**Title**
Stress and Human-Computer Interaction at the Workplace: Unobtrusive Tracking With Wearable Sensors and Computer Logs



**Author**
Akbar, Fatema

**Publication Date**
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Stress and Human-Computer Interaction at the Workplace:
Unobtrusive Tracking With Wearable Sensors and Computer Logs

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Informatics

by

Fatema Akbar

Dissertation Committee:
Professor Gloria Mark, Chair
Associate Professor Sarah Pressman
Assistant Professor Daniel Epstein

2021

# DEDICATION

To my parents, who gave me roots, and to my teachers, who gave me wings.

# TABLE OF CONTENTS


Page

**LIST OF FIGURES** **vii**

**LIST OF TABLES** **viii**

**ACKNOWLEDGMENTS** **ix**

**CURRICULUM VITAE** **x**

**ABSTRACT OF THE DISSERTATION** **xii**

**1 Introduction** **1**
1.1 Motivation . . . . . . . . . . 2
1.2 Thesis Statement . . . . . . . . . . 4
1.3 Research objectives . . . . . . . . . . 5
1.4 Dissertation outline . . . . . . . . . . 5

**2 Background: Stress and Computer Use in the Workplace** **8**
2.1 Defining stress . . . . . . . . . . 8
2.2 Stress measurement approaches . . . . . . . . . . 11
2.2.1 Stress measurement through self-reports . . . . . . . . . . 12
2.2.2 Physiological sensors . . . . . . . . . . 14
2.3 Theoretical frameworks for stress research . . . . . . . . . . 17
2.3.1 Individual differences in the experience of stress . . . . . . . . . . 17
2.3.2 Stress, motivation and performance . . . . . . . . . . 18
2.3.3 Stress and affect . . . . . . . . . . 19
2.3.4 Workplace stress . . . . . . . . . . 20
2.4 Modeling stress in the workplace with unobtrusive sensors . . . . . . . . . . 21
2.5 Computer-use factors potentially associated with stress . . . . . . . . . . 26
2.5.1 Email . . . . . . . . . . 26
2.5.2 Attention switching . . . . . . . . . . 28
2.5.3 After-hours work connectivity . . . . . . . . . . 28
2.6 Physician stress related to EHR systems and EHR inbox . . . . . . . . . . 29
2.7 Summary . . . . . . . . . . 31

**3 Methods: Unobtrusive Tracking of Stress and Computer Use 32**
3.1 Physiological stress through a wearable sensor . . . . . 32
3.1.1 Validation of the Garmin Stress Score . . . . . 33
3.1.2 Technical setup for data collection and storage . . . . . 36
3.2 Perceived stress through Ecological Momentary Assessments . . . . . 38
3.3 Computer activity logging . . . . . 38
3.4 Security and privacy . . . . . 39

**4 Information Workers' Stress and Computer Interaction 41**
4.1 Introduction . . . . . 41
4.2 Methods . . . . . 45
4.2.1 Recruitment . . . . . 45
4.2.2 Study protocol . . . . . 46
4.2.3 Computer activity logging . . . . . 47
4.2.4 Measures . . . . . 48
4.2.5 Analysis . . . . . 53
4.3 Results . . . . . 55
4.3.1 Daily stress . . . . . 55
4.3.2 Duration of computer work . . . . . 55
4.3.3 Computer work strategies and patterns . . . . . 58
4.3.4 Computer activity types . . . . . 60
4.3.5 Variability of computer work duration . . . . . 61
4.4 Discussion . . . . . 62
4.4.1 What computer use factors are associated with daily stress at the workplace? How do individual differences affect those factors? . . . . 62
4.4.2 How does variability in computer use patterns affect stress at the workplace? . . . . . 68
4.4.3 To what extent does unobtrusive monitoring of workplace computer use help identify stress levels? . . . . . 69
4.4.4 Limitations . . . . . 71
4.5 Conclusion . . . . . 72

**5 Information Workers' Perspectives on Technology-Supported Stress Tracking 73**
5.1 Introduction . . . . . 73
5.2 Methods . . . . . 75
5.3 Results . . . . . 77
5.3.1 Perceived benefits . . . . . 77
5.3.2 Perceived challenges . . . . . 80
5.3.3 Conflicting preferences for stress-tracking at the workplace . . . . . 84
5.4 Discussion . . . . . 86
5.4.1 Design implications for the validation vs. reflection on stress . . . . . 86
5.4.2 Trade-offs between unobtrusiveness, engagement, and value . . . . . 88
5.4.3 Designing for varying and conflicting preferences . . . . . 90
5.4.4 Limitations . . . . . 92

