s (a great deal of stress).

The responses to the question "How stressful was that time?" were distributed as follows: 36 responses were 1s (no stress at all); 254 responses were 2s (very little stress); 732 responses were 3s (some stress), 71 responses were 4s (a lot of stress); and 280 were 5s (a great deal of stress).

From the timings provided by participants, we calculated the duration of the reported most stressful time of the day, as well as the length of time between the end of that moment and when the participant answered the survey. We refer to the stress question asked in the same way as in the main study, as perceived stress at the time of survey response, whereas we refer to the item introduced in the follow-up study as perceived stress at the reported most stressful time of the day. Figures 2 and 3 provide the distribution of responses, and Table 3 shows the correlation of the responses [80].

Figure 2. Distribution of the duration of perceived stress at the reported most stressful time of the day. Note that in some cases, this time overlapped with the survey response time.
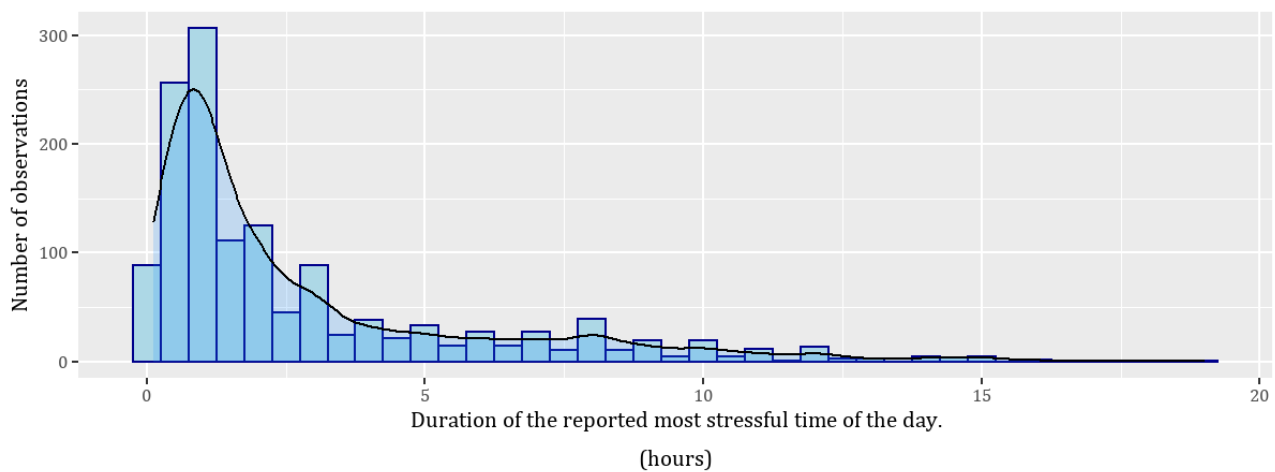

Figure 3. Distribution of the time between the reported most stressful time of the day and the survey response time. Negative values are because of when participants anticipated that the most stressful time of the day would end after the survey response time. Positive times indicate that the most stressful time of the day started and ended before the survey was answered, and negative times indicate the most stressful time of the day at least ended after the survey was answered.
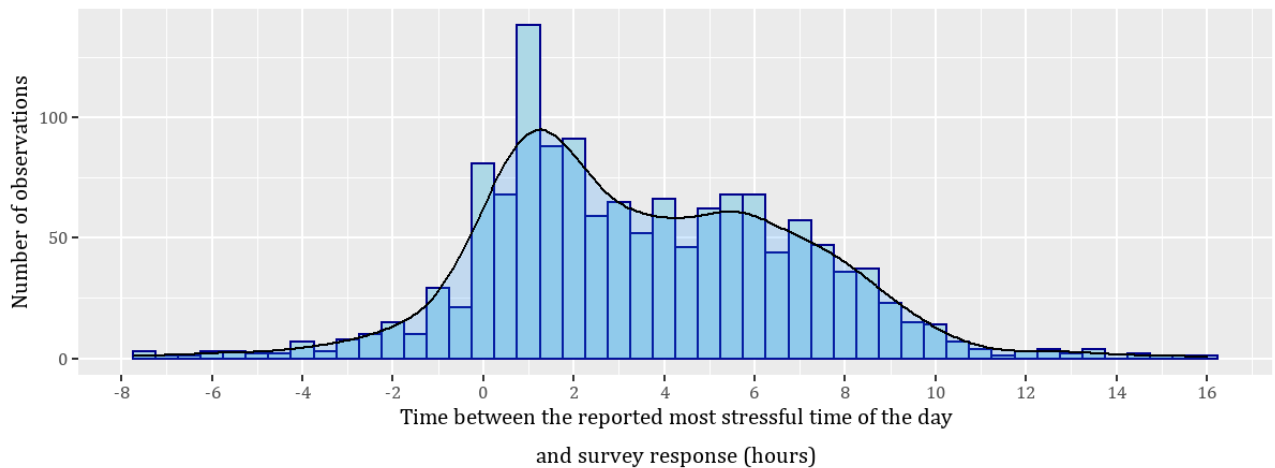
Table 3. Repeated-measures correlations of the responses in the follow-up study and 95% CI.

Measures	Perceived stress at the time of survey response, r_{rm} (95% CI)	Perceived stress at the reported most stressful time of the day, r_{rm} (95% CI)	Duration of the most stressful time, r_{rm} (95% CI)
Perceived stress at the time of survey response	1	0.5 (0.45 to 0.54)	0.33 (0.27 to 0.38)
Perceived stress at the reported most stressful time of the day	0.5 (0.45 to 0.54)	1	0.17 (0.11 to 0.22)
Duration of most stressful time	0.33 (0.27 to 0.38)	0.17 (0.11 to 0.22)	1
Time between most stressful time and survey response	−0.29 (−0.34 to −0.23)	−0.12 (−0.18 to −0.06)	−0.43 (−0.48 to −0.38)

Physiological Measures

Wearables can accurately detect HR, especially in conditions of rest or mild exercise [81], although they can have missing data [82]. To measure HR and BBI, from which HRV is computed, participants wore the Garmin vivosmart 3 fitness band (24/7) for the duration of their participation. The same sensors were used in the main study and the follow-up.

In both studies, we examined the associations between HRV and the psychological measures in our sample. To do so, we derived a series of HRV features by adopting standards for the measurement, physiological interpretation, and clinical use of HRV from the North American Society of Pacing and Electrophysiology [29]. In total, we computed 16 HRV features across different time windows using the "hrvanalysis" python library [83], each with a minimum and maximum recording time within the recommended ranges established by Shaffer

and Ginsberg [84]. Of these features, 5 were from time domain analyses, which measure variation in HR over time, or the intervals between HR cycles [29]. Triangular index was the single geometric method used [85]. A total of 7 features were from frequency domain analyses [24] where the power spectral density analysis of the HRV frequency domain provides information about how power in a signal is distributed as a function of frequency, which allows the autonomic balance to be quantified at a specific time [29]. The remaining 3 features were nonlinear HRV features, which characterize changes in HRV [86-88]. In this study, we focused on features derived from the Poincaré plot (ie, the scatter plot of successive BBIs: BBI_n vs BBI_{n+1}). Table 4 shows the mean and SD of the features across 3 different time windows.

Table 4. Mean and SD of heart rate variability features in the main study by window size.

Feature	Type	Description	Values by window size, mean (SD)		
			5-minute	8 AM to 6 PM	24-hour
Mean BBI[a]	TD[b]	The mean BBI for a period	758.1 (130.3)	755 (87.1)	797.8 (90.7)
SDNN[c]	TD	The SD of NN[d] intervals for a period	87.6 (34.2)	135.8 (37.1)	156.0 (45.3)
RMSSD[e]	TD	The square root of the mean of the squares of the successive differences between adjacent NN intervals for a period	68.7 (24.1)	71.7 (18.1)	64.7 (16.1)
PNN50[f]	TD	The number of interval differences of successive NN intervals >50 milliseconds (NN50) divided by the total number of all NN intervals	33.3 (14.4)	33.3 (10.1)	27.8 (9.2)
SDANN[g]	TD	The SD of the averages of NN intervals in all 5-minute segments of a period	N/A[h]	99.5 (32.0)	130.4 (41.4)
Triangular index	GM[i]	The number of total NN intervals/number of NN intervals in the modal bin	16.1 (5.2)	35.4 (10.4)	41.9 (13.5)
HF[j]	FD[k]	Spectral density power in the HF range	1184.9 (688.1)	1244.3 (500.4)	991.0 (404.0)
LF[l]	FD	Spectral density power in the LF range	1637.3 (1129.7)	1779.9 (799.3)	1628.1 (707.0)
LFnu[m]	FD	LF power in normalized units: LF/(total power – VLF[n]) × 100	56.8 (8.4)	58.4 (3.11)	61.8 (4.0)
HFnu[o]	FD	HF power in normalized units: HF/(total power – VLF) × 100	43.2 (8.4)	41.6 (3.11)	38.2 (4.0)
LF/HF	FD	Ratio of LF/HF	1.43 (0.7)	1.42 (0.2)	1.65 (0.3)
Total power	FD	The variance of NN intervals over the temporal segment	4224.1 (2859.9)	4517.9 (2026.3)	4035.4 (1730.7)
VLF	FD	Spectral density power in the VLF range	1401.9 (1363.6)	1493.7 (765.4)	1416.4 (670.2)
SD1	NL[p]	The SD of the Poincaré plot perpendicular to the line of identity	48.68 (17.1)	50.7 (12.8)	45.7 (11.4)
SD2	NL	The SD of the Poincaré plot along the line of identity	113.2 (47.1)	185 (51.8)	215.6 (63.8)
SD2/SD1	NL	Ratio of SD2 and SD1	2.39 (0.78)	3.7 (0.8)	4.8 (1.1)

[a]BBI: beat-to-beat intervals.

