

## Article

# Wearable-Measured Sleep and Resting Heart Rate Variability as an Outcome of and Predictor for Subjective Stress Measures: A Multiple N-of-1 Observational Study

Herman J. de Vries <sup>1,2,3,\*</sup> , Helena J. M. Pennings <sup>2,4</sup>, Cees P. van der Schans <sup>5,6</sup>, Robbert Sanderma <sup>3,7</sup>, Hilbrand K. E. Oldenhuis <sup>1,†</sup> and Wim Kamphuis <sup>2,†</sup> 

- <sup>1</sup> Research Group Digital Transformation, Hanze University of Applied Sciences, 9747 AS Groningen, The Netherlands
  - <sup>2</sup> Department of Human Behaviour & Training, Netherlands Organization for Applied Scientific Research (TNO), 3769 DE Soesterberg, The Netherlands
  - <sup>3</sup> Department of Health Psychology, University Medical Center Groningen, 9700 AB Groningen, The Netherlands
  - <sup>4</sup> Utrecht Center for Research and Development of Health Professions Education, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands
  - <sup>5</sup> Department of Rehabilitation Medicine, University Medical Center Groningen, 9700 AB Groningen, The Netherlands
  - <sup>6</sup> Research Group Healthy Ageing Allied Health Care and Nursing, Hanze University of Applied Sciences, 9747 AS Groningen, The Netherlands
  - <sup>7</sup> Department of Psychology, Health and Technology, University of Twente, 7522 NB Enschede, The Netherlands
- \* Correspondence: herman.devries@tno.nl  
† These authors contributed equally to this work.



**Citation:** de Vries, H.J.; Pennings, H.J.M.; van der Schans, C.P.; Sanderma, R.; Oldenhuis, H.K.E.; Kamphuis, W. Wearable-Measured Sleep and Resting Heart Rate Variability as an Outcome of and Predictor for Subjective Stress Measures: A Multiple N-of-1 Observational Study. *Sensors* **2023**, *23*, 332. <https://doi.org/10.3390/s23010332>

Academic Editor: Marco Altini

Received: 25 November 2022

Revised: 22 December 2022

Accepted: 26 December 2022

Published: 28 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The effects of stress may be alleviated when its impact or a decreased stress-resilience are detected early. This study explores whether wearable-measured sleep and resting HRV in police officers can be predicted by stress-related Ecological Momentary Assessment (EMA) measures in preceding days and predict stress-related EMA outcomes in subsequent days. Eight police officers used an Oura ring to collect daily Total Sleep Time (TST) and resting Heart Rate Variability (HRV) and an EMA app for measuring demands, stress, mental exhaustion, and vigor during 15–55 weeks. Vector Autoregression (VAR) models were created and complemented by Granger causation tests and Impulse Response Function visualizations. Demands negatively predicted TST and HRV in one participant. TST negatively predicted demands, stress, and mental exhaustion in two, three, and five participants, respectively, and positively predicted vigor in five participants. HRV negatively predicted demands in two participants, and stress and mental exhaustion in one participant. Changes in HRV lasted longer than those in TST. Bidirectional associations of TST and resting HRV with stress-related outcomes were observed at a weak-to-moderate strength, but not consistently across participants. TST and resting HRV are more consistent predictors of stress-resilience in upcoming days than indicators of stress-related measures in prior days.

**Keywords:** wearables; heart rate variability; sleep; stress; time series analysis; police; ecological momentary assessment; resilience

## 1. Introduction

Stress is associated with an increased risk of numerous diseases [1–7] and mental disorders [8,9]. Besides these adverse effects on individuals, it also imposes a large financial burden on society via absenteeism, healthcare costs, and productivity loss [10,11]. Personalized just-in-time interventions may be able to prevent or alleviate some of these burdens [12]. To do this, either the negative impact of stress or a decreased resilience to

cope with stress should be detected early, preferably via unobtrusive monitoring. For instance, unobtrusive detection of the negative impact of stress (e.g., on sleep or physiological systems) in an early state may help increase awareness that current circumstances may be causing wear and tear on bodily systems (allostatic load) that may be contributing to health-related problems if sustained over time [13]. Similarly, recognition of the potential depletion of resources that are needed for resiliently coping with challenges could be used to trigger feedback to take it easy that day and avoid overly challenging circumstances where possible. Recent developments in wearable sensor technology introduce promising opportunities for this type of unobtrusive monitoring [14,15].

When the first modern wearables came to market around 2009 (e.g., the Fitbit Classic), these devices initially became popular as pedometers or activity trackers but were already able to estimate sleep duration via accelerometry as well [16]. Since then, consumer wearable-based sleep tracking has improved to a point where it is considered proficient for measuring the Total Sleep Time (TST), while the detection of sleep stages needs further work [17]. Sleep deprivation is known to have a reciprocal relationship with stress, meaning that it is both caused and can be caused by stress [18]. Longitudinal studies with repeated daily measures confirm this bidirectional association [19–21] but tend to rely on subjective TST measures (e.g., measured via questionnaires) and need verification using objective sleep measurements [22]. Wearable-based research can therefore contribute to this body of knowledge and explore the potential of wearables to unobtrusively monitor for signs of the negative impact of stress or decreased resilience.

Besides behavioral outcomes such as physical activity and sleep, around 2015 (e.g., the Fitbit Charge HR) consumer wearables started measuring heart rate after photoplethysmography (PPG) sensors were included [23]. Today, PPG sensors are also used to track physiological outcomes such as heart rate, blood oxygen saturation, blood pressure, and respiration [24]. Perhaps the most important PPG-based innovation in the context of stress and resilience is the measurement of Heart Rate Variability (HRV), which can now be accurately measured using wearables or even camera-based smartphone apps in a resting state or during sleep [25]. HRV is a measure of the variation in heartbeats and is a proxy for autonomous nervous system functioning [26]. HRV acutely declines during stress [27] and afterward can remain suppressed during subsequent sleep [28,29]. Consequently, individuals with a low resting HRV are more likely to interpret seemingly mild stimuli as significant stressors [30–32] and have suboptimal emotion regulation [33,34]. Although these findings are based on population studies that investigated between-subject differences, the reciprocal nature of these findings illustrates that an initial decline in resting HRV could potentially cascade into subsequent days and thus have downstream effects.

A recent paper introduced a conceptual model in which the potential underlying mechanism for such a cascading effect of an initial decline in resting HRV was described [14]. The model suggests that resting HRV buffers against the impact of demands on stress by making potentially stressful situations seem less stressful [30–32], as well as against the impact of stress on mental exhaustion via more optimal emotion regulation [33,34]. Since this model also proposes that the need for recovery (e.g., increased mental exhaustion and/or decreased vigor) negatively influences resting HRV [28,29], a potential negative feedback loop is formed. This aligns with the conservation of resources theory, which states that since resources are needed to cope with demands, an initial loss of resources may result in a loss spiral [35]. Finally, the model hypothesizes stress to both be negatively impacted by stress [18–21], as well as to buffer against the negative impact of an increased need for recovery on resting HRV due to its restorative properties [36,37]. A study was then performed to test these hypotheses by utilizing wearables to measure TST and resting HRV, as well as an Ecological Momentary Assessment (EMA) smartphone app to measure subjective demands, stress, and mental exhaustion [38]. The study confirmed that resting HRV is both negatively impacted by mental exhaustion and buffers against the negative associations between demands and stress, as well as stress and exhaustion. Day-to-day changes in resting HRV may therefore be both indicative of the negative impact of stress and

predictive of stress-resilience, potentially even on a multi-day level. Further exploration of these potential multi-day bidirectional associations will improve our understanding of the degree to which day-to-day changes in wearable-measured resting HRV can be interpreted as potentially stress-related and in which they should be expected to reflect a state of lowered resilience.

To summarize: wearable-measured sleep and resting HRV have both been bidirectionally associated with subjective stress-related outcomes, but within-subject research investigating the potential patterns in multi-day associations in a real-world context is lacking. Increased insight into the degree to which these relationships are consistently observed in individuals may help improve models for the early recognition of the negative impact of stress and of lowered resilience. Such insights could contribute to the development of automated resilience interventions that may help to prevent stress-related problems. These interventions are especially relevant for individuals working in safety-critical professions, such as police officers [39]. Therefore, this study explores whether wearable-measured TST and resting HRV in police officers (1) can be predicted by stress-related EMA outcomes (demands, stress, mental exhaustion and vigor) in the preceding days, and (2) predict stress-related EMA outcomes in the subsequent days.

## 2. Materials and Methods

### 2.1. Study Design

An observational multiple n-of-1 study design was used [40], where individuals collected data on a daily basis, which were then individually assessed as independent time series. The results are therefore presented as a series ( $n$  = number of participants) of independent quantitative analyses on within-subject associations (e.g., a case series) based on samples with a high number of observations per participant ( $N$  = number of observations per participant) that can be relatively well-intercompared due to consistency in the applied methods. The current design is therefore optimized to provide a first exploration of possible multi-day associations at a within-subject level based on high-quality data, as is the aim of this study. Additionally, a rough estimate of the extent to which the respective associations may be found across individuals can be described in order to guide future studies. The current methods were based on a prior study that investigated nested within-day associations [14,38]. Missing data are problematic for time-series analysis. To limit missing data, we made several adaptations to optimize the previously used research design. We included automatic resting HRV measurements, a shorter daily Ecological Momentary Assessment (EMA) questionnaire, and an improved reward for adhering to the measurement protocol (participants were allowed to keep the wearable if they collected at least 100 complete daily observations). Data were collected for two purposes: (1) comparing longitudinal (5-week) trends in daily resting HRV and fluctuations therein to full questionnaire outcomes for stress, somatization, anxiety, and depression, and (2) the assessment of potential bidirectional and/or multi-day associations of sleep and resting HRV with stress-related EMA outcomes. The results of the former are published elsewhere [41], whereas the results of the latter are presented in this paper. The study protocol was approved by the ethical committee of the Hanze University of Applied Sciences Groningen (heac.2020.012).