5.5 Conclusion . . . . . . . . . . 93

6 **Physicians' Electronic Inbox Work Patterns and Factors Associated With High Inbox Work Duration** **95**
6.1 Introduction . . . . . . . . . . 95
6.2 Study setting . . . . . . . . . . 96
6.3 EHR system logs . . . . . . . . . . 97
6.4 Statistical analyses . . . . . . . . . . 99
6.5 Results . . . . . . . . . . 100
6.5.1 Participants . . . . . . . . . . 100
6.5.2 Time spent on inbox management . . . . . . . . . . 100
6.5.3 Daily patterns of electronic inbox work compared with other EHR work 102
6.5.4 Factors associated with high duration of time on inbox work . . . . . 102
6.6 Discussion . . . . . . . . . . 106
6.6.1 Major findings . . . . . . . . . . 106
6.6.2 Interpretation and comparison with past studies . . . . . . . . . . 107
6.6.3 System design and organizational implications . . . . . . . . . . 109
6.6.4 Limitations . . . . . . . . . . 110
6.7 Conclusion . . . . . . . . . . 111

7 **Physicians' Stress and EHR Inbox Work Patterns** **112**
7.1 Introduction . . . . . . . . . . 112
7.2 Recruitment and protocol . . . . . . . . . . 113
7.3 Analysis . . . . . . . . . . 114
7.4 Results . . . . . . . . . . 116
7.4.1 Participants . . . . . . . . . . 116
7.4.2 Three distinct patterns of EHR inbox work . . . . . . . . . . 116
7.4.3 Stress patterns . . . . . . . . . . 119
7.4.4 EHR use characteristics associated with stress . . . . . . . . . . 121
7.5 Discussion . . . . . . . . . . 122
7.5.1 Principal findings . . . . . . . . . . 122
7.5.2 Comparison with previous work . . . . . . . . . . 123
7.5.3 Limitations . . . . . . . . . . 125
7.6 Conclusion . . . . . . . . . . 127

8 **Discussion and Conclusion** **128**
8.1 Summary of findings . . . . . . . . . . 128
8.2 Research contributions . . . . . . . . . . 137
8.3 Implications . . . . . . . . . . 139
8.3.1 Users . . . . . . . . . . 139
8.3.2 System designers . . . . . . . . . . 139
8.3.3 Personal Informatics community . . . . . . . . . . 140
8.3.4 Affective Computing and Context-Aware Computing . . . . . . . . . . 141
8.3.5 Mental health research community . . . . . . . . . . 141
8.3.6 Organizations . . . . . . . . . . 141

8.4 Limitations . . . 142
8.5 Challenges and ethical considerations . . . 144
8.6 Conclusion . . . 146

**Bibliography** **148**

**A Recruitment flyer** **176**

**B Screener survey** **178**

# LIST OF FIGURES


| | | Page |
|---|---|---|
| 2.1 | Components of stress and their traditional measurement approaches. | 11 |
| 2.2 | The Yerkes–Dodson Law showing the relationship between arousal and performance (Cohen, 2011). | 19 |
| 3.1 | The display of tracked stress scores on the wearable device and its app. | 35 |
| 3.2 | The mobile EMA questions and interface. | 39 |
| 4.1 | Samples of workdays where email work was batched or consistent. Green bars represent the duration on email work and the gray bars represent other computer work. | 51 |
| 4.2 | Interaction between non-workhours computer work duration and work-life balance on HRV-stress. | 59 |
| 4.3 | Interaction between batching and neuroticism on HRV-stress. | 60 |
| 4.4 | Interaction between window switching rate and neuroticism on HRV-stress. | 61 |
| 6.1 | Time spent on (A) the EHR inbox and (B) EHR functionality other than the inbox. Top figures show daily averages for each user (1257 users) and the bottom figures show overall average across user averages. | 103 |
| 7.1 | Temporal patterns of inbox and other EHR work. The green background indicates work hours. | 117 |
| 7.2 | Workday stress pattern per group. Error bars represent the standard error of the mean (SE). | 120 |

# LIST OF TABLES

Page

2.1 Sensors and signals of the reviewed studies. . . . . . . . . . . . . . . . . . . . . 22
2.2 Computer tasks/stressors of the reviewed studies. . . . . . . . . . . . . . . . . 23
2.3 Summary of reviewed studies. . . . . . . . . . . . . . . . . . . . . . . . . . . 24

4.1 Summary information about participants. . . . . . . . . . . . . . . . . . . . . 46
4.2 Nested model and Likelihood Ratio Test for information worker's model of daily HRV-based stress and computer use. . . . . . . . . . . . . . . . . . . . . 56
4.3 Generalized linear mixed model for information worker's daily self-reported stress and computer use. . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
4.4 Descriptive statistics of daily averages of computer use measures. N=51. . . 58
4.5 Bivariate correlations of stress measures (HRV-based, PSS, job stress surveys) and within-person regularity measures (SD, RI mean, RI range, FRI mean, FRI range). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63