[b]TD: time domain.

[c]SDNN: SD of normal-to-normal intervals.

[d]NN: normal-to-normal.

[e]RMSSD: root mean square of successive differences.

[f]PNN50: proportion of normal-to-normal intervals that differ by >50 milliseconds.

[g]SDANN: SD of the averages of normal-to-normal intervals

[h]N/A: not applicable.

[i]GM: geometric method.

[j]HF: high frequency.

[k]FD: frequency domain.

[l]LF: low frequency.

[m]LFnu: low frequency in normalized units.

[n]VLF: very low frequency.

[o]HFnu: high frequency in normalized units.

[p]NL: nonlinear.

XSL•FO
RenderX

As HRV features have different applications but are nevertheless correlated among themselves to varying degrees [84,89], we examined previous studies to select which features to include in our modeling. We started by selecting the three time domain features and one geometric method feature recommended by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [85]: SD of normal-to-normal intervals (SDNN), root mean square of successive differences (RMSSD), SDANN, and triangular index. As RMSSD and SD1 are identical, as are SDNN and SD2, we only entered RMSSD and SDNN in the models [84]. LF power in normalized units and HF power in normalized units are identical measures that capture the same information as LF/HF; therefore, we only included LF/HF in the models to estimate the ratio between SNS and PNS activity [84,90]. HF is also strongly correlated with PNN50 and RMSSD; therefore, we did not include it in the models. Despite eliminating SD1, SD2, HF, and LF, we decided to keep the ratios as SD2/SD1 and LF/HF as they could capture additional information compared with the individual measures [84]. The correlations among the final set of features across long-term (24 hours) and short-term (5 minutes) windows are shown in Tables 5 and 6. Finally, as HRV measurements explain different phenomena depending on the time window, we decided to use variance inflation factor (VIF) feature elimination [91] to determine the set of features for each particular model and time window and address concerns with multicollinearity.

Table 5. Repeated-measures correlations among the final set of features calculated during the 24 hours of the day when participants answered the surveys and 95% CI (N=14,695 observations from 657 participants).

Measures	Correlations, r_{rm} (95% CI)								
	SDNN[a]	RMSSD[b]	MRRI[c]	PNN50[d]	SDANN[e]	Triangular index	LF[f]/HF[g]	Total power	VLF[h]
SDNN	__[i]	—	—	—	—	—	—	—	—
RMSSD	0.2 (0.19 to 0.22)	—	—	—	—	—	—	—	—
MRRI	0.27 (0.25 to 0.28)	0.49 (0.48 to 0.51)	—	—	—	—	—	—	—
PNN50	0.21 (0.19 to 0.23)	0.93 (0.93 to 0.93)	0.45 (0.43 to 0.46)	—	—	—	—	—	—
SDANN	0.82 (0.81 to 0.82)	0.02 (0 to 0.03)	0.02 (0 to 0.04)	0.04 (0.03 to 0.06)	—	—	—	—	—
Triangular index	0.67 (0.66 to 0.68)	0.26 (0.24 to 0.27)	0.29 (0.28 to 0.31)	0.3 (0.29 to 0.32)	0.5 (0.49 to 0.51)	—	—	—	—
LF/HF	0.24 (0.23 to 0.26)	−0.33 (−0.34 to −0.31)	0.33 (0.31 to 0.34)	−0.35 (−0.37 to −0.34)	0.17 (0.15 to 0.18)	0.14 (0.12 to 0.16)	—	—	—
Total power	0.36 (0.35 to 0.38)	0.84 (0.83 to 0.84)	0.62 (0.61 to 0.63)	0.84 (0.84 to 0.84)	0.13 (0.11 to 0.14)	0.37 (0.36 to 0.38)	0.01 (−0.01 to 0.03)	—	—
VLF	0.42 (0.41 to 0.44)	0.69 (0.68 to 0.7)	0.63 (0.62 to 0.64)	0.68 (0.67 to 0.69)	0.18 (0.17 to 0.2)	0.39 (0.37 to 0.4)	0.15 (0.14 to 0.17)	0.94 (0.94 to 0.95)	—
SD2/SD1	0.69 (0.68 to 0.7)	−0.5 (−0.52 to −0.49)	−0.14 (−0.16 to −0.13)	−0.48 (−0.5 to −0.47)	0.67 (0.66 to 0.68)	0.37 (0.35 to 0.38)	0.47 (0.46 to 0.48)	−0.29 (−0.3 to −0.27)	−0.14 (−0.16 to −0.13)

[a]SDNN: SD of normal-to-normal intervals.

[b]RMSSD: root mean square of successive differences.

[c]MRRI: mean RR interval.

[d]PNN50: proportion of normal-to-normal intervals that differ by >50 milliseconds.

[e]SDANN: SD of the averages of normal-to-normal intervals.

[f]LF: low frequency.

[g]HF: high frequency.

[h]VLF: very low frequency.

[i]Upper triangle of the correlation matrix was omitted for simplicity and readability.

Table 6. Repeated-measures correlations among the final set of features calculated on the 5 minutes centered on the time when participants started answering the surveys and 95% CI (N=14,695 observations from 657 participants)[a].

Measure	Correlations, r_{rm} (95% CI)						
	SDNN[b]	RMSSD[c]	MRRI[d]	PNN50[e]	LF[f]/HF[g]	Total power	VLF[h]
SDNN	__[i]	—	—	—	—	—	—
RMSSD	0.63 (0.62 to 0.64)	—	—	—	—	—	—
MRRI	0.27 (0.25 to 0.28)	0.61 (0.6 to 0.62)	—	—	—	—	—
PNN50	0.59 (0.58 to 0.6)	0.95 (0.94 to 0.95)	0.53 (0.52 to 0.54)	—	—	—	—
LF/HF	0.02 (0 to 0.03)	−0.11 (−0.13 to −0.1)	0.22 (0.2 to 0.24)	−0.15 (−0.16 to −0.13)	—	—	—
Total power	0.75 (0.74 to 0.76)	0.81 (0.8 to 0.81)	0.55 (0.54 to 0.57)	0.73 (0.72 to 0.74)	0.13 (0.11 to 0.14)	—	—
VLF	0.73 (0.72 to 0.74)	0.53 (0.52 to 0.54)	0.34 (0.32 to 0.35)	0.46 (0.45 to 0.47)	0.12 (0.1 to 0.13)	0.88 (0.88 to 0.88)	—
SD2/SD1	0.52 (0.51 to 0.53)	−0.27 (−0.29 to −0.26)	−0.28 (−0.3 to −0.27)	−0.28 (−0.29 to −0.26)	0.24 (0.22 to 0.25)	0.05 (0.03 to 0.07)	0.28 (0.27 to 0.3)

[a]SD of the averages of normal-to-normal intervals and triangular index are not included as they should not be calculated in a single 5-minute time window.

[b]SDNN: SD of normal-to-normal intervals.

[c]RMSSD: root mean square of successive differences.

[d]MRRI: mean RR interval.

[e]PNN50: proportion of normal-to-normal intervals that differ by >50 milliseconds.

[f]LF: low frequency.

[g]HF: high frequency.

[h]VLF: very low frequency.

[i]Upper triangle of the correlation matrix was omitted for simplicity and readability.

Data Exclusion

To account for missing EMA or smartwatch data during both studies (eg, dead battery or device not worn), days were excluded from the sample if any value was missing from the predictors for that day. This resulted in a final data set of 14,695 entries in the main study and 1373 in the follow-up study of matching psychological and physiological measures.