### 2.2. Participants

Police officers working in a large Dutch city and possessing an Android- or iOS-based smartphone were invited to participate by the human resources bureau of their office. Interested respondents received the study information via e-mail. Participation was voluntary. The data collection period lasted a minimum of 15 weeks but could be extended with a number of additional 5-week periods. Extending the data-collection period was optional. Since at least 20 but preferably 50 observations with limited missing data are needed for accurate time series analysis [42,43], this data collection period (105–140 days) was expected to be appropriate to collect sufficient data. To also minimize potential missing

data participants could keep the wearable and received a personal feedback report after the study as a reward if they collected complete daily data for at least 100 days and completed all baseline and 5-weekly questionnaires. The recruitment period lasted until the maximum capacity of 10 participants was reached (i.e., it ran from June 2020 and July 2020). Before the start of their data collection, participants had a conversation with the first author during which the study requirements were explained, and participants gave their written informed consent. Due to COVID-19 restrictions, all contact with the participants occurred via teleconferencing and e-mail. After data collection, one participant was excluded because they were diagnosed with atrial fibrillation. This participant's data were excluded from the study, because this may have interfered with the accuracy of the HRV measurements. Another participant of whom only 56.3% of the daily observations were available was also excluded from the analysis. The remaining eight analyzed participants were predominantly male ( $n = 6$ ), had an average age of 37.0 years (range: 29–51), and contributed at least 80% (range: 80.7–96.8%) of complete observations.

### 2.3. Data-Collection

#### 2.3.1. Baseline Questionnaires

Immediately after consent was provided, participants were asked to fill in a baseline questionnaire. The baseline questionnaire included two items on gender and birthdate, as well as full questionnaires on personality traits (the Big Five Inventory; BFI) [44], symptoms of distress, somatization, depression, and anxiety (the Four-Dimensional Symptom Questionnaire; 4DSQ) [45], burnout (the Oldenburg Burnout Inventory; OLBI) [46] and work engagement (the Utrecht Work Engagement Scale; UWES) [47]. The outcomes of the baseline questionnaires and the mean values of the daily wearable and EMA outcomes were summarized per participant and on aggregate and used to describe the current sample for generalization purposes and as background information on the characteristics of the participants (Table 1). The age and gender of the individual participants were not described out of privacy considerations.

#### 2.3.2. Wearable-Based Variables

The Oura ring (generation 2, Oura Ring, Oulu, Finland) was used to measure TST and resting HRV during sleep. The Oura ring is a consumer wearable that measures sleep, physical activity, temperature, heart rate, and HRV. The consumer-available ring contains 2 infrared Light-Emitting Diode (LED) sensors, 2 Negative Temperature Coefficient (NTC) thermistor sensors, a tri-axial accelerometer, and a gyroscope. Although the algorithms that are used by the Oura ring to classify sleep and HRV based on the outputs of these sensors are proprietary, the ring (generation 2) has been confirmed to provide valid measurements of TST [48,49] and HRV [25,50,51] in independent research. Participants used a ring-size kit to determine their correct ring size to optimize fit for both user comfort and measurement accuracy and were allowed to choose a ring color of their preference. To preserve privacy, anonymized Oura accounts were created by using e-mail addresses on a custom domain to create accounts without the participants' names. The Oura-reported TST was used, which represents the total Duration of the Sleep Episode (DSE) minus the Sleep Onset Latency (SOL) and Wake-time After Sleep Onset (WASO). Similarly, the Oura-reported HRV was used, which represents the root Mean Square of the Successive Differences (rMSSD) in the inter-beat-intervals. This metric was then logarithmically transformed (lnrMSSD) to improve its distribution for statistical modeling, which is a common procedure in HRV research [52]. Finally, the Moderate-to-Vigorous Physical Activity (MVPA) was used as a control variable during analysis [53].

**Table 1.** Participant characteristics.

Baseline Questionnaires	Participant							
	1	2	3	4	5	6	7	8
Extraversion (1–5)	2.8	3.3	3.3	2.5	3.8	4.4	3.8	4.1
Agreeableness (1–5)	3.3	3.9	3.8	3.6	3.4	3.8	3.9	3.8
Conscientiousness (1–5)	3.9	3.8	3.3	3.3	3.8	3.0	3.8	4.0
Neuroticism (1–5)	2.1	2.5	2.4	2.4	2.5	2.3	2.9	2.4
Openness (1–5)	4.1	2.8	3.7	3.6	3.4	3.8	2.5	3.4
Distress	Low	Moderate	Low	Low	Low	Low	Moderate	Low
Depression	Low							
Anxiety	Low							
Somatization	Low							
Exhaustion (1–4)	3.0	2.5	3.3	2.5	3.0	2.6	2.6	2.6
Disengagement (1–4)	2.4	2.6	3.3	2.9	2.9	2.3	2.4	3.8
Work engagement (1–7)	5.4	4.6	5.3	4.5	5.2	5.1	4.2	5.9
<b>Daily Measurements</b>								
Total observations	147	125	386	150	285	143	147	140
% complete observations	81.6	94.4	88.3	80.7	96.8	93.0	95.9	80.7
TST (hours)	7.0 (1.5)	6.5 (0.8)	7.4 (1.5)	5.5 (1.3)	6.7 (1.3)	7.6 (1.5)	7.4 (1.3)	6.5 (1.2)
HRV (milliseconds)	51.3 (15.6)	43.8 (5.1)	54.8 (15.9)	72.8 (9.8)	47.4 (15.2)	29.6 (4.3)	39.6 (6.9)	26.8 (5.1)
Demands (0–10)	4.4 (3.0)	4.7 (2.0)	3.2 (1.6)	5.1 (2.7)	4.1 (1.2)	2.4 (1.9)	4.3 (2.7)	3.4 (2.2)
Stress (0–10)	3.9 (2.7)	3.5 (1.9)	2.6 (1.8)	3.3 (2.0)	1.5 (0.9)	1.6 (1.8)	1.3 (1.9)	2.8 (1.6)
Mental exhaustion (0–10)	2.8 (2.7)	4.8 (2.0)	3.7 (1.7)	4.2 (2.4)	2.1 (1.0)	1.4 (1.9)	2.6 (2.9)	3.3 (2.1)
Vigor (0–10)	5.3 (2.8)	5.6 (1.4)	5.8 (2.1)	5.0 (2.1)	4.9 (1.1)	6.0 (1.7)	5.5 (2.5)	7.4 (0.8)
MVPA (minutes)	54.2 (33.4)	49.3 (24.2)	45.4 (35.6)	28.0 (26.5)	49.7 (28.7)	33.4 (30.3)	25.7 (19.2)	18.5 (14.2)
Alcohol use (units)	0.4 (0.8)	0.5 (0.9)	0.4 (0.7)	0.1 (0.6)	0.5 (0.9)	0.1 (0.4)	0.9 (1.3)	0.1 (0.5)

Note. For the baseline questionnaires, the observed values are reported. For the daily measurements, mean and standard deviation are reported. TST: Total Sleep Time; HRV: Heart Rate Variability; rMSSD: root Mean Square of the Successive Differences; MVPA: Moderate-to-Vigorous Physical Activity.

### 2.3.3. Ecological Momentary Assessment-Based Variables

Every day at 7 PM, participants received a notification that a new EMA questionnaire was available on their smartphone app. Participants were instructed to complete the EMA before they went to bed. Since participants regularly worked night shifts, the EMA was available until 3 PM on the next day while participants received a reminder at noon to fill in their previous-day questionnaire if they had not finished it already. The EMA items were based on items used in a similar study [14,38]. The EMA measured: demands (“How demanding was your day?”), stress (“How much stress did you perceive today?”), mental exhaustion (“I felt mentally exhausted as a result of my activities”), vigor (“Do you feel like undertaking activities?”), and alcohol intake (“I consumed . . . alcoholic beverages today”). The demands, stress, and vigor items were scored on an 11-point Numeric Rating Scale (NRS), ranging from 0 (“Not at all”) to 10 (“Extremely”). Mental exhaustion was scored on an 11-point NRS ranging from 0 (“Strongly disagree”) to 10 (“Strongly agree”). The item for stress was based on a validated single-item scale [54], the item for mental exhaustion on an item of the Need For Recovery Scale [55], the item for vigor on an item of the Utrecht Work Engagement Scale [47], whereas the item for demands was self-composed in a similar style as the item for stress. The number of alcoholic beverages participants consumed during the passing day was included for use as a control variable during analysis and based on the AUDIT-C questionnaire [56], since alcohol consumption is known to impact resting HRV [57].

### 2.4. Data-Analysis

All analyses were performed in RStudio version 2022.7.1.554 [58] using R version 4.2.1 [59]. The ‘*zoo*’ package was used for linear interpolation of missing data [60], the ‘*vars*’ package was used for Vector Auto-Regression (VAR) modeling, Granger causation

testing, and Impulse Response Function (IRF) calculation [61]. Finally, 'ggplot2' was used to visualize the IRFs [62].

#### 2.4.1. Data Preparation

First, descriptive statistics on the number of observations, the percentage of complete observations, and the wearable- and EMA-variables were calculated based on the full set of collected data. Since VAR analyses do not allow for missing data, missing data were imputed via linear interpolation. Rows, where data were missing at the beginning or end of the time series, were removed, as these could not be imputed. All values were standardized (by first subtracting the within-subject mean from each daily value and then dividing it by the within-subject standard deviation) to optimize the inter-comparability of beta-coefficients and prevent multicollinearity. Finally, two versions of the vectors with the four core EMA variables (demands, stress, mental exhaustion, and vigor), two core wearable variables (TST and resting HRV), and two control variables (MVPA and alcohol consumption) were constructed to answer both research questions. The vector for the first analysis contained rows with values for the passing night's TST and nocturnal HRV, combined with the EMA items of the subsequent evening so that the lagged values of the EMA items (the values on the previous row that represent the EMA of the previous day) could be interpreted as predictors for TST and HRV (the values on the current row that represent the values for the passing night). For analysis 2, the TST and HRV data were shifted to the previous day, so that the lagged values of the TST and HRV (the values on the previous row, representing the passing night) could be interpreted as predictors for the EMA items (the values on the current row, representing the current day).