6.1 Electronic inbox message views and time spent per day by primary care physicians in the Permanente Medical Group, March 2018. . . . . . . . . . . . . . 101
6.2 Average (SD) time spent on each message type during workdays and non-workdays. The overall column indicates mean percentages of total inbox time over the month. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102
6.3 Comparisons of high-duration and low-duration users of the electronic health record–based inbox among primary care physicians in the Permanente Medical Group, 2018. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105
6.4 Regression models predicting duration of all-day inbox work (per 24-hour period) and after-hours inbox work. . . . . . . . . . . . . . . . . . . . . . . 106

7.1 Comparing inbox use characteristics across three work patterns. . . . . . . . 119
7.2 Generalized linear mixed effects regression model. . . . . . . . . . . . . . . . 122

# ACKNOWLEDGMENTS

The work presented in this dissertation is the result of collaborations with, and support from, many extraordinary people. I would like to express my deepest gratitude and thanks for my advisor, Prof. Gloria Mark for her mentorship and support. Her experience and her passion for research have been so inspiring to me. You helped me grow as a person and as a scholar. Thanks also to my committee members and Dr. Tracy Lieu for their insightful feedback on my research.

I want to express my most sincere appreciation for Prof. Melissa Mazmanian, who provided unwavering support throughout my PhD journey.

I was fortunate to have friends with whom I shared this journey, with all its challenges and celebrations: Krithika Jagannath, Eugenia Rho and Elvan Bayraktaroglu, I am grateful to you. My brilliant lab-mates: Ted Grover, Yu Chen, Yiran Wang, Kevin Storer, Ge Gao, Alex Williams, Judith Borghouts and Thomas Breideband, thank you for the many interesting conversations and ideas. Thanks also to the research collaborators I worked with from different universities and institution.

Pursuing a PhD was a difficult decision to make. I must thank Dr. Ayman Bassil and others from Qatar Foundation, Dr. Ingmar Weber and others from the Qatar Computing Research Institute, and my friends for their advice, encouragement and support throughout this decision-making process. I am forever grateful for the Qatar Research Leadership Program, which truly opened many doors and opportunities for me.

Research presented in this dissertation was supported by the National Science Foundation, the Permanente Medical Group, and the Qatar National Research Fund (QNRF).

# CURRICULUM VITAE

## Fatema Akbar

### EDUCATION

**Ph.D. in Informatics** **2016-2021**
University of California, Irvine

**MSc. in Social Science of the Internet** **2014-2015**
University of Oxford

**B.S. in Information Systems and Business Administration** **2009-2013**
Carnegie Mellon University

### PUBLICATIONS

F. Akbar and Yasseri, T., (2021). “Engagement and Progression in Open Online ’Micro Lessons’: An Analysis of Learners’ Log Data from an Online Learning Platform,” *Proceedings of the AUBH E-Learning Conference 2021: Innovative Learning & Teaching - Lessons from COVID-19*, Available at SSRN: https://ssrn.com/abstract=3874415

F. Akbar, Mark, G., Prausnitz, S., et al., (2021). “Physician Stress During Electronic Health Record Inbox Work: In Situ Measurement With Wearable Sensors,” *JMIR Medical Informatics*, 9(4), e24014. doi: 10.2196/24014

T. A. Lieu et al., “Evaluation of Attention Switching and Duration of Electronic Inbox Work Among Primary Care Physicians,” *JAMA Network Open* 2021;4(1):e2031856. doi:10.1001/jamanetworkopen.2020.31856

F. Akbar et al., “Physicians’ electronic inbox work patterns and factors associated with high inbox work duration,” *Journal of the American Medical Informatics Association*, 2020;, ocaa229, https://doi.org/10.1093/jamia/ocaa229

S. Zaman, et al,.“Stress and productivity patterns of interrupted, synergistic, and antagonistic office activities,” *Scientific Data* 6, 264 (2019). https://doi.org/10.1038/s41597-019-0249-5

F. Akbar, G. Mark, I. Pavlidis and R. Gutierrez-Osuna, “An Empirical Study Comparing Unobtrusive Physiological Sensors for Stress Detection in Computer Work,” *Sensors*, 2019; 19(17):3766. https://doi.org/10.3390/s19173766

F. Akbar et al., "Email Makes You Sweat: Examining Email Interruptions and Stress Using Thermal Imaging, "*ACM Conference on Human Factors in Computing Systems (CHI'19)*, Glasgow, UK, 2019. Paper 668, 1–14. doi:10.1145/3290605.3300898