HRV Analysis

Main Study

The main purpose of this study was to examine the relationship between HRV and perceived stress as assessed by a daily stress survey. Many of the HRV features calculated are suited for short time frame measurements (eg, 2 minutes), as well as the long term (eg, 24 hours); however, Shaffer and Ginsberg [84] cautioned that these are not to be used interchangeably. Therefore, given the conflicting evidence presented in the related works as to when it is best to measure HRV in relation to a stressful event, we tested a series of models for predicting the daily stress survey response, with HRV features derived (1) 5 minutes before completing the survey, (2) 30 minutes before, (3) 5 minutes after, (4) 30 minutes after, (5) using time windows of varying length (5 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, 8 hours, and 24 hours) centered on the moment the survey

was started, (6) during the entire 24 hours on the day a participant answered the survey, and (7) during the "work day" from 8 AM to 6 PM. For sake of brevity, we report all the coefficients only for the model using the time frame with the best fit in the main results, whereas the coefficients of the models across all other time windows are reported in the form of density plots in Figure S5 in Multimedia Appendix 2. Finally, we examined the overall variance explained in the outcome measure of daily perceived stress from the HRV features.

To determine whether HRV specifically predicts stress or simply indicates arousal, which correlates with other psychological measures, we first built models to examine whether our derived HRV features predicted other survey measures that are known to have a relationship with psychological stress or arousal: positive affect, negative affect, and anxiety [92,93]. Then, to understand whether there is specificity in predicting perceived stress, we further built two models: a model predicting stress using anxiety, positive affect, and negative affect as predictors, and a second model incorporating HRV as an additional predictor.

Follow-up Study

In the analysis conducted in the follow-up study, we leveraged the additional information gained from participants related to their perceived stress duration and evaluated how well the HRV

features can predict perceived stress at the reported most stressful time of the day and again predict perceived stress at the time of survey response (the same question asked in the main study). For predicting perceived stress at the reported most stressful time of the day and perceived stress at the time of survey response, in this study, we computed the HRV features in the same manner as in the main study and used the best performing time window found earlier while also considering HRV features calculated during participants' reported most stressful periods for that day. We proceeded to compare these 2 models in predicting both perceived stress at the reported most stressful time of the day and perceived stress at the time of survey response. In addition, we considered the duration of perceived stress at the reported most stressful time of the day as an outcome measure in itself to better understand whether HRV is related to the saliency (score) of the stress events or the duration.

Modeling Strategy

As our data comprises repeated observations for each participant, and stress and anxiety are ordinal variables, we used cumulative link mixed-effects models [94] using a random intercept for the participant. We considered using random slopes in our models but decided against it because of model convergence issues in the main study and not having enough observations to support such random effects structure in the follow-up study. In the cases of predicting positive affect and the duration of perceived stress at the reported most stressful time of the day (follow-up study), we used linear mixed-effects models [95,96] as the variables can be considered continuous. In the case of negative affect, we used a negative binomial generalized linear mixed-effects model, given the distribution of the variable (Figure 1). As stated earlier, we used VIF [97] feature elimination to iteratively remove VIFs >3 to address multicollinearity [91,98]. As the predictors were on vastly different scales, all predictor variables were z score standardized before being entered into the models. Pseudo R^2 values for both marginal (fixed effects alone) and conditional (random and

fixed) effects are reported using the method described by Nakagawa and Schielzeth [99].

Ethics Approval

The study protocol was approved by the University of Notre Dame Institutional Review Board (17-5-3870).

Results

Main Study

Figure 4 provides a density plot of the variance explained (pseudo R^2) by the HRV features across all periods. On average, HRV explained a small portion (approximately 1%) of the variability in perceived stress. We also found that the model with features computed during the hours of 8 AM to 6 PM had the lowest Akaike information criterion (AIC) and explained the highest variance (Figure 4), although this was still modest (2.2%). Coefficients for this model are reported in Table 7, whereas density plots of coefficients across all time windows are included in Figure S5 (Multimedia Appendix 2).

Regarding whether HRV predicts perceived stress specifically or simply predicts arousal, we found that the directionality of most of the associations was the same for stress, anxiety, positive affect, and negative affect (Tables 7 and 8). Mean RR interval was a significant predictor of anxiety and positive affect but was not significant in predicting stress. LF to HF ratio and triangular index were both significant predictors of stress; however, LF to HF ratio was not a significant predictor of negative affect, and triangular index was not a significant predictor of positive affect.

In addition, after controlling for positive affect, negative affect, and anxiety, most HRV features were still significant predictors of perceived stress, and when compared against a model that only considers the measures of affect and anxiety, a model containing HRV provided a better fit (Table 7), as confirmed by likelihood ratio tests and AIC (χ^2_5=157.8; P<.001; AIC 23,561 vs 23,709).

Figure 4. Density plot of marginal R2 across time windows from 5 minutes to 24 hours.
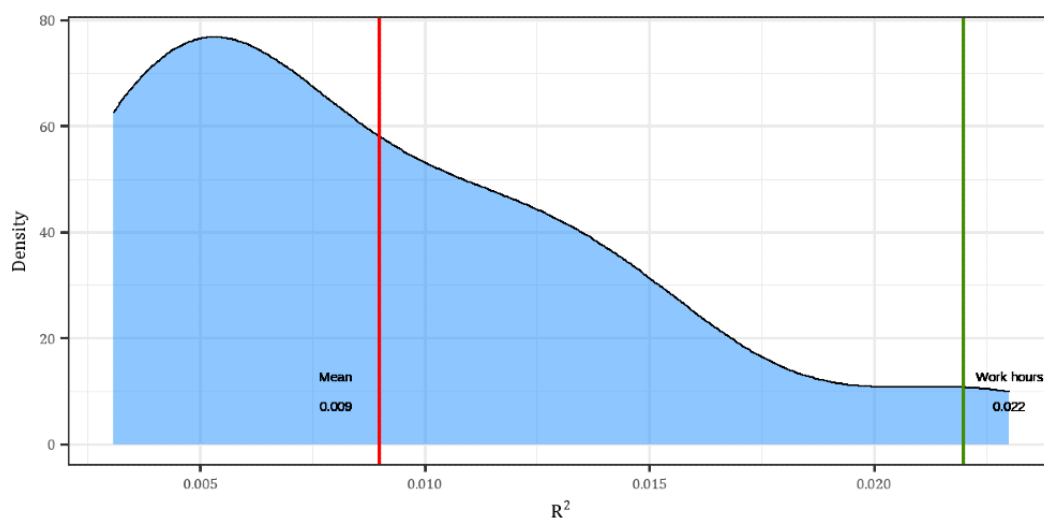
Table 7. Model for perceived stress with variance inflation factor–reduced HRV[a] features derived from beat-to-beat interval data during normal work hours of 8 AM to 6 PM[b].

Predictors	Perceived stress at the time of survey response from HRV[c]		Perceived stress at the time of survey response from anxiety, positive affect, and negative affect[d]		Perceived stress at the time of survey response from anxiety, positive affect, negative affect, and HRV[e]	
	OR[f] (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
MRRI[g]	0.95 (0.89-1.02)	.16	__[h]	—	1.01 (0.94-1.09)	.75
LF[i]/HF[j]	0.86 (0.82-0.91)	*<.001*[k]	—	—	0.85 (0.81-0.90)	*<.001* [k]
VLF[l]	1.54 (1.42-1.67)	*<.001* [k]	—	—	1.31 (1.20-1.43)	*<.001* [k]
Triangular index	0.88 (0.83-0.94)	*<.001* [k]	—	—	0.94 (0.88-1.01)	.10
SDANN[m]	0.74 (0.69-0.78)	*<.001* [k]	—	—	0.81 (0.76-0.86)	*<.001* [k]
Anxiety	—	—	5.38 (5.05-5.73)	*<.001* [k]	5.30 (4.97-5.64)	*<.001* [k]
Positive affect	—	—	0.96 (0.91-1.01)	.11	0.94 (0.89-0.99)	*.01*
Negative affect	—	—	2.52 (2.37-2.68)	*<.001* [k]	2.53 (2.38-2.69)	*<.001* [k]

[a]HRV: heart rate variability.

[b]Model fit on 14,695 observations from 657 participants. Cumulative link mixed-effects model thresholds are omitted for brevity. An extended version with threshold values is available in Table S15 in Multimedia Appendix 5.

[c]Random effects: σ^2=3.29; τ_{00}=2.25; participant intraclass correlation coefficient 0.41; marginal R^2/conditional R^2=0.022/0.420; Akaike information criterion 31,602.

[d]Random effects: σ^2=3.29; τ_{00}=1.48; participant intraclass correlation coefficient 0.31; marginal R^2/conditional R^2=0.547/0.688; Akaike information criterion 23,709.

[e]Random effects: σ^2=3.29; τ_{00}=1.52; participant intraclass correlation coefficient 0.32; marginal R^2/conditional R^2=0.548/0.691; Akaike information criterion 23,561.