#### 2.4.2. Vector Auto-Regressive Modeling

To assess the stationarity of the time series as a prerequisite to performing VAR analysis, the Phillips-Perron (PP) unit root test was used on all variables [63]. All-time series were stationary (PP  $p < 0.05$ ). Next, the number of lags (i.e., number of preceding days included as predictor values) to include in the VAR model was determined. This was completed via the 'VARselect' function, which calculates models up to 7 lags (i.e., one full week's worth of lags). The most optimal lag order is based on four information criteria corresponding to the different models (i.e., Akaike Information Criterion (AIC), Hannan-Quinn (HQ) criterion, Schwarz Criterion (SC), and Final Prediction Error (FPE) criterion). The mode of these four information criteria was selected as the most optimal lag order used in the VAR model. In the case of a tie, the most conservative estimate was chosen. Assumptions were tested on the residuals of the VAR model. The residuals were assessed for autocorrelation via an asymptotic multivariate Portmanteau Test (PT) [64], for heterogeneity via an ARCH-LM test [65], and for normality via a Jarque-Bera (JB) test [66].

#### 2.4.3. Granger Causation Testing

To increase confidence in the predictive value of core EMA variables that were found to be statistically significant predictors of wearable variables (or vice versa) in the full VAR models, Granger causation tests were applied [67]. Granger causation tests assess if the inclusion of a predictor significantly improves a VAR. To isolate the direct relationships between these associations of interest from interrelations with the other variables in the vector, the Granger causation tests were applied to vectors with only the core predictor and outcome variables. Therefore, significant Granger causation tests showed that the predictor variable itself explains meaningful variance in the outcome variable and is not just significant in the VAR due to interrelations with other variables in the vector.

#### 2.4.4. Impulse Response Function Visualization

An IRF is the reaction of a dynamic system in response to an external change [68]. IRF visualizations of relevant predictors on the outcomes can illustrate how the outcome varies on subsequent days after being faced with an increase in a predictor variable. The

IRF of predictors that were both statistically significant in the full VAR model and in the additional Granger causation test were visualized. The IRF visualizations consisted of an overlay of participants where the respective association was observed. The IRFs with the same predictor were grouped in a grid in order to cluster visualizations of the multi-day impact of a predictor on all relevant outcomes (including bootstrapped 95% confidence interval (CI) based on 1000 runs).

### 3. Results

#### 3.1. Participant Characteristics

The eight participants, who were 29.4 to 51.1 years old (median = 36.8) and predominantly male (75%), collected 125 to 386 observations per person (median = 147) of which 80.7 to 96.8% (median = 90.7%) contained complete data on the EMA outcomes, as well as daytime and nighttime wearable outcomes. The average TST ranged from 5.5 to 7.6 h (median = 6.8), during which they had an average resting HRV (rMSSD) of 26.8 to 72.8 milliseconds (median = 45.6). The participants were moderate-to-vigorously physically active for 18.5 to 54.4 min per day (median = 39.4). The median reported daily scores on the stress-related outcomes was in the lower half of the scale (0–10) for demands (median = 4.2, range = 2.4–5.1), stress (median = 2.7, range = 1.3–3.9) and mental exhaustion (median = 3.0, range = 1.4–4.8). The mean reported daily scores on vigor were in the upper half of the scale (median = 5.6, range = 4.9–7.4). On average, the participants consumed between 0.1 and 0.9 (median = 0.4) alcoholic beverages per day. An overview of all participant characteristics is presented in Table 1.

#### 3.2. Analysis 1: Predicting TST and HRV by EMA

All analyzed time series were found to be stationary (PP unit root test  $p < 0.05$ ). The AIC, HQ, SC, and FPE information criteria that were used to determine the lag order for the VAR models unanimously suggested an optimal lag order of 1 in all participants, with exception of participant 7, where 2 out of 4 information criteria suggested a lag order of 2. Since the conservative option was chosen in case of a tie (§2.4.2), VAR models with 1 lag were created for all participants. No heterogeneity (ARCH-LM test  $p > 0.05$ ) was found in the residuals of any model. The residuals also contained no autocorrelation (PT  $p > 0.05$ ) in most participants, except for participant 5. This autocorrelation could not be resolved (e.g., by adding additional lags to the model), and suggests that an unobserved but relevant factor was not included in the model, which therefore may be useful but not complete. Finally, none of the residuals of any model were found to be normally distributed (JB-test  $p < 0.05$ ). This was likely attributable to the distribution of some of the EMA items, which were occasionally skewed or even bimodal. Since simulation studies showed that non-normally distributed residuals are not problematic in analyses with a sample of at least 100 observations [69], this was not considered to be a problem for the interpretation of these results.

The results of the VAR models on TST are presented in Table 2. Demands was a statistically significant ( $p < 0.05$ ) negative predictor of TST for three participants (4, 5, 7). For participant 5 this finding was confirmed by a statistically significant Granger causation test. Mental exhaustion was a significant positive predictor of TST in participant 4, but this was not confirmed in Granger causation testing and therefore interpreted as a potentially spurious relationship. Stress and vigor were not statistically significant predictors of TST in any model. The explained variance in the TST of the participant (5) in which demands was a significant predictor that was confirmed by a significant Granger causation test was 9%.

**Table 2.** Vector autoregression models for Total Sleep Time (TST) per participant (1–8).

Independent Variable	TST							
	1 $\beta$	2 $\beta$	3 $\beta$	4 $\beta$	5 $\beta$	6 $\beta$	7 $\beta$	8 $\beta$
Constant	0.00	0.03	−0.01	−0.00	0.01	−0.02	0.02	0.00
TST (lag 1)	0.08	0.06	−0.02	0.21 **	0.04	−0.10	−0.08	−0.05
HRV (lag 1)	0.07	0.01	−0.07	0.10	−0.10	0.04	−0.12	−0.08
MVPA (lag 1)	0.02	−0.09	0.05	0.18 *	0.04	−0.02	0.07	0.07
Alcohol intake (lag 1)	0.17	0.08	0.09	0.02	0.08	0.04	0.08	0.01
<u>Demands (lag 1)</u>	0.24	−0.02	0.05	−0.40 **	<u>−0.39 ***</u>	0.01	−0.21 *	0.02
Stress (lag 1)	−0.16	−0.13	−0.12	0.03	0.08	−0.07	−0.00	−0.12
Mental exhaustion (lag 1)	0.24	0.14	0.02	0.32 *	0.01	−0.14	0.10	−0.07
Vigor (lag 1)	−0.01	−0.06	−0.10	−0.10	0.10	0.16	−0.07	0.02
<i>N</i>	142	123	385	148	283	141	146	138
<i>Adjusted R</i> <sup>2</sup>	0.03	−0.04	0.02	0.11	0.09	0.03	0.01	−0.03
<i>F</i> -statistic	1.49	0.48	1.76	3.26 **	4.53 ***	1.47	1.19	0.54

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .  $p < 0.1$ ; Underlined: both the beta-coefficient and Granger causation test  $p < 0.05$ . TST: Total Sleep Time; HRV: Heart Rate Variability; MVPA: Moderate-to-Vigorous Physical Activity.

The results of the VAR models on HRV are presented in Table 3. Demands was a significant negative predictor of HRV in participant 3, which was also confirmed via Granger causation testing. Mental exhaustion was a significant positive predictor of HRV in participant 4, but it was in Granger causation testing and therefore interpreted as a potentially spurious relationship. Stress and vigor were not statistically significant predictors of HRV in any model. The explained variance in the HRV of the participant (3) in which demands was a significant predictor that was confirmed by a significant Granger causation test was 22%.

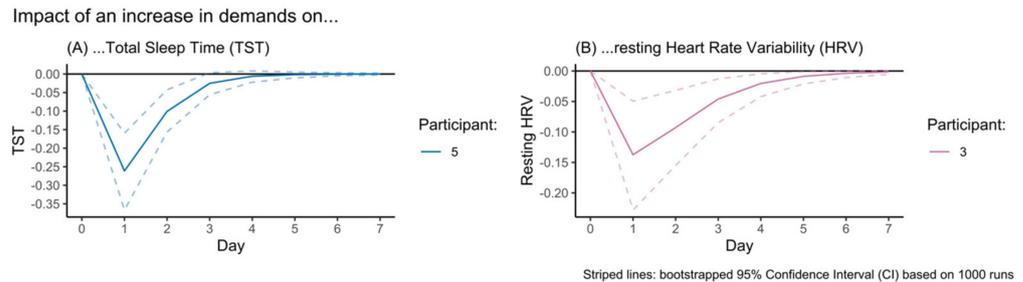
**Table 3.** Vector autoregression models for Heart Rate Variability (HRV) per participant (1–8).