F. Akbar and I. Weber,"#Sleep_as_Android: Feasibility of Using Sleep Logs on Twitter for Sleep Studies," *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, Chicago, IL, 2016, pp. 227-233. doi: 10.1109/ICHI.2016.32

**CONFERENCE PRESENTATIONS**
**(Posters and Extended Abstracts)**

F. Akbar and K. Jagannath, "Personalized Models for Health and Wellbeing: Insights from an Ongoing Project on Unobtrusive Stress Tracking with Smartphones," *CHI Workshop on Next Steps Towards Long Term Tracking*, Toronto, Canada, 2018

F. Akbar, T. Grover, G. Mark and M. Zhou, "The Effects of Virtual Agents' Characteristics on User Impressions and Language Use," *ACM Intelligent User Interfaces (IUI 2018)*, Tokyo, Japan, 2018.

F. Akbar and L. Fernandez-Luque, "What's in the Store? A Review of Arabic Medical and Health Apps in the App Store," *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, Chicago, IL, 2016, pp. 413-413. doi: 10.1109/ICHI.2016.77

F. Akbar and I. Weber, "When Did They Join Twitter? Patterns of Twitter Adoption for Different Nationality Groups in The Arab Gulf Countries," *2nd International Conference on Computational Social Science (IC2S2 2016)*, Chicago, IL, 2016.

L. Fernandez-Luque and F. Akbar, "Arabic Health Social Media: Challenges, Opportunities and The Experience of QCRI," *International Saudi Health Informatics Conference (ISHIC 2016)*, Riyadh, Saudi Arabia, 2016.

# ABSTRACT OF THE DISSERTATION

Stress and Human-Computer Interaction at the Workplace:
Unobtrusive Tracking With Wearable Sensors and Computer Logs

By

Fatema Akbar

Doctor of Philosophy in Informatics

University of California, Irvine, 2021

Professor Gloria Mark, Chair

The relationship between workplace stress and computer use has mostly been investigated with self-reports or in controlled environments. However, self-report methods are prone to memory and emotion expression biases, and can be interruptive to employees when implemented for continuous stress tracking in real workplace environments. Researchers have explored the use of wearable sensors for unobtrusive and continuous stress tracking, but mostly in controlled laboratory settings, which limit the understanding of factors influencing stress in real-workplace environments, and the extent to which passive sensing can reveal information about stress during uncontrolled computer interactions.

This dissertation presents novel findings on computer use and stress at the workplace by employing computational methods leveraging computer activity logging and wearable devices that unobtrusively and continuously measured physiological stress through heart-rate variability in two real-world workplace settings: information work and medical work.

In the first part of the dissertation, fifty office employees were tracked for three to four weeks. Time spent on the work computer during and outside workhours, email work strategy, window switching, and computer activity types explained 14% of the variance in the daily stress duration. Individual differences (personality and work-life balance) moderate the relationship between workplace computer use factors and stress. A novel measure of variability in daily computer work was associated with perceived job demands, effort and overcommitment and arousal. Employees' perspectives on technology-supported stress tracking at the workplace indicated trust in algorithmic output, confirmation bias, and challenges balancing unobtrusiveness and engagement.

The second part of the dissertation analyzed how physicians use Electronic Health Record (EHR) systems and measured their physiological stress throughout the workday. One month of EHR logs of 1275 physicians were analyzed to characterize EHR use. Temporal patterns of EHR inbox use were found to be different from other EHR functions in their distribution throughout the day. Factors associated with high EHR inbox use were identified. Physiological stress data were collected for 47 physicians for a week and paired with their EHR logs. Among three patterns of EHR inbox work identified, the pattern characterized by working mostly outside of workhours had the longest average stress duration. Inbox work duration, the rate of EHR window switching, working outside of workhours and batching inbox work were associated with physicians' daily stress duration.

By evaluating a range of computer use factors and their association with daily physiological stress, the dissertation extends previous work that often focused on specific computer tasks or used self-reports. I provide recommendations and design implications for supporting different personal and organizational technology-supported stress tracking goals, and suggest future areas of work.

# Chapter 1

# Introduction

Imagine your computer knowing that you are experiencing stress and communicating with your virtual assistant (e.g. Amazon's Alexa or Apple's Siri) to play soothing music, or to inform your partner that you had a stressful day at work. For decades, researchers and science fiction writers have envisioned that future computers will be able to recognize our feelings and adapt to them. In the last 30 years, advances in affective computing [224] showed the possibility and importance of recognizing affective states for human-computer interaction. In this dissertation, I explore the relationship between stress and how employees interact with their computers at the workplace. Many individuals spend an increasingly significant proportion of their workday at