[f]OR: odds ratio.

[g]MRRI: mean RR interval.

[h]The predictor was not included in this model.

[i]LF: low frequency.

[j]HF: high frequency.

[k]P values lower than .05 are highlighted in italics.

[l]VLF: very low frequency.

[m]SDANN: SD of the averages of normal-to-normal intervals.

Table 8. Model for anxiety (cumulative link mixed-effects model) and negative affect (linear mixed-effects model) with variance inflation factor–reduced heart rate variability features derived from beat-to-beat interval data during normal work hours of 8 AM to 6 PM[a].

Predictors	Positive affect[b]			Negative affect[c]			Anxiety[d]	
	Standardized β	Standardized 95% CI	P value	IRR[e]	95% CI	P value	OR[f] (95% CI)	P value
Intercept	−.01	−0.07 to 0.05	<.001	6.32	6.22 to 6.43	<.001	__[g]	—
MRRI[h]	−.15	−0.17 to −0.12	<.001	0.99	−0.06 to −0.004	.06	0.90 (0.83 to 0.97)	.004 [i]
LF[j]/HF[k]	−.08	−0.10 to −0.07	<.001	1.00	−0.03 to 0.01	.88	0.92 (0.87 to 0.97)	.002 [i]
VLF[l]	.12	0.09 to 0.15	<.001	1.04	0.06 to 0.12	<.001	1.51 (1.39 to 1.65)	<.001 [i]
Triangular index	.00	−0.02 to 0.02	.91	0.98	−0.08 to −0.03	.001	0.91 (0.85 to 0.97)	.005 [i]
SDANN[m]	−.03	−0.05 to −0.01	.002	0.98	−0.07 to −0.03	<.001	0.76 (0.72 to 0.81)	<.001 [i]

[a]Models fit on 14,695 observations from 657 participants. P values <.05 are highlighted in italics. Cumulative link mixed-effects model thresholds are omitted for brevity. An extended version with threshold values is available in Table S16 in Multimedia Appendix 5.

[b]Random effects: σ^2=9.03; τ_{00}=9.69; participant intraclass correlation coefficient 0.52; marginal R^2/conditional R^2=0.020/0.527.

[c]Random effects: σ^2=0.15; τ_{00}=0.03; participant intraclass correlation coefficient 0.19; marginal R^2/conditional R^2=0.004/0.191.

[d]Random effects: σ^2=3.29; τ_{00}=2.51; participant intraclass correlation coefficient 0.43; marginal R^2/conditional R^2=0.015/0.441.

[e]IRR: incidence rate ratio.

[f]OR: odds ratio.

[g]The predictor was not included in this model.

[h]MRRI: mean RR interval.

[i]P values lower than .05 are highlighted in italics.

[j]LF: low frequency.

[k]HF: high frequency.

[l]VLF: very low frequency.

[m]SDANN: SD of the averages of normal-to-normal intervals.

Follow-up Study

We first assessed whether using the context provided by participants to determine an HRV window to calculate the features provided a benefit over the previously found best time window of work hours of the day. Our outcome variables were perceived stress at the time of survey response and perceived stress at the reported most stressful time of the day. In the case of perceived stress at the time of survey response, the model of HRV during work hours (reported in Table 9) achieved the best fit with an R^2 of 0.032 versus 0.022 and AIC of 3465 versus 3475, therefore favoring the model with HRV features calculated during the workday, as in the main study. It also replicates findings from the main study, which found an R^2 of 0.022. Similar results are obtained when predicting perceived stress at the reported most stressful time of the day (R^2 of 0.023 vs 0.015), with the model based on HRV during work hours reported in Table 9 and a full comparison available in Tables S11 to S12 in Multimedia Appendix 3. Thus, we did not observe

benefits from computing HRV features based on self-reported most stressful time of the day compared with the entire workday. We also found that HRV during work hours was predictive of the duration of perceived stress at the reported most stressful time of the day (Table 9), although the fit was quite small.

As the duration of perceived stress at the reported most stressful time of the day was correlated with perceived stress at the time of survey response and perceived stress at the reported most stressful time of the day scores (Table 3), we conducted a post hoc analysis to investigate whether HRV could predict the saliency of the perceived stress while controlling for the effects of the duration of the event and elapsed time since it occurred—contextual features provided through self-report. Including HRV features along with contextual features provided a better fit (R^2 of 0.064 vs 0.050) over simply using the contextual features. This was further confirmed by likelihood ratio tests and AIC (χ^2_5=22.9; P<.001; AIC 3242 vs 3255; see Tables S13 and S14 in Multimedia Appendix 4 for the full models).

XSL•FO

RenderX

Table 9. Prediction of perceived stress at the time of survey response, perceived stress at the reported most stressful time of the day, and duration of perceived stress at the reported most stressful time of the day with the same predictors—heart rate variability during work hours—as in the best model in the main study[a].

Predictors	Perceived stress at the time of survey response[b]		Perceived stress at the reported most stressful time of the day[c]		Duration of perceived stress at the reported most stressful time of the day[d]		
	OR[e]	*P* value	OR	*P* value	β	95% CI	*P* value
Intercept	__[f]	—	—	—	.02	−0.06 to 0.10	.64
MRRI[g]	0.98 (0.79 to 1.23)	.89	0.86 (0.70 to 1.07)	.18	−.03	−0.12 to 0.07	.59
LF[h]/HF[i]	0.84 (0.73 to 0.98)	*.03* [j]	0.85 (0.73 to 0.99)	*.04*	−.02	−0.09 to 0.05	.57
VLF[k]	1.56 (1.22 to 1.99)	*<.001* [j]	1.54 (1.21 to 1.97)	*<.001*	.15	0.05 to 0.25	*.005*
Triangular index	0.79 (0.63 to 0.99)	*.04* [j]	0.98 (0.79 to 1.24)	.93	−.11	−0.19 to −0.03	*.03*
SDANN[l]	0.75 (0.61 to 0.91)	*.003* [j]	0.73 (0.60 to 0.89)	*.002*	−.10	−0.20 to −0.01	*.008*

[a]The models were fit with 1373 observations from 327 participants. Cumulative link mixed-effects models threshold values are omitted for brevity. An extended version with threshold values is available in Table S17 in Multimedia Appendix 5.

[b]Random effects: σ^2=3.29; τ_{00}=1.21; participant intraclass correlation coefficient 0.27; marginal R^2/conditional R^2=0.032/0.292.

[c]Random effects: σ^2=3.29; τ_{00}=0.97; participant intraclass correlation coefficient 0.23; marginal R^2/conditional R^2=0.023/0.245.

[d]Random effects: σ^2=0.60; τ_{00}=0.40; participant intraclass correlation coefficient 0.40; marginal R^2/conditional R^2=0.019/0.414.

[e]OR: odds ratio.

[f]Cumulative Link Mixed Models have multiple thresholds rather than one intercept. Therefore, no value for an intercept is included in this table.

[g]MRRI: mean RR interval.

[h]LF: low frequency.

[i]HF: high frequency.

[j]*P* values lower than .05 are highlighted in italics.

[k]VLF: very low frequency.

[l]SDANN: SD of the averages of normal-to-normal intervals.

Discussion

Principal Findings

Stress is associated with many negative outcomes [3-6], thereby making accurate measurement and management of it an important aspect of improving both physical and mental health outcomes. To this end, the ubiquitous computing and mobile health communities have turned to wearables and, more specifically, identified wearable-sensed HRV as an attractive method for passively sensing stress [12,23,24,29]. However, does the evidence support associating HRV—as measured with wearables in the wild—with stress, as perceived by the user?

We found that the best model yielded a marginal R^2 of 2.2%, which approximately corresponds to a correlation of 0.15 and a Cohen *d* of 0.30, which lies between a small (Cohen *d*=0.20) to medium (Cohen *d*=0.50) effect [100,101]. Thus, HRV was weakly, although significantly, associated with perceived stress when measured using a wearable in naturalistic settings. The size of this effect is, to some degree, expected, given that HRV only measures ANS activity and not HPA activity, thus being an incomplete assessment of stress, even in ideal conditions. That said, we would have expected a stronger relationship between perceived stress and HRV a priori, given its popular use in assessing stress [8,34-37]. Nevertheless, despite the small magnitude of the effect, we also found some evidence for incremental prediction in that HRV uniquely predicted perceived

stress above and beyond self-reported positive affect, negative affect, and anxiety (Table 7).

We do not believe the small effect size is because of how perceived stress was assessed, as using validated assessments of related constructs, such as negative affect and anxiety, yielded similar results (Table 8) and was highly correlated with stress (Table 2). Our findings suggest that the signal provided by wearable-measured HRV is of limited use in predicting perceived stress in the wild in the absence of clear and isolated stressors (such as those provided in laboratory studies).