Independent Variable	HRV							
	1 $\beta$	2 $\beta$	3 $\beta$	4 $\beta$	5 $\beta$	6 $\beta$	7 $\beta$	8 $\beta$
Constant	−0.01	0.02	−0.01	0.03	−0.01	−0.01	0.00	0.03
TST (lag 1)	0.04	0.00	−0.02	−0.03	−0.04	0.01	0.11	0.10
HRV (lag 1)	0.74 ***	0.06	0.40 ***	0.25 **	0.59 ***	0.30 ***	0.32 ***	0.36 ***
MVPA (lag 1)	0.03	0.03	−0.18 ***	0.10	−0.11 *	0.04	−0.14	−0.12
<u>Alcohol intake (lag 1)</u>	<u>−0.12 *</u>	0.06	<u>−0.19 ***</u>	−0.11	<u>−0.18 ***</u>	−0.10	−0.07	0.03
<u>Demands (lag 1)</u>	−0.01	0.01	<u>−0.16 *</u>	−0.20	0.01	0.16	−0.01	−0.08
Stress (lag 1)	0.05	0.03	−0.01	−0.22	−0.04	−0.23	0.04	−0.05
Mental exhaustion (lag 1)	0.05	−0.08	0.04	0.33 *	0.07	0.09	−0.07	−0.07
Vigor (lag 1)	0.05	−0.12	−0.02	0.15	−0.04	0.02	−0.06	0.06
<i>N</i>	142	123	385	148	283	141	146	138
<i>Adjusted R</i> <sup>2</sup>	0.59	−0.05	0.22	0.12	0.40	0.07	0.12	0.17
<i>F</i> -statistic	26.61 ***	0.34	14.20 ***	3.46 **	24.15 ***	2.41 *	3.40 **	4.51 ***

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .  $p < 0.1$ ; Underlined: both the beta-coefficient and Granger causation test  $p < 0.05$ . TST: Total Sleep Time; HRV: Heart Rate Variability; MVPA: Moderate-to-Vigorous Physical Activity.

To support the interpretation of the temporal associations where both the beta-coefficient and Granger causation tests were significant, IRF visualizations for the impact of an increase in demands on (A) TST and (B) HRV are displayed in Figure 1. In both outcomes, an increase in demands results in a sudden drop in the outcome variable, which then gradually recovers in subsequent days. However, the recovery of HRV takes longer (0 enters the 95% CI on the sixth day) than that of TST (0 enters the 95% CI on the third day). This difference can be attributed to the highly significant autoregression component in HRV ( $p < 0.001$ ), which is not observed in TST. This means that resting HRV values are relatively likely to

be similar to the previous day (e.g., if yesterday’s resting HRV value was relatively low, today’s value is likely to be relatively low again), whereas TST values have little to no association to the value of the previous day. The impact of demands, therefore, appears to be more long-lasting on HRV than on TST—at least in these participants.



**Figure 1.** Visualization of the Impulse Response Function (IRF) of the impact of an increase in demands on the Total Sleep Time (TST) and resting Heart Rate Variability (HRV) during the subsequent week.

3.3. Analysis 2 Predicting EMA by TST and HRV

The outcomes of the pre- and post-model diagnostic tests of analysis 2 were similar to those of analysis 1. The only difference in the pre- and post-model diagnostic tests of analysis 1 is that in analysis 2, participant 7 had just 1 out of 4 information criteria suggesting an optimal lag order of 2 instead of 2 out of 4. Therefore, VAR models with 1 lag were again created for all participants.

The results of the VAR models on demands are presented in Table 4. TST was a statistically significant negative predictor of demands in two participants (1, 2), which was confirmed with the Granger causation test in both cases. HRV was a significant negative predictor of demands in two participants (7, 8), also confirmed via Granger causation tests. The explained variance in the demands of these participants was 16% and 23%, respectively.

**Table 4.** Vector autoregression models for demands per participant (1–8).

Independent Variable	Demands							
	1 β	2 β	3 β	4 β	5 β	6 β	7 β	8 β
Constant	−0.00	0.01	−0.01	0.03	0.00	−0.01	0.02	0.02
Demands (lag 1)	0.03	0.17	0.21 **	0.49 ***	0.34 ***	−0.09	0.21 *	0.31 **
Stress (lag 1)	<u>0.25 *</u>	−0.05	−0.06	−0.02	<u>−0.17 *</u>	0.17	−0.10	−0.06
Mental exhaustion (lag 1)	0.03	−0.03	0.08	0.06	0.08	0.18	0.05	0.13
Vigor (lag 1)	0.06	0.04	0.02	−0.06	<u>0.15 *</u>	0.03	−0.05	−0.04
MVPA (lag 1)	0.09	−0.06	0.08	0.00	<u>0.12 *</u>	0.06	0.01	<u>0.19 *</u>
Alcohol intake (lag 1)	−0.13	−0.07	−0.04	−0.10	−0.04	0.00	<u>−0.24 **</u>	−0.12
TST (lag 1)	<u>−0.37 ***</u>	<u>−0.42 ***</u>	−0.09	−0.06	0.00	−0.12	−0.11	0.08
HRV (lag 1)	0.06	0.14	−0.02	0.03	−0.07	0.11	<u>−0.21 **</u>	<u>−0.18 *</u>
N	142	123	385	148	283	141	146	138
Adjusted R <sup>2</sup>	0.24	0.15	0.04	0.28	0.14	0.05	0.16	0.23
F-statistic	6.50 ***	3.63 **	3.23 **	8.18 ***	6.72 ***	1.94	4.41 ***	6.01 ***

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ,  $p < 0.1$ ; Underlined: both the beta-coefficient and Granger causation test  $p < 0.05$ . MVPA: Moderate-to-Vigorous Physical Activity; HRV: Heart Rate Variability; TST: Total Sleep Time.

Table 5 contains the results of the VAR models on stress. TST was a significant negative predictor of stress in three participants (1, 2, 7), all confirmed via the Granger causation test. HRV was a significant negative predictor of stress in one participant (8), again confirmed via Granger causation tests. The explained variance in the stress of these participants ranged from 2% to 23%.

**Table 5.** Vector autoregression models for stress per participant (1–8).

Independent Variable	Stress							
	1 $\beta$	2 $\beta$	3 $\beta$	4 $\beta$	5 $\beta$	6 $\beta$	7 $\beta$	8 $\beta$
Constant	−0.00	−0.00	−0.01	0.03	−0.00	0.01	−0.01	0.03
Demands (lag 1)	−0.03	0.01	−0.06	0.04	−0.02	−0.04	0.01	−0.10
Stress (lag 1)	0.32 *	0.06	0.22 **	0.49 ***	0.19 *	0.13	0.18	0.30 **
Mental exhaustion (lag 1)	0.02	0.06	0.05	0.01	0.07	0.14	−0.01	−0.10
Vigor (lag 1)	0.11	−0.03	−0.05	−0.00	0.11	0.06	0.01	0.00
MVPA (lag 1)	0.04	0.03	0.07	−0.07	0.09	0.01	−0.02	0.15
Alcohol intake (lag 1)	<u>−0.18 *</u>	−0.13	−0.08	−0.09	−0.02	−0.00	−0.01	−0.04
TST (lag 1)	<u>−0.19 *</u>	<u>−0.49 ***</u>	−0.04	0.01	−0.01	−0.05	<u>−0.19 *</u>	0.05
HRV (lag 1)	−0.12	0.08	0.03	−0.00	−0.02	0.08	0.06	<u>−0.29 **</u>
<i>N</i>	142	123	385	148	283	141	146	138
<i>Adjusted R<sup>2</sup></i>	0.18	0.23	0.04	0.23	0.05	0.01	0.02	0.12
<i>F-statistic</i>	4.88 ***	5.44 ***	3.17 **	6.56 ***	2.80 **	1.09	1.39	3.32 **

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .  $p < 0.1$ ; Underlined: both the beta-coefficient and Granger causation test  $p < 0.05$ . MVPA: Moderate-to-Vigorous Physical Activity; HRV: Heart Rate Variability; TST: Total Sleep Time.

The results of the VAR models on mental exhaustion are presented in Table 6. TST was a significant negative predictor of mental exhaustion in five participants (1, 2, 3, 5, 7), all confirmed via Granger causation tests. HRV was a significant negative predictor of mental exhaustion in one participant (8), again confirmed via Granger causation testing. The explained variance in the mental exhaustion of these participants ranged from 3% to 36%.

**Table 6.** Vector autoregression models for mental exhaustion per participant (1–8).

Independent Variable	Mental Exhaustion							
	1 $\beta$	2 $\beta$	3 $\beta$	4 $\beta$	5 $\beta$	6 $\beta$	7 $\beta$	8 $\beta$
Constant	−0.00	0.01	−0.01	0.02	−0.00	0.00	0.01	0.03
Demands (lag 1)	0.08	0.05	0.13	<u>0.25 *</u>	−0.03	−0.14	0.05	0.17
Stress (lag 1)	−0.07	−0.18	−0.17 *	0.17	0.02	0.22	0.08	0.02
Mental exhaustion (lag 1)	0.23	0.27 *	0.13	0.17	0.30 ***	0.04	0.16	0.01
Vigor (lag 1)	0.03	−0.07	−0.04	−0.14	0.00	−0.11	0.05	−0.05
MVPA (lag 1)	0.04	−0.06	0.01	−0.07	−0.05	0.08	−0.07	0.11
Alcohol intake (lag 1)	−0.17	−0.10	−0.02	−0.05	0.11	−0.05	−0.14	−0.09
TST (lag 1)	<u>−0.17 *</u>	<u>−0.41 ***</u>	<u>−0.12 *</u>	−0.05	<u>−0.25 ***</u>	−0.11	<u>−0.17 *</u>	0.02
HRV (lag 1)	0.10	0.16	−0.06	0.02	−0.00	0.04	−0.12	<u>−0.22 *</u>
<i>N</i>	142	123	385	148	283	141	146	138
<i>Adjusted R<sup>2</sup></i>	0.07	0.17	0.03	0.36	0.15	0.02	0.08	0.08
<i>F-statistic</i>	2.37 *	4.05 ***	2.68 **	11.15 ***	7.04 ***	1.44	2.61 *	2.47 *

Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .  $p < 0.1$ ; Underlined: both the beta-coefficient and Granger causation test  $p < 0.05$ . MVPA: Moderate-to-Vigorous Physical Activity; HRV: Heart Rate Variability; TST: Total Sleep Time.

Finally, the results of the VAR models on vigor are presented in Table 7. TST was a significant positive predictor of vigor in five participants (1, 3, 4, 5, 7), all confirmed via Granger causation tests. HRV did not predict vigor in any participant. The explained variance in the vigor of these participants ranged from 8% to 34%.

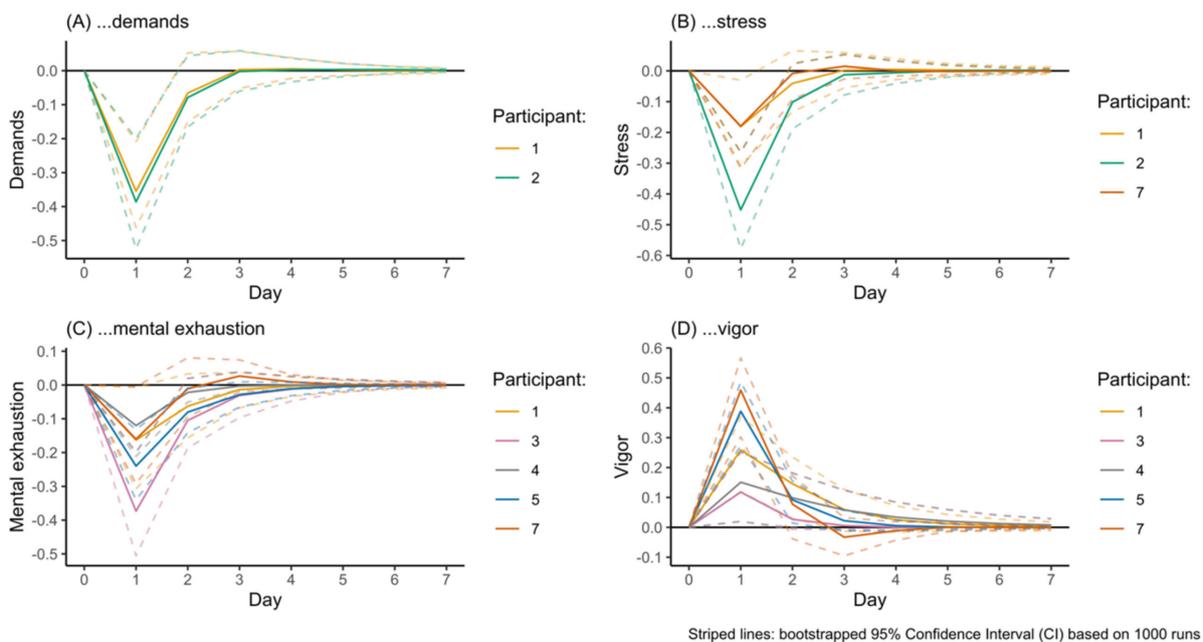
**Table 7.** Vector autoregression models for vigor per participant (1–8).

Independent Variable	Vigor							
	1 $\beta$	2 $\beta$	3 $\beta$	4 $\beta$	5 $\beta$	6 $\beta$	7 $\beta$	8 $\beta$
Constant	0.06	−0.00	−0.01	0.00	−0.00	−0.00	−0.01	0.02
Demands (lag 1)	−0.24 *	−0.10	−0.04	−0.16	−0.05	0.02	−0.09	0.07
Stress (lag 1)	<u>0.32 **</u>	−0.12	<u>0.14 *</u>	−0.03	0.09	0.04	−0.07	0.11
Mental exhaustion (lag 1)	−0.06	−0.16	0.04	0.01	−0.03	−0.14	0.02	−0.13
Vigor (lag 1)	0.37 ***	0.19 *	0.31 ***	0.48 ***	0.18 **	0.19 *	0.22 **	0.22 *
MVPA (lag 1)	−0.02	−0.04	0.02	0.05	0.06	0.05	0.10	0.13
Alcohol intake (lag 1)	0.11	−0.06	0.02	0.10	−0.04	0.12	0.10	−0.04
TST (lag 1)	<u>0.27 ***</u>	0.11	<u>0.12 *</u>	<u>0.17 *</u>	<u>0.41 ***</u>	0.13	<u>0.49 ***</u>	−0.17
HRV (lag 1)	0.06	−0.01	−0.01	−0.07	0.10	0.11	0.04	0.06
N	142	123	385	148	283	141	146	138
Adjusted R <sup>2</sup>	0.27	0.11	0.08	0.34	0.20	0.08	0.34	0.05
F-statistic	7.45 ***	2.93 **	5.23 ***	10.60 ***	9.62 ***	2.44	10.39 ***	1.81

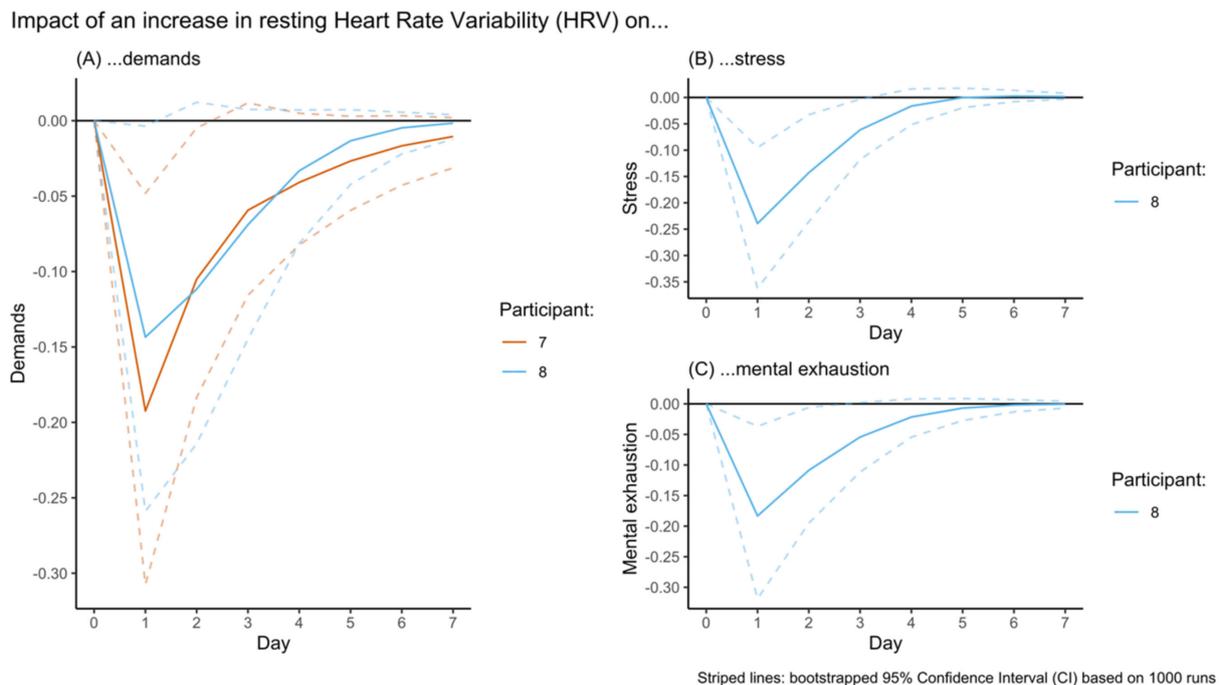
Note. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ ,  $p < 0.1$ ; Underlined: both the beta-coefficient and Granger causation test  $p < 0.05$ . MVPA: Moderate-to-Vigorous Physical Activity; HRV: Heart Rate Variability; TST: Total Sleep Time.

IRF visualizations for the impact of an increase in TST on each of the four EMA outcomes (A–D) are displayed in Figure 2. In all outcomes, an increase in TST resulted in a sudden decline (or incline in the case of vigor) that recovered (0 enters the 95% CI) in the subsequent 1 or 2 days. The IRF visualizations for the impact of an increase in HRV on the four EMA outcomes (Figure 3A–C) is similar for demands (1–2 days), although recovery from the impact on stress (2–3 days) and mental exhaustion (2–3 days) appears to take a bit longer. It appears that in these participants, the impact of changes in HRV is more long-lasting than for changes in TST.

Impact of an increase in Total Sleep Time (TST) on...



**Figure 2.** Visualization of the Impulse Response Function (IRF) of the impact of an increase in Total Sleep Time (TST) on the subsequent week's demands, stress, mental exhaustion, and vigor.



**Figure 3.** Visualization of the Impulse Response Function (IRF) of the impact of an increase in resting Heart Rate Variability (HRV) on the subsequent week's demands, stress, and mental exhaustion.

#### 4. Discussion

This study aimed to explore to what degree wearable-measured sleep and resting HRV in police officers (1) can be predicted by stress-related EMA outcomes in the preceding days, and (2) predict stress-related EMA outcomes in the subsequent days. After performing a time series analysis on eight participants, the results showed that associations in both directions of modest strength were observed and that TST and resting HRV were more consistent predictors for the next day's perceived demands, stress, mental exhaustion, and vigor than the other way around. Demands was a negative predictor of TST of one participant, and for resting HRV in another. Mental exhaustion predicted both resting HRV and TST in the same participant. Especially, TST seemed a strong predictor of stress-related EMA outcomes. TST negatively predicted demands in two participants, stress in three participants, mental exhaustion in five participants, and positively predicted vigor in five participants. Resting HRV negatively predicted demands in two participants, and both stress and mental exhaustion in one participant.

This study led to three key findings that will first be reflected upon, followed by a discussion of the strengths and limitations of the study, and finally a summary of the main conclusions and recommendations for future research.