Regarding the optimal temporal association between HRV and perceived stress, we found that HRV features measured around the time of the survey response—when participants were assessing their current stress level—yielded a lower fit than a generic time window covering the workday (ie, between 8 AM to 6 PM). This is different from the results in laboratory settings, which suggest the optimal time window to be shorter and closer to the assessment of stress, given the quick SNS response to induced stress. Although the length of the time window in which HRV is measured can affect what contributes to the changes in the HRV features (eg, circadian rhythms might be captured with longer-term HRV but not short term [84]), the estimates found within the "workday" time window of 8 AM to 6 PM were generally consistent in directionality with previous literature for changes in HRV because of stress.

Specifically, triangular index and SDANN were both negatively associated with perceived stress. Both of these match the expectation that lower HRV would indicate higher stress [29]. VLF was positively associated with perceived stress, which is to be expected as SNS activity because of stress (among other reasons) modulates the amplitude and frequency of HRV measured in this band [84,102]. Finally, the ratio of LF to HF was negatively associated with perceived stress in the work hours time window, which might be considered counterintuitive. In controlled conditions, LF/HF can be used as a measure of autonomic balance; that is, it is assumed that PNS and SNS activity contributes to LF, and PNS largely contributes to HF [84]. Therefore, one could have expected a higher LF/HF ratio to equate to higher perceived stress, as it would indicate more SNS than PNS activity. Nevertheless, as highlighted in the study by Shaffer and Ginsberg [84], because of the complex relationship between SNS and PNS activity, LF/HF ratio will not always index autonomic balance. Thus, it is possible that in the conditions of this study, either a higher LF/HF was an indicator of higher PNS activity over SNS activity, or a higher PNS activity was a better marker for the saliency of a previous stressful event from which the participant was recovering at the time of the survey response.

In the follow-up study, our modified stress survey aimed to identify and compute HRV based on participants' most stressful time of the day. Although this is impractical for a real-world use case, it does allow measurement of HRV closer to the stressor, as in many laboratory studies. Nevertheless, measuring HRV during the most stressful time of the day yielded a lower model fit than using the generic 8 AM to 6 PM time window (Multimedia Appendix 3). Therefore, we believe the small effect of HRV as a predictor of stress ostensibly resides in the conditions of measurement themselves. Specifically, in laboratory-based studies, the measurements of changes in HRV because of stress occur in the presence of clear and isolated stressors (eg, stress being induced by the study conditions, causing an increase in SNS activity), which, in turn, implies that HRV changes because stress, and these changes can often cease with the end of a stressor [51]. Discrete and isolated stressors in controlled laboratory studies may not be as common in naturalistic settings, making results from these studies under controlled conditions not fully applicable to daily life settings.

In naturalistic settings, identifying perceived stress at the precise moment of a clear and isolated stressor would be difficult to achieve from HRV alone for several reasons. First, physiological stress is different from perceived stress. For instance, physical exertion or exercise is generally classified as a physiological stressor (and would exhibit increases in HR, decreased HRV, and increases in cortisol); however, it is well known that exercise can reduce perceived stress [103] and generally would not be reported as *stressful* by participants. Second, self-reports are subject to emotional perception and expression biases [104-107], as well as memory biases and/or coping responses [73,74]. Finally, EMAs are designed to measure stress at either random or specific times, although participants may not respond at the designated time (eg, at the end of a stressor as opposed to the middle of a stressor).

In summary, our main conclusion is that the reported association between HRV and perceived stress may depend on laboratory conditions. In naturalistic studies, there are no clear and direct links between isolated stressors and SNS responses. Although there is still an observable association between wearables and perceived stress, it is weak, and it suggests that HRV alone should not be considered a valid proxy measure of perceived stress in naturalistic studies.

Implications of This Study

Although HRV has been shown to be a useful biomarker of perceived stress in laboratory studies, we have shown that in the wild, perceived stress does not always align strongly with physiological stress. This is of special importance as an increasing number of studies and commercial applications in the ubiquitous computing community use wearables to measure stress using HRV, sometimes under the assumption that there is a very strong alignment between the two, when, in fact, the alignment is more tenuous. Although it is beneficial to have wearables capable of providing continuous measurement of HRV unobtrusively, we caution against the use of HRV features as sole or main indicators of "stress" in user-facing applications, as the results may not align with perceived stress. This level of inaccuracy risks an increase of distrust in health and well-being applications at a minimum. It can have more profound negative effects as well, and based on the present findings, labeling HRV as "stress" without proper validity data would be highly suspect. Therefore, we would encourage future work in the scientific community to investigate complementary sensing streams that could serve as markers of stress and use those in conjunction with HRV.

To realize the goal of monitoring the health of individuals, such sensing streams should be rigorously vetted through longitudinal studies to appropriately measure their predictive power for capturing intraindividual differences over time. Nevertheless, it is unlikely that any single physiological sensing stream would be able to perfectly align with perceived stress. Therefore, rather than looking at a single biomarker of the ANS, as is HRV, a more complete view of the ANS response could perhaps delineate a viable strategy for health monitoring unobtrusively in the wild. More broadly, approaches based on multimodality are more likely to yield successful outcomes in health monitoring, as recent studies show in other fields such as sleep monitoring [108], job performance monitoring [109,110], and personality prediction [111].

Limitations

It is important to note that this study has limitations. First, our sample comprised information workers who might be less likely to have movement artifacts that could affect the wearable measurements of HRV. Second, our sample was fairly homogenous, with participants whose income and education levels were above the US average (low-income and lower education populations were underrepresented). Third, we are unable to determine the accuracy of self-reported stress durations and timing of stress. Similarly, the duration of the most stressful time of the day was correlated with the perceived stress at that time, and it is possible that participants' response to one question influenced the answer to the other (ie, judging stressors that last

longer as more intense). Finally, the items introduced in the follow-up study were not validated in this or other studies. Addressing these limitations is a goal for future work.

Conclusions

We examined the alignment of physiological stress (HRV), as measured with a consumer-grade wearable device, and perceived stress in an 8-week study with information workers from multiple organizations across the United States. We found a weak but significant association between HRV and perceived stress, which was replicated in a week-long follow-up study a year later. Computing HRV across the workday outperformed other time windows, including self-reported stressful events. Overall, our findings suggest that wearable-based HRV should not be used as a sole biomarker for perceived stress in naturalistic settings. Instead, it might best be used in conjunction with other measures to measure this complex phenomenon in the wild.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Correlation of the stress item with other validated measures.
[DOCX File , 59 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Summary of odds ratios of predictors across different time windows.
[DOCX File , 120 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Comparison between models, including heart rate variability calculated during the workday and during the reported most stressful time of the day.
[DOCX File , 18 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Additive value of heart rate variability features over considering only the participants' subjective assessment.
[DOCX File , 18 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Expanded versions of the tables included in the main paper, including the threshold values for the ordinal models.
[DOCX File , 23 KB-Multimedia Appendix 5]

References

1. Murray CJ, Lopez AD, World Health Organization, World Bank, Harvard School of Public Health. The Global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020 : summary / edited by Christopher J. L. Murray, Alan D. Lopez. World Health Organization. 1996. URL: https://apps.who.int/iris/bitstream/handle/10665/41864/0965546608_eng.pdf?sequence=1&isAllowed=y [accessed 2020-11-30]
2. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York, NY, USA: Springer Publishing Company; 1984.
3. Herbert J. Fortnightly review. Stress, the brain, and mental illness. BMJ 1997 Aug 30;315(7107):530-535 [FREE Full text] [doi: 10.1136/bmj.315.7107.530] [Medline: 9329310]
4. Colligan TW, Higgins EM. Workplace stress. J Workplace Behav Health 2006 Jul 25;21(2):89-97. [doi: 10.1300/j490v21n02_07]

5. Lagraauw HM, Kuiper J, Bot I. Acute and chronic psychological stress as risk factors for cardiovascular disease: insights gained from epidemiological, clinical and experimental studies. Brain Behav Immun 2015 Nov;50:18-30. [doi: 10.1016/j.bbi.2015.08.007] [Medline: 26256574]

6. Krantz DS, Kop WJ, Santiago HT, Gottdiener JS. Mental stress as a trigger of myocardial ischemia and infarction. Cardiol Clin 1996 May;14(2):271-287. [Medline: 8724559]

7. Dishman RK, Nakamura Y, Garcia ME, Thompson RW, Dunn AL, Blair SN. Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. Int J Psychophysiol 2000 Aug;37(2):121-133. [doi: 10.1016/s0167-8760(00)00085-4] [Medline: 10831999]

8. Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability – Firstbeat White Paper. Firstbeat. URL: https://www.firstbeat.com/en/stress-recovery-analysis-method-based-24-hour-heart-rate-variability-firstbeat-white-paper-2/ [accessed 2020-10-13]

9. Kirschbaum C, Hellhammer DH. Salivary cortisol in psychobiological research: an overview. Neuropsychobiology 1989;22(3):150-169. [doi: 10.1159/000118611] [Medline: 2485862]

10. Setz C, Arnrich B, Schumm J, La Marca R, Troster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. IEEE Trans Inform Technol Biomed 2010 Mar;14(2):410-417. [doi: 10.1109/titb.2009.2036164]

11. Koldijk S, Sappelli M, Neerincx M, Kraaij W. Unobtrusive Monitoring of Knowledge Workers for Stress Self-regulation. In: Proceedings of the 21th International Conference on the User Modeling, Adaption, and Personalization. Berlin, Germany: Springer; 2013 Presented at: UMAP '13; June 10-14, 2013; Rome, Italy p. 335-337. [doi: 10.1007/978-3-642-38844-6_31]

12. Muaremi A, Arnrich B, Tröster G. Towards measuring stress with smartphones and wearable devices during workday and sleep. Bionanoscience 2013;3(2):172-183 [FREE Full text] [doi: 10.1007/s12668-013-0089-2] [Medline: 25530929]

13. Sano A, Picard RW. Stress recognition using wearable sensors and mobile phones. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Presented at: ACII '13; September 2-5, 2013; Geneva, Switzerland p. 671-676. [doi: 10.1109/acii.2013.117]

14. Hernandez J, Morris RR, Picard RW. Call center stress recognition with person-specific models. In: Proceedings of the 4th International Conference on the Affective Computing and Intelligent Interaction. Berlin, Germany: Springer; 2011 Presented at: ACII '11; October 9-12, 2011; Memphis, TN, USA p. 125-134. [doi: 10.1007/978-3-642-24600-5_16]

15. Smith CA. Dimensions of appraisal and physiological response in emotion. J Pers Soc Psychol 1989 Mar;56(3):339-353. [doi: 10.1037//0022-3514.56.3.339] [Medline: 2926633]

16. Tomaka J, Blascovich J, Kibler J, Ernst JM. Cognitive and physiological antecedents of threat and challenge appraisal. J Pers Soc Psychol 1997 Jul;73(1):63-72. [doi: 10.1037//0022-3514.73.1.63] [Medline: 9216079]

17. Neufeld RW. Evidence of stress as a function of experimentally altered appraisal of stimulus aversiveness and coping adequacy. J Pers Soc Psychol 1976 May;33(5):632-646. [doi: 10.1037//0022-3514.33.5.632] [Medline: 1271228]

18. Hjortskov N, Garde AH, Ørbæk P, Hansen ÅM. Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. Stress Health 2004 Apr 07;20(2):91-98. [doi: 10.1002/smi.1000]

19. Kalman BA, Grahn RE. Measuring salivary cortisol in the behavioral neuroscience laboratory. J Undergrad Neurosci Educ 2004;2(2):A41-A49 [FREE Full text] [Medline: 23493518]

20. Garbarino M, Lai M, Bender D, Picard RW, Tognetti S. Empatica E3 — A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In: Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies. 2014 Presented at: MOBIHEALTH '14; November 3-5, 2014; Athens, Greece p. 39-42. [doi: 10.4108/icst.mobihealth.2014.257418]

21. Haescher M, Matthies DJ, Trimpop J, Urban B. SeismoTracker: upgrade any smart wearable to enable a sensing of heart rate, respiration rate, and microvibrations. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery; 2016 Presented at: CHI EA '16; May 7-12, 2016; San Jose, CA, USA p. 2209-2216. [doi: 10.1145/2851581.2892279]

22. Aly H, Youssef M. Zephyr: ubiquitous accurate multi-sensor fusion-based respiratory rate estimation using smartphones. In: The 35th Annual IEEE International Conference on Computer Communications. 2016 Presented at: IEEE INFOCOM '16; April 10-14, 2016; San Francisco, CA, USA p. 1-9. [doi: 10.1109/infocom.2016.7524401]

23. Hernandez Rivera J. Towards wearable stress measurement. Massachusetts Institute of Technology. 2015. URL: https://dspace.mit.edu/handle/1721.1/101849 [accessed 2020-11-17]

24. Wu M, Cao H, Nguyen HL, Surmacz K, Hargrove C. Modeling perceived stress via HRV and accelerometer sensor streams. Annu Int Conf IEEE Eng Med Biol Soc 2015 Aug;2015:1625-1628. [doi: 10.1109/EMBC.2015.7318686] [Medline: 26736586]

25. Taelman J, Vandeput S, Spaepen A, Van Huffel S. Influence of mental stress on heart rate and heart rate variability. In: Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering. 2009 Presented at: IFMBE '08; November 23-27, 2008; Antwerp, Belgium p. 1366-1369. [doi: 10.1007/978-3-540-89208-3_324]

26. Salahuddin L, Cho J, Jeong MG, Kim D. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. Annu Int Conf IEEE Eng Med Biol Soc 2007;2007:4656-4659. [doi: 10.1109/IEMBS.2007.4353378] [Medline: 18003044]

27.   Sun FT, Kuo C, Cheng HT, Buthpitiya S, Collins P, Griss M. Activity-aware mental stress detection using physiological sensors. In: Proceedings of the 2nd International ICST Conference on the Mobile Computing, Applications, and Services. 2012 Presented at: MobiCASE '10; October 25-28, 2010; Santa Clara, CA, USA p. 211-230. [doi: 10.1007/978-3-642-29336-8_12]

28.   Brugnera A, Zarbo C, Tarvainen MP, Marchettini P, Adorni R, Compare A. Heart rate variability during acute psychosocial stress: a randomized cross-over trial of verbal and non-verbal laboratory stressors. Int J Psychophysiol 2018 May;127:17-25. [doi: 10.1016/j.ijpsycho.2018.02.016] [Medline: 29501671]

29.   Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]

30.   Kageyama T, Nishikido N, Kobayashi T, Kurokawa Y, Kaneko T, Kabuto M. Self-reported sleep quality, job stress, and daytime autonomic activities assessed in terms of short-term heart rate variability among male white-collar workers. Ind Health 1998 Jul;36(3):263-272 [FREE Full text] [doi: 10.2486/indhealth.36.263] [Medline: 9701906]

31.   Sin NL, Sloan RP, McKinley PS, Almeida DM. Linking daily stress processes and laboratory-based heart rate variability in a national sample of midlife and older adults. Psychosom Med 2016 Jun;78(5):573-582 [FREE Full text] [doi: 10.1097/PSY.0000000000000306] [Medline: 26867082]

32.   Hynynen E, Konttinen N, Kinnunen U, Kyröläinen H, Rusko H. The incidence of stress symptoms and heart rate variability during sleep and orthostatic test. Eur J Appl Physiol 2011 May;111(5):733-741. [doi: 10.1007/s00421-010-1698-x] [Medline: 20972879]

33.   Rieger A, Stoll R, Kreuzfeld S, Behrens K, Weippert M. Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress. Int Arch Occup Environ Health 2014 Feb;87(2):165-174. [doi: 10.1007/s00420-013-0847-z] [Medline: 23370764]

34.   Hantono BS, Nugroho LE, Santosa PI. Mental stress detection via heart rate variability using machine learning. Int J Electr Eng Inform 2020 Sep 30;12(3):431-444. [doi: 10.15676/ijeei.2020.12.3.3]

35.   Van Deusen M. Heart Rate Variability | The Ultimate Guide to HRV. WHOOP. URL: https://www.whoop.com/thelocker/heart-rate-variability-hrv/ [accessed 2020-11-17]

36.   What is Heart Rate Variability (HRV)? And how it can enhance your training. Myithlete. URL: https://www.myithlete.com/what-is-hrv/ [accessed 2020-11-17]

37.   Menon V. Reducing Stress With HRV for a happier, healthier life. Elite HRV. 2019 Apr 18. URL: https://elitehrv.com/reducing-stress-with-heart-rate-variability [accessed 2020-11-17]

38.   de Kloet ER, Joëls M, Holsboer F. Stress and the brain: from adaptation to disease. Nat Rev Neurosci 2005 Jun;6(6):463-475. [doi: 10.1038/nrn1683] [Medline: 15891777]

39.   Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: implications for the field of brain and cognition. Brain Cogn 2007 Dec;65(3):209-237. [doi: 10.1016/j.bandc.2007.02.007] [Medline: 17466428]

40.   McEwen BS. Physiology and neurobiology of stress and adaptation: central role of the brain. Physiol Rev 2007 Jul;87(3):873-904 [FREE Full text] [doi: 10.1152/physrev.00041.2006] [Medline: 17615391]

41.   Roozendaal B, McEwen BS, Chattarji S. Stress, memory and the amygdala. Nat Rev Neurosci 2009 Jun;10(6):423-433. [doi: 10.1038/nrn2651] [Medline: 19469026]

42.   Tsigos C, Chrousos GP. Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. J Psychosom Res 2002 Oct;53(4):865-871. [doi: 10.1016/s0022-3999(02)00429-4] [Medline: 12377295]

43.   Berridge CW, Waterhouse BD. The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. Brain Res Brain Res Rev 2003 Apr;42(1):33-84. [doi: 10.1016/s0165-0173(03)00143-7] [Medline: 12668290]

44.   Morilak DA, Barrera G, Echevarria DJ, Garcia AS, Hernandez A, Ma S, et al. Role of brain norepinephrine in the behavioral response to stress. Prog Neuropsychopharmacol Biol Psychiatry 2005 Dec;29(8):1214-1224. [doi: 10.1016/j.pnpbp.2005.08.007] [Medline: 16226365]

45.   Brindle RC, Ginty AT, Phillips AC, Carroll D. A tale of two mechanisms: a meta-analytic approach toward understanding the autonomic basis of cardiovascular reactivity to acute psychological stress. Psychophysiology 2014 Oct;51(10):964-976. [doi: 10.1111/psyp.12248] [Medline: 24924500]

46.   Gunnar M, Quevedo K. The neurobiology of stress and development. Annu Rev Psychol 2007;58:145-173. [doi: 10.1146/annurev.psych.58.110405.085605] [Medline: 16903808]

47.   Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. Nat Rev Neurosci 2009 Jun;10(6):434-445. [doi: 10.1038/nrn2639] [Medline: 19401723]

48.   Kaltsas GA, Chrousos GP. The neuroendocrinology of stress. In: Cacioppo JT, Tassinary LG, Berntson G, editors. Handbook of Psychophysiology. 3rd edition. Cambridge, UK: Cambridge University Press; 2007:303-318.

49.   Sternberg EM. Neuroendocrine regulation of autoimmune/inflammatory disease. J Endocrinol 2001 Jun;169(3):429-435. [doi: 10.1677/joe.0.1690429] [Medline: 11375112]

50.   Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004 May;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]

XSL•FO
RenderX

51. Hellhammer J, Schubert M. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. Psychoneuroendocrinology 2012 Jan;37(1):119-124. [doi: 10.1016/j.psyneuen.2011.05.012] [Medline: 21689890]

52. Kirschbaum C, Pirke KM, Hellhammer DH. The 'Trier Social Stress Test'--a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 1993;28(1-2):76-81. [doi: 10.1159/000119004] [Medline: 8255414]

53. Atienza AA, Henderson PC, Wilcox S, King AC. Gender differences in cardiovascular response to dementia caregiving. Gerontologist 2001 Aug;41(4):490-498. [doi: 10.1093/geront/41.4.490] [Medline: 11490047]

54. Cooper C, Balamurali TB, Livingston G. A systematic review of the prevalence and covariates of anxiety in caregivers of people with dementia. Int Psychogeriatr 2007 Apr;19(2):175-195. [doi: 10.1017/S1041610206004297] [Medline: 17005068]

55. Stalder T, Tietze A, Steudte S, Alexander N, Dettenborn L, Kirschbaum C. Elevated hair cortisol levels in chronically stressed dementia caregivers. Psychoneuroendocrinology 2014 Sep;47:26-30. [doi: 10.1016/j.psyneuen.2014.04.021] [Medline: 25001953]

56. Chiodini I, Adda G, Scillitani A, Coletti F, Morelli V, Di Lembo S, et al. Cortisol secretion in patients with type 2 diabetes: relationship with chronic complications. Diabetes Care 2007 Jan;30(1):83-88. [doi: 10.2337/dc06-1267] [Medline: 17192338]

57. Dekker MJ, Koper JW, van Aken MO, Pols HA, Hofman A, de Jong FH, et al. Salivary cortisol is related to atherosclerosis of carotid arteries. J Clin Endocrinol Metab 2008 Oct;93(10):3741-3747. [doi: 10.1210/jc.2008-0496] [Medline: 18682510]

58. Vogelzangs N, Beekman AT, Milaneschi Y, Bandinelli S, Ferrucci L, Penninx BW. Urinary cortisol and six-year risk of all-cause and cardiovascular mortality. J Clin Endocrinol Metab 2010 Nov;95(11):4959-4964 [FREE Full text] [doi: 10.1210/jc.2010-0192] [Medline: 20739384]

59. Beauchaine TP, Thayer JF. Heart rate variability as a transdiagnostic biomarker of psychopathology. Int J Psychophysiol 2015 Nov;98(2 Pt 2):338-350. [doi: 10.1016/j.ijpsycho.2015.08.004] [Medline: 26272488]

60. Castaldo R, Melillo P, Bracale U, Caserta M, Triassi M, Pecchia L. Acute mental stress assessment via short term HRV analysis in healthy adults: a systematic review with meta-analysis. Biomed Signal Process Control 2015 Apr;18:370-377. [doi: 10.1016/j.bspc.2015.02.012]

61. Gouin JP, Wenzel K, Boucetta S, O'Byrne J, Salimi A, Dang-Vu TT. High-frequency heart rate variability during worry predicts stress-related increases in sleep disturbances. Sleep Med 2015 May;16(5):659-664. [doi: 10.1016/j.sleep.2015.02.001] [Medline: 25819418]

62. Holzman JB, Bridgett DJ. Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: a meta-analytic review. Neurosci Biobehav Rev 2017 Mar;74(Pt A):233-255. [doi: 10.1016/j.neubiorev.2016.12.032] [Medline: 28057463]

63. Järvelin-Pasanen S, Sinikallio S, Tarvainen MP. Heart rate variability and occupational stress-systematic review. Ind Health 2018 Nov 21;56(6):500-511 [FREE Full text] [doi: 10.2486/indhealth.2017-0190] [Medline: 29910218]

64. Nolan RP, Jong P, Barry-Bianchi SM, Tanaka TH, Floras JS. Effects of drug, biobehavioral and exercise therapies on heart rate variability in coronary artery disease: a systematic review. Eur J Cardiovasc Prev Rehabil 2008 Aug;15(4):386-396. [doi: 10.1097/HJR.0b013e3283030a97] [Medline: 18677161]

65. Pradhapan P, Tarvainen MP, Nieminen T, Lehtinen R, Nikus K, Lehtimäki T, et al. Effect of heart rate correction on pre- and post-exercise heart rate variability to predict risk of mortality-an experimental study on the FINCAVAS cohort. Front Physiol 2014 Jun 3;5:208 [FREE Full text] [doi: 10.3389/fphys.2014.00208] [Medline: 24917825]

66. Johnstone JA, Ford PA, Hughes G, Watson T, Garrett AT. Bioharness(™) multivariable monitoring device: part. I: validity. J Sports Sci Med 2012 Sep 1;11(3):400-408 [FREE Full text] [Medline: 24149346]

67. Johnstone JA, Ford PA, Hughes G, Watson T, Garrett AT. Bioharness(™) multivariable monitoring device: part. II: reliability. J Sports Sci Med 2012 Sep 1;11(3):409-417 [FREE Full text] [Medline: 24149347]

68. Pinheiro N, Couceiro R, Henriques J, Muehlsteff J, Quintal I, Goncalves L, et al. Can PPG be used for HRV analysis? Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:2945-2949. [doi: 10.1109/EMBC.2016.7591347] [Medline: 28268930]

69. Weiler DT, Villajuan SO, Edkins L, Cleary S, Saleem JJ. Wearable heart rate monitor technology accuracy in research: a comparative study between PPG and ECG technology. Proc Hum Factors Ergon Soc Annu Meet 2017 Oct 20;61(1):1292-1296. [doi: 10.1177/1541931213601804]

70. Schlotz W, Kumsta R, Layes I, Entringer S, Jones A, Wüst S. Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: a question of timing. Psychosom Med 2008 Sep;70(7):787-796. [doi: 10.1097/PSY.0b013e3181810658] [Medline: 18725434]

71. Gaab J, Rohleder N, Nater UM, Ehlert U. Psychological determinants of the cortisol stress response: the role of anticipatory cognitive appraisal. Psychoneuroendocrinology 2005 Jul;30(6):599-610. [doi: 10.1016/j.psyneuen.2005.02.001] [Medline: 15808930]

72. Oldehinkel AJ, Ormel J, Bosch NM, Bouma EM, Van Roon AM, Rosmalen JG, et al. Stressed out? Associations between perceived and physiological stress responses in adolescents: the TRAILS study. Psychophysiology 2011 Apr;48(4):441-452. [doi: 10.1111/j.1469-8986.2010.01118.x] [Medline: 21361964]

73. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996 Jul;66(1):3-8. [doi: 10.1016/0304-3959(96)02994-6] [Medline: 8857625]

74. Scheier MF, Weintraub JK, Carver CS. Coping with stress: divergent strategies of optimists and pessimists. J Pers Soc Psychol 1986 Dec;51(6):1257-1264. [doi: 10.1037//0022-3514.51.6.1257] [Medline: 3806361]

75. Mattingly SM, Gregg JM, Audia P, Bayraktaroglu AE, Campbell AT, Chawla NV, et al. The tesserae project: large-scale, longitudinal, in situ, multimodal sensing of information workers. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI EA '19; May 4-9, 2019; Glasgow, UK p. 1-8. [doi: 10.1145/3290607.3299041]

76. The MITRE Corporation. URL: https://www.mitre.org/ [accessed 2020-12-03]

77. Mackinnon A, Jorm AF, Christensen H, Korten AE, Jacomb PA, Rodgers B. A short form of the Positive and Negative Affect Schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. Pers Individ Dif 1999 Sep;27(3):405-416. [doi: 10.1016/s0191-8869(98)00251-7]

78. Merz EL, Malcarne VL, Roesch SC, Ko CM, Emerson M, Roma VG, et al. Psychometric properties of Positive and Negative Affect Schedule (PANAS) original and short forms in an African American community sample. J Affect Disord 2013 Dec;151(3):942-949 [FREE Full text] [doi: 10.1016/j.jad.2013.08.011] [Medline: 24051099]

79. Davey HM, Barratt AL, Butow PN, Deeks JJ. A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. J Clin Epidemiol 2007 Apr;60(4):356-360. [doi: 10.1016/j.jclinepi.2006.07.015] [Medline: 17346609]

80. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol 2017 Apr 7;8:456 [FREE Full text] [doi: 10.3389/fpsyg.2017.00456] [Medline: 28439244]

81. Georgiou K, Larentzakis AV, Khamis NN, Alsuhaibani GI, Alaska YA, Giallafos EJ. Can wearable devices accurately measure heart rate variability? A systematic review. Folia Med (Plovdiv) 2018 Mar 01;60(1):7-20. [doi: 10.2478/folmed-2018-0012] [Medline: 29668452]

82. Martinez GJ, Mattingly SM, Mirjafari S, Nepal SK, Campbell AT, Dey AK, et al. On the quality of real-world wearable data in a longitudinal study of information workers. In: Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops. 2020 Presented at: PerCom Workshops '20; March 23-27, 2020; Austin, TX, USA p. 1-6. [doi: 10.1109/percomworkshops48775.2020.9156113]

83. Champseix R. hrv-analysis: a package to calculate features from Rr Interval for HRV analyses. GitHub. 2021 Apr 12. URL: https://github.com/Aura-healthcare/hrvanalysis [accessed 2022-01-13]

84. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep 28;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]

85. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 1996 Mar 01;93(5):1043-1065. [doi: 10.1161/01.CIR.93.5.1043] [Medline: 8598068]

86. Francesco B, Maria Grazia B, Emanuele G, Valentina F, Sara C, Chiara F, et al. Linear and nonlinear heart rate variability indexes in clinical practice. Comput Math Methods Med 2012;2012:219080 [FREE Full text] [doi: 10.1155/2012/219080] [Medline: 22400047]

87. Roy B, Ghatak S. Nonlinear methods to assess changes in heart rate variability in type 2 diabetic patients. Arq Bras Cardiol 2013 Oct;101(4):317-327 [FREE Full text] [doi: 10.5935/abc.20130181] [Medline: 24008652]

88. Suzuki M, Hiroshi T, Aoyama T, Tanaka M, Ishii H, Kisohara M, et al. Nonlinear measures of heart rate variability and mortality risk in hemodialysis patients. Clin J Am Soc Nephrol 2012 Sep;7(9):1454-1460 [FREE Full text] [doi: 10.2215/CJN.09430911] [Medline: 22723446]

89. Salai M, Vassányi I, Kósa I. Stress detection using low cost heart rate sensors. J Healthc Eng 2016;2016:5136705 [FREE Full text] [doi: 10.1155/2016/5136705] [Medline: 27372071]

90. Burr RL. Interpretation of normalized spectral heart rate variability indices in sleep research: a critical review. Sleep 2007 Jul;30(7):913-919 [FREE Full text] [doi: 10.1093/sleep/30.7.913] [Medline: 17682663]

91. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 2010 Mar;1(1):3-14. [doi: 10.1111/j.2041-210x.2009.00001.x]

92. Folkman S, Moskowitz JT. Positive affect and the other side of coping. Am Psychol 2000 Jun;55(6):647-654. [doi: 10.1037//0003-066x.55.6.647] [Medline: 10892207]

93. Du J, Huang J, An Y, Xu W. The Relationship between stress and negative emotion: the Mediating role of rumination. Clin Res Trial 2018 Jan 31;4(1):1-5. [doi: 10.15761/crt.1000208]

94. Christensen RH. Cumulative Link Models for Ordinal Regression with the R Package ordinal. Technical University of Denmark & Christensen Statistics. 2019. URL: https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf [accessed 2021-09-08]

95. Kuznetsova A, Brockhoff PB, Christensen RH. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]

96. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]

97. Fox J, Weisberg S. An R Companion to Applied Regression. 3rd edition. Thousand Oaks, CA, USA: Sage Publications; 2019.

98. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. New York, NY, USA: Springer; 2013.

99. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol Evol 2012 Dec 03;4(2):133-142 [FREE Full text] [doi: 10.1111/j.2041-210x.2012.00261.x]

100. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY, USA: Routledge; 1988.

101. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. Law Hum Behav 2005 Oct;29(5):615-620. [doi: 10.1007/s10979-005-6832-7] [Medline: 16254746]

102. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol 2014 Sep 30;5:1040 [FREE Full text] [doi: 10.3389/fpsyg.2014.01040] [Medline: 25324790]

103. Mikkelsen K, Stojanovska L, Polenakovic M, Bosevski M, Apostolopoulos V. Exercise and mental health. Maturitas 2017 Dec;106:48-56. [doi: 10.1016/j.maturitas.2017.09.003] [Medline: 29150166]

104. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol 2003 Aug;85(2):348-362. [doi: 10.1037/0022-3514.85.2.348] [Medline: 12916575]

105. Gross JJ, John OP. Revealing feelings: facets of emotional expressivity in self-reports, peer ratings, and behavior. J Pers Soc Psychol 1997 Feb;72(2):435-448. [doi: 10.1037//0022-3514.72.2.435] [Medline: 9107009]

106. Ciuk D, Troy A, Jones M. Measuring Emotion: Self-Reports vs. Physiological Indicators. Social Science Research Network. 2015. URL: https://ssrn.com/abstract=2595359 or http://dx.doi.org/10.2139/ssrn.2595359 [accessed 2020-12-18]

107. Bardwell WA, Dimsdale JE. The impact of ethnicity and response bias on the self-report of negative affect. J Appl Biobehav Res 2001 Jan;6(1):27-38 [FREE Full text] [doi: 10.1111/j.1751-9861.2001.tb00105.x]

108. Martinez GJ, Mattingly SM, Young J, Faust L, Dey AK, Campbell AT, et al. Improved sleep detection through the fusion of phone agent and wearable data streams. In: Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops. 2020 Presented at: PerCom Workshops '20; March 23-27, 2020; Austin, TX, USA p. 1-6. [doi: 10.1109/percomworkshops48775.2020.9156211]

109. Das Swain V, Saha K, Rajvanshy H, Sirigiri A, Gregg JM, Lin S, et al. A multisensor person-centered approach to understand the role of daily activities in job performance with organizational personas. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec 11;3(4):1-27. [doi: 10.1145/3369828]

110. Mirjafari S, Masaba K, Grover T, Wang W, Audia P, Campbell AT, et al. Differentiating higher and lower job performers in the workplace using mobile sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-24. [doi: 10.1145/3328908]

111. Alam F, Riccardi G. Predicting personality traits using multimodal information. In: Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition. 2014 Presented at: WCPR '14; November 7, 2014; Orlando, FL, USA p. 15-18. [doi: 10.1145/2659522.2659531]

Abbreviations

AIC: Akaike information criterion

ANS: autonomic nervous system

BBI: beat-to-beat intervals

ECG: electrocardiogram

EMA: ecological momentary assessment

HF: high frequency

HPA: hypothalamic-pituitary-adrenal

HR: heart rate

HRV: heart rate variability

LF: low frequency

PNS: parasympathetic nervous system

RMSSD: root mean square of successive differences

SDANN: SD of the averages of normal-to-normal intervals

SDNN: SD of normal-to-normal intervals

SNS: sympathetic nervous system

VIF: variance inflation factor

VLF: very low frequency

XSL•FO
RenderX