##### 4.1. Associations between TST, HRV and EMA Outcomes Are Not Consistently Observed

Although TST was a negative predictor of mental exhaustion and a positive predictor of vigor in the majority (62.5%) of the participants, no association between a wearable- and an EMA-based item was consistently observed in all participants. No convincing explanations for the prevalence of these associations were identified after inspection of differences in the participant characteristics (Table 1).

The number of participants in this study ( $n = 8$ ) was too low to meaningfully assess to what extent between-subject differences in participant characteristics could predict the prevalence of these associations. Future studies with a larger sample size are recommended to explore if the occurrence or strength of these associations may be explained by participant characteristics, for instance via multilevel VAR [70]. If these differences can be explained

in future studies, they may be used to further personalize wearable-based models for stress-resilience.

It is also possible that the strength of these associations does not (only) depend on differences between individuals, but (also) on differences within individuals or in their environment. However, it may be difficult to determine beforehand what these influencing factors may be. It is possible to first explore if the strength of these relationships changes over time, for example via time-varying VAR models [71]. Detecting such changes over time is particularly feasible in datasets with a larger number of observations and/or more granular data. If these associations do change over time, it is possible that they may be actually relevant for all participants, but only under certain circumstances. Depending on the outcomes of such studies, it could provide new insights into the internal or external factors that determine when these associations are observed.

#### *4.2. The Impact of Changes in HRV Appears to Be More Abiding than That of Changes in TST*

The IRF visualizations in Figure 1 demonstrated that a demand-induced decline of resting HRV appears to have a longer recovery time (5–6 days) than a demand-induced decline of TST (2–3 days). Similarly, the impact of a change in resting HRV on stress-related EMA outcomes (Figure 3) appears to also be more long-lasting (1–3 days) than that of a change in TST (1 day) (Figure 2). This was attributed to the significant autoregression component that was observed in resting HRV, but not in TST. The strong autoregression component in the resting HRV model means that resting HRV values are relatively likely to be similar to those of the previous day(s). Therefore, a demand-induced decline in resting HRV (analysis 1) may take several days to recover from. Similarly, the impact of a decline in resting HRV on demands, stress, mental exhaustion, and vigor is likely to spill over into subsequent days, as it means that resting HRV is likely to remain suppressed for another few days.

This observation may be explained by the fundamentally different nature of the concepts resting HRV and TST. Resting HRV is a quantification of a physiological state that is continuously striving to maintain stability (homeostasis) despite disruptive challenges (allostasis) [13]. The recovery from a stressor that has a physiological impact (allostatic load) could take longer depending on the intensity and frequency of the stressor, as well as the quality and quantity of the subsequent recovery [36,37]. As such, a large decline in resting HRV can logically be expected to take some time as well. TST, on the other hand, is a quantification of the recovery process itself. Stress can negatively influence TST on the following night [18–21] and can therefore also impact TST on subsequent nights in the case of a recurring or sustained stressor. However, when this is not the case, it is also possible that the individual compensates for the previous sleep loss via recovery sleep [72], which would mean that TST on a subsequent night is no longer suppressed but actually increased. From this perspective, TST values can be expected to be more volatile than changes in resting HRV and thus have a weaker autoregression component. However, it is possible that changes in TST do have a longer-lasting impact on relevant underlying (psycho)physiological states such as vigor, which was observed to consistently have a significant autoregression component (Table 4).

The seemingly more abiding impact of a change in resting HRV on the resting HRV of the subsequent days may also be influenced by the development of a negative feedback loop. A previous study showed that evening mental exhaustion negatively impacted subsequent resting HRV and that resting HRV itself buffered against the positive association between demands and stress, as well as between stress and mental exhaustion [38]. This aligns with the Conservation of Resources Theory, which describes that an initial loss of resources could lead to a negative feedback loop. This means that fewer resources are available to handle upcoming challenges, which leads to lower resilience [35]. However, in the current study, no bidirectional association between a stress-related EMA item and resting HRV was observed within a single participant. Future studies with a larger sample are needed to increase insight into the multi-day impact of stress-related changes in resting HRV.

#### 4.3. TST and HRV Are More Consistent Predictors of Stress-Related Outcomes than Vice Versa

These findings indicate that wearable-measured TST and HRV seemed better predictors of stress-related EMA outcomes than the other way around. EMA-based predictions of TST and resting HRV were only observed in two participants, who had relatively large samples of observations ( $N = 385$  and  $N = 283$ ) compared to the median ( $N = 144$ ). Additionally, these relationships were not consistently observed in both participants. These differences cannot merely be explained by statistical power. Nevertheless, these models explained a modest amount of variance in TST (9%) and resting HRV (22%) in some participants. It is possible that these relationships are relatively small in nature and can only be observed in larger samples.

The finding that TST is a more consistent predictor of stress-related outcomes than that it can be predicted by stress-related outcomes aligns with prior research [22]. For instance, a lower TST has consistently been shown to predict increased stress [19–21,73]. Conversely, in the same studies, the opposite is regularly associated with smaller effect sizes [19,21], but in another study, TST was not associated with stress-related outcomes [73].

Similar scientific findings on the combination of both the predictive power and predictability of resting HRV in the context of stress-related outcomes are limited. However, the current findings do align with prior research, which has shown that stress-related outcomes negatively affect resting HRV [27,28,38] and that a relatively lower resting HRV than an individual's normal resting HRV can negatively impact stress-related outcomes on the following day [38,74].

One of the implications of this finding is that a decrease in wearable-measured TST or resting HRV does not necessarily point toward the occurrence of stress-related outcomes. Although the observed decrease in TST or resting HRV might have been caused by subsequent high demands or stress, this outcome may have been confounded by other factors. In situations where sudden extreme demands or stress occur, this might in some cases directly cause a decreased TST or resting HRV. However, in these circumstances, the wearable-user is likely already aware of the impact of such events. In such instances, the wearable-user less likely needs objective feedback to confirm this short-term effect.

Based on these findings, wearable-measured TST and resting HRV are not necessarily usable as a direct indication of the negative impact of stress but hold more promise to function as potential predictors to estimate one's resilience. For instance, these insights could be implemented in resilience interventions in the form of a readiness score that gives the user feedback on his or her expected readiness to handle mental and physical challenges that day [75]. Ideally, these factors will be expanded upon in future research (e.g., by also assessing behavioral outcomes such as smartphone usage, geolocation, or patterns in communication) that also explore different modeling approaches (e.g., machine learning) in order to improve the performance of these models.

#### 4.4. Strengths and Limitations of the Current Study

This study applied a novel research design and recruited a motivated number of participants that resulted in a series ( $n = 8$ ) of sizable datasets ( $N = 125$ – $386$ ) with mostly (80.7–96.8%) complete observations. By utilizing a consumer-available wearable that is validated for both TST and resting HRV measurements to collect observational data in a real-life environment, the generalizability of the findings to practical settings is relatively good. However, three limitations of the current study should also be considered during the interpretation of the presented results.

First, the multiple n-of-1 study design with a small number of participants ( $n$ ) but a large number of observations per participant ( $N$ ) was optimized as a first exploration of the potential existence of the hypothesized multi-day associations at a within-subject level based on high-quality data but limits the generalizability of the current findings to a broader target population. Therefore, future research with a larger number of participants is needed to increase confidence that the found associations are indeed relevant for larger groups of people. Future research is also needed to better understand why the identified

relationships are prevalent in some cases, but not in others. For instance, it is possible that studies with a larger number of observations per participant can unveil to what extent associations with a smaller strength can be observed in other participants, and to what extent the strength of these associations may change over time (e.g., via time-varying VAR).

Second, the included healthy participants and data collection during the COVID-19 lockdown might have affected the participants' perceptions of demanding and stressful situations. Their daily practice may not have been very demanding, which may have resulted in relatively low variance in the data. This aligns with the findings of a study on 2567 European police officers, which reported decreased strain during the pandemic [76]. The analyzed participants all scored relatively well on the mental well-being questionnaires (Table 1). Another article that was based on data from this same study population showed that some participants reported moderately elevated stress and somatization throughout the study period, but that there were no clinically relevant signs of anxiety and depression [41]. Future studies with a more mentally challenged sample need to verify the current findings for more challenging conditions.

Finally, some of the statistical assumptions of the created VAR models were technically violated. Most notably, none of the VAR models had normally distributed residuals, which was likely the result of sometimes skewed or bimodally distributed EMA items. Since simulation studies have shown that this assumption is particularly relevant when relatively small samples are assessed but not problematic when a sample of at least 100 observations is analyzed [70], this was not considered to be a problem for the interpretation of the results. The VAR model of participant 5 was also found to have autocorrelated residuals, which could not be resolved (e.g., by adding additional lags). Although this does not necessarily limit the interpretability of the model and related findings, it does show that the model is incomplete, and at least one unobserved but relevant factor was not included in the present study.

## 5. Conclusions

This multiple n-of-1 study showed that in relatively healthy police officers, demands were occasionally observed to be a negative predictor of wearable-measured TST and resting HRV. TST and resting HRV were more regularly observed to be negative predictors of demands, stress, or mental exhaustion, whereas TST also positively predicted vigor in several participants. The presented results illustrate that caution is needed when interpreting changes in TST and resting HRV to be potentially stress-related and that TST and resting HRV are more likely to be useful as predictors of stress-resilience (e.g., expressed as a readiness score).

However, since the identified associations were not consistently observed amongst participants, further research is necessary to better understand the underlying mechanism. For instance, future studies with a larger sample of participants, which is also needed to improve the generalizability of the current findings, could consider assessing if these between-subject differences could be explained by participant characteristics (e.g., via multilevel VAR). Another direction could be to explore if the strength of these associations' changes over time in samples with a larger number or more granular data (e.g., via time-varying VAR). Finally, future studies should explore if predictive models with a higher explained variance can be achieved by including additional data sources (e.g., smartphone usage, geolocation, or patterns in communication) or utilizing more inductive methods (e.g., machine learning approaches).

**Author Contributions:** Conceptualization, H.J.d.V., C.P.v.d.S., R.S., H.K.E.O., W.K.; methodology, all authors; formal analysis, H.J.d.V. and H.J.M.P.; investigation, H.J.d.V.; resources, C.P.v.d.S., R.S., H.K.E.O., W.K.; data curation, H.J.d.V.; writing—original draft preparation, H.J.d.V.; writing—review and editing, H.J.M.P., C.P.v.d.S., R.S., H.K.E.O., W.K.; visualization, H.J.d.V.; supervision, C.P.v.d.S., R.S., H.K.E.O., W.K.; project administration, H.J.d.V.; funding acquisition, C.P.v.d.S., R.S., H.K.E.O., W.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of Hanze University of Applied Sciences Groningen (protocol code heac.2020.012 and date of approval 11 June 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data of this article are not publicly available due to the personal nature of the data in combination with the sensitive professions of the participants in this study.

**Acknowledgments:** The authors thank Roos Delahajj (TNO) and Liselotte Eikenhout, (TNO, Open University), for their help the recruitment of participants. The Dutch police officers that participated are also thanked for their efforts during the relatively long data collection for this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Booth, J.; Connelly, L.; Lawrence, M.; Chalmers, C.; Joice, S.; Becker, C.; Dougall, N. Evidence of perceived psychosocial stress as a risk factor for stroke in adults: A meta-analysis. *BMC Neurol.* **2015**, *15*, 233. [\[CrossRef\]](#)
- Buscemi, V.; Chang, W.-J.; Liston, M.; McAuley, J.; Schabrun, S.M. The Role of Perceived Stress and Life Stressors in the Development of Chronic Musculoskeletal Pain Disorders: A Systematic Review. *J. Pain* **2019**, *20*, 1127–1139. [\[CrossRef\]](#)
- Decker, A.; Askar, H.; Tattan, M.; Taichman, R.; Wang, H.-L. The assessment of stress, depression, and inflammation as a collective risk factor for periodontal diseases: A systematic review. *Clin. Oral Investig.* **2020**, *24*, 1–12. [\[CrossRef\]](#)
- Li, J.; Zhang, M.; Loerbroks, A.; Angerer, P.; Siegrist, J. Work stress and the risk of recurrent coronary heart disease events: A systematic review and meta-analysis. *Int. J. Occup. Med. Environ. Health* **2014**, *28*, 8–19. [\[CrossRef\]](#)
- Magnavita, N.; Capitanelli, I.; Garbarino, S.; Pira, E. Work-related stress as a cardiovascular risk factor in police officers: A systematic review of evidence. *Int. Arch. Occup. Environ. Health* **2018**, *91*, 377–389. [\[CrossRef\]](#)
- Sui, H.; Sun, N.; Zhan, L.; Lu, X.; Chen, T.; Mao, X. Association between Work-Related Stress and Risk for Type 2 Diabetes: A Systematic Review and Meta-Analysis of Prospective Cohort Studies. *PLoS ONE* **2016**, *11*, e0159978. [\[CrossRef\]](#)
- Yang, T.; Qiao, Y.; Xiang, S.; Li, W.; Gan, Y.; Chen, Y. Work stress and the risk of cancer: A meta-analysis of observational studies. *Int. J. Cancer* **2019**, *144*, 2390–2400. [\[CrossRef\]](#)
- Van der Molen, H.F.; Nieuwenhuijsen, K.; Frings-Dresen, M.H.; de Groene, G. Work-related psychosocial risk factors for stress-related mental disorders: An updated systematic review and meta-analysis. *BMJ Open* **2020**, *10*, e034849. [\[CrossRef\]](#)
- Chirico, F. Job stress models for predicting burnout syndrome: A review. *Ann. Dell'istituto Super. Sanita* **2016**, *52*, 443–456.
- Hassard, J.; Teoh, K.R.H.; Visockaite, G.; Dewe, P.; Cox, T. The cost of work-related stress to society: A systematic review. *J. Occup. Health Psychol.* **2018**, *23*, 1. [\[CrossRef\]](#)
- Sultan-Taïeb, H.; Béjean, S. Modeling the economic burden of diseases imputable to stress at work. *Eur. J. Health Econ.* **2005**, *6*, 16–23.
- Wang, L.; Miller, L. Just-in-the-Moment Adaptive Interventions (JITAI): A Meta-Analytical Review. *Health Commun.* **2020**, *35*, 1531–1544. [\[CrossRef\]](#)
- McEwen, B.S. Stress, adaptation, and disease: Allostasis and allostatic load. *Ann. N. Y. Acad. Sci.* **1998**, *840*, 33–44. [\[CrossRef\]](#)
- De Vries, H.; Kamphuis, W.; Oldenhuis, H.; van der Schans, C.; Sanderman, R. Modelling employee resilience using wearables and apps: A conceptual framework and research design. *Int. J. Adv. Life Sci.* **2019**, *11*, 110–118.
- Drury, R.L. Wearable biosensor systems and resilience: A perfect storm in health care? *Front. Psychol.* **2014**, *5*, 853. [\[CrossRef\]](#)
- Leung, W.; Case, L.; Sung, M.-C.; Jung, J. A meta-analysis of Fitbit devices: Same company, different models, different validity evidence. *J. Med. Eng. Technol.* **2022**, *46*, 102–115. [\[CrossRef\]](#)
- De Zambotti, M.; Cellini, N.; Menghini, L.; Sarlo, M.; Baker, F.C. Sensors Capabilities, Performance, and Use of Consumer Sleep Technology. *Sleep Med. Clin.* **2020**, *15*, 1–30. [\[CrossRef\]](#)
- Nollet, M.; Wisden, W.; Franks, N.P. Sleep deprivation and stress: A reciprocal relationship. *Interface Focus* **2020**, *10*, 20190092. [\[CrossRef\]](#)
- Narmandakh, A.; Oldehinkel, A.J.; Masselink, M.; de Jonge, P.; Roest, A.M. Affect, worry, and sleep: Between- and within-subject associations in a diary study. *J. Affect. Disord. Rep.* **2021**, *4*, 100134. [\[CrossRef\]](#)
- Sperry, S.H.; Kwapil, T.R. Variability in Sleep Is Associated with Trait-Based and Daily Measures of Bipolar Spectrum Psychopathology. *Behav. Sleep Med.* **2022**, *20*, 90–99. [\[CrossRef\]](#)
- Yap, Y.; Slavish, D.C.; Taylor, D.J.; Bei, B.; Wiley, J.F. Bi-directional relations between stress and self-reported and actigraphy-assessed sleep: A daily intensive longitudinal study. *Sleep* **2020**, *43*, zsz250. [\[CrossRef\]](#)
- Ekuni, R.; Souza, B.M.N.; Cogo-Moreira, H.; Lourenço, F.C.; Pompeia, S. Bidirectional longitudinal dynamics of self-reported total sleep time and perceived stress: Establishing potential causal relationships. *Sleep Health* **2022**, *8*, 406–409. [\[CrossRef\]](#)
- Stahl, S.E.; An, H.-S.; Dinkel, D.M.; Noble, J.M.; Lee, J.-M. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exerc. Med.* **2016**, *2*, e000106. [\[CrossRef\]](#)
- Lee, I.; Park, N.; Lee, H.; Hwang, C.; Kim, J.; Park, S. Systematic Review on Human Skin-Compatible Wearable Photoplethysmography Sensors. *Appl. Sci.* **2021**, *11*, 2313. [\[CrossRef\]](#)

25. Stone, J.D.; Ulman, H.K.; Tran, K.; Thompson, A.G.; Halter, M.D.; Ramadan, J.H.; Stephenson, M.; Finomore, V.S.J.; Galster, S.M.; Rezai, A.R.; et al. Assessing the Accuracy of Popular Commercial Technologies That Measure Resting Heart Rate and Heart Rate Variability. *Front. Sports Act. Living* **2021**, *3*, 37. [[CrossRef](#)]
26. Thayer, J.F.; Åhs, F.; Fredrikson, M.; Sollers, J.J., III; Wager, T.D. A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* **2012**, *36*, 747–756. [[CrossRef](#)]
27. Kim, H.-G.; Cheon, E.-J.; Bai, D.-S.; Lee, Y.H.; Koo, B.-H. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investig.* **2018**, *15*, 235–245. [[CrossRef](#)]
28. Hall, M.; Vasko, R.; Buysse, D.; Ombao, H.; Chen, Q.; Cashmere, J.D.; Kupfer, D.; Thayer, J.F. Acute Stress Affects Heart Rate Variability During Sleep. *Psychosom. Med.* **2004**, *66*, 56–62. [[CrossRef](#)]
29. 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**, *111*, 733–741. [[CrossRef](#)]
30. Park, G.; Vasey, M.; Van Bavel, J.J.; Thayer, J.F. When tonic cardiac vagal tone predicts changes in phasic vagal tone: The role of fear and perceptual load. *Psychophysiology* **2014**, *51*, 419–426. [[CrossRef](#)]
31. Park, G.; Thayer, J.F. From the heart to the mind: Cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Front. Psychol.* **2014**, *5*, 278. [[CrossRef](#)]
32. Gaebler, M.; Daniels, J.K.; Lamke, J.-P.; Fydrich, T.; Walter, H. Heart rate variability and its neural correlates during emotional face processing in social anxiety disorder. *Biol. Psychol.* **2013**, *94*, 319–330. [[CrossRef](#)]
33. Mather, M.; Thayer, J.F. How heart rate variability affects emotion regulation brain networks. *Curr. Opin. Behav. Sci.* **2018**, *19*, 98–104. [[CrossRef](#)]
34. Holzman, J.B.; Bridgett, D.J. Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review. *Neurosci. Biobehav. Rev.* **2017**, *74*, 233–255. [[CrossRef](#)]
35. Hobfoll, S.E. The influence of culture, community, and the nested-self in the stress process: Advancing conservation of resources theory. *Appl. Psychol.* **2001**, *50*, 337–421. [[CrossRef](#)]
36. McEwen, B.S.; Karatsoreos, I.N. Sleep deprivation and circadian disruption: Stress, allostasis, and allostatic load. *Sleep Med. Clin.* **2015**, *10*, 1–10. [[CrossRef](#)]
37. McEwen, B.S. Sleep deprivation as a neurobiologic and physiologic stressor: Allostasis and allostatic load. *Metabolism* **2006**, *55*, S20–S23. [[CrossRef](#)]
38. De Vries, H.; Kamphuis, W.; Oldenhuis, H.; van der Schans, C.; Sanderman, R. Moderation of the Stressor-Strain Process in Interns by Heart Rate Variability Measured with a Wearable and Smartphone App: Within-Subject Design Using Continuous Monitoring. *JMIR Cardio* **2021**, *5*, e28731. [[CrossRef](#)]
39. Collins, P.; Gibbs, A. Stress in police officers: A study of the origins, prevalence and severity of stress-related symptoms within a county police force. *Occup. Med.* **2003**, *53*, 256–264. [[CrossRef](#)]
40. De Groot, M.; Drangsholt, M.; Martin-Sanchez, F.J.; Wolf, G. Single Subject (N-of-1) Research Design, Data Processing, and Personal Science. *Methods Inf. Med.* **2017**, *56*, 416–418. [[CrossRef](#)]
41. De Vries, H.; Kamphuis, W.; van der Schans, C.; Sanderman, R.; Oldenhuis, H. Trends in Daily Heart Rate Variability Fluctuations Are Associated with Longitudinal Changes in Stress and Somatisation in Police Officers. *Healthcare* **2022**, *10*, 144. [[CrossRef](#)]
42. Jebb, A.T.; Tay, L.; Wang, W.; Ehuang, Q. Time series analysis for psychological research: Examining and forecasting change. *Front. Psychol.* **2015**, *6*, 727. [[CrossRef](#)]
43. Warner, R.M. *Spectral Analysis of Time-Series Data*; Guilford Press: New York, NY, USA, 1998.
44. Denissen, J.J.A.; Geenen, R.; Van Aken, M.A.G.; Gosling, S.; Potter, J. Development and Validation of a Dutch Translation of the Big Five Inventory (BFI). *J. Pers. Assess.* **2008**, *90*, 152–157. [[CrossRef](#)]
45. Terluin, B.; van Marwijk, H.W.; Adèr, H.J.; de Vet, H.C.; Penninx, B.W.; Hermens, M.L.; van Boeijen, C.A.; van Balkom, A.J.L.M.; van der Klink, J.J.L.; Stalman, W.A.B. The Four-Dimensional Symptom Questionnaire (4DSQ): A validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry* **2006**, *6*, 34. [[CrossRef](#)]
46. Halbesleben, J.R.B.; Demerouti, E. The construct validity of an alternative measure of burnout: Investigating the English translation of the Oldenburg Burnout Inventory. *Work. Stress* **2005**, *19*, 208–220. [[CrossRef](#)]
47. Schaufeli, W.B.; Bakker, A.B.; Salanova, M. The measurement of work engagement with a short questionnaire: A cross-national study. *Educ. Psychol. Meas.* **2006**, *66*, 701–716. [[CrossRef](#)]
48. De Zambotti, M.; Rosas, L.; Colrain, I.M.; Baker, F.C. The Sleep of the Ring: Comparison of the ÖURA Sleep Tracker Against Polysomnography. *Behav. Sleep Med.* **2019**, *17*, 124–136. [[CrossRef](#)]
49. Chee, N.I.; Ghorbani, S.; Golkashani, H.A.; Leong, R.L.; Ong, J.L.; Chee, M.W. Multi-Night Validation of a Sleep Tracking Ring in Adolescents Compared with a Research Actigraph and Polysomnography. *Nat. Sci. Sleep* **2021**, *13*, 177–190. [[CrossRef](#)]
50. Kinnunen, H.; Rantanen, A.; Kenttä, T.; Koskimäki, H. Feasible assessment of recovery and cardiovascular health: Accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. *Physiol. Meas.* **2020**, *41*, 04NT01. [[CrossRef](#)]
51. Cao, R.; Azimi, I.; Sarhaddi, F.; Niela-Vilen, H.; Axelin, A.; Liljeborg, P.; Rahmani, A.M. Accuracy Assessment of Oura Ring Nocturnal Heart Rate and Heart Rate Variability in Comparison with Electrocardiography in Time and Frequency Domains: Comprehensive Analysis. *J. Med. Internet Res.* **2022**, *24*, e27487. [[CrossRef](#)]
52. Shaffer, F.; Ginsberg, J.P. An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* **2017**, *5*, 258. [[CrossRef](#)]

53. Henriksen, A.; Svartdal, F.; Grimsgaard, S.; Hartvigsen, G.; Hopstock, L.A. Polar Vantage and Oura Physical Activity and Sleep Trackers: Validation and Comparison Study. *JMIR Form. Res.* **2022**, *6*, e27248. [CrossRef]
54. Littman, A.J.; White, E.; Satia, J.A.; Bowen, D.J.; Kristal, A.R. Reliability and validity of 2 single-item measures of psychosocial stress. *Epidemiology* **2006**, *17*, 398–403. [CrossRef]
55. Van Veldhoven, M.; Broersen, S. Measurement quality and validity of the “need for recovery scale”. *Occup. Environ. Med.* **2003**, *60* (Suppl. 1), i3–i9. [CrossRef]
56. Bush, K.; Kivlahan, D.R.; McDonell, M.B.; Fihn, S.D.; Bradley, K.A. The AUDIT Alcohol Consumption Questions (AUDIT-C). An Effective Brief Screening Test for Problem Drinking. *Arch. Intern. Med.* **1998**, *158*, 1789–1795. [CrossRef]
57. Romanowicz, M.; Schmidt, J.E.; Bostwick, J.M.; Mrazek, D.A.; Karpyak, V.M. Changes in Heart Rate Variability Associated with Acute Alcohol Consumption: Current Knowledge and Implications for Practice and Research. *Alcohol. Clin. Exp. Res.* **2011**, *35*, 1092–1105. [CrossRef]
58. RStudio Team. *RStudio: Integrated Development Environment for R*; BPC: Boston, MA, USA, 2022; Available online: <https://www.rstudio.com/> (accessed on 30 July 2022).
59. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022; Available online: <https://www.R-project.org/> (accessed on 30 July 2022).
60. Zeileis, A.; Grothendieck, G. zoo: S3Infrastructure for Regular and Irregular Time Series. *J. Stat. Softw.* **2005**, *14*, 1–27. [CrossRef]
61. Pfaff, B. VAR, SVAR and SVEC Models: Implementation within R Package vars. *J. Stat. Softw.* **2008**, *27*, 1–32. Available online: <https://www.jstatsoft.org/v27/i04/> (accessed on 30 July 2022). [CrossRef]
62. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; Available online: <https://ggplot2.tidyverse.org> (accessed on 30 July 2022).
63. Phillips, P.C.; Perron, P. Testing for a unit root in time series regression. *Biometrika* **1988**, *75*, 335–346. [CrossRef]
64. Hosking, J.R. The multivariate portmanteau statistic. *J. Am. Stat. Assoc.* **1980**, *75*, 602–608. [CrossRef]
65. Engle, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **1982**, *50*, 987–1007. [CrossRef]
66. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *Int. Stat. Rev./Rev. Int. Stat.* **1987**, *55*, 163–172. [CrossRef]
67. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **1969**, *37*, 424–438. [CrossRef]
68. Lütkepohl, H. Impulse response function. In *Macroeconometrics and Time Series Analysis*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 145–150.
69. Lumley, T.; Diehr, P.; Emerson, S.; Chen, L. The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* **2002**, *23*, 151–169. [CrossRef]
70. Epskamp, S.; Deserno, M.K.; Bringmann, L.F. mlVAR: Multi-Level Vector Autoregression. 2021. Available online: <https://CRAN.R-project.org/package=mlVAR> (accessed on 30 July 2022).
71. Haslbeck, J.M.B.; Bringmann, L.F.; Waldorp, L.J. A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivar. Behav. Res.* **2021**, *56*, 120–149. [CrossRef]
72. Carskadon, M.; Acebo, C.; Seifer, R. Extended nights, sleep loss, and recovery sleep in adolescents. *Arch. Ital. Biol.* **2001**, *139*, 301–312.
73. Lee, S.; Crain, T.L.; McHale, S.M.; Almeida, D.M.; Buxton, O.M. Daily antecedents and consequences of nightly sleep. *J. Sleep Res.* **2017**, *26*, 498–509. [CrossRef]
74. Da Estrela, C.; MacNeil, S.; Gouin, J.-P. Heart rate variability moderates the between- and within-person associations between daily stress and negative affect. *Int. J. Psychophysiol.* **2021**, *162*, 79–85. [CrossRef]
75. Carper, B.; McGowan, D.; Miller, S.; Nelson, J.; Palombi, L.; Romeo, L.; Spiegelman, K.; Doryab, A. Modeling Biological Rhythms to Predict Mental and Physical Readiness. In Proceedings of the 2020 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA, 24–24 April 2020; pp. 1–6.
76. Frenkel, M.O.; Giessing, L.; Egger-Lampl, S.; Hutter, V.; Oudejans, R.R.; Kleygrewe, L.; Jaspaert, E.; Plessner, H. The impact of the COVID-19 pandemic on European police officers: Stress, demands, and coping resources. *J. Crim. Justice* **2021**, *72*, 101756. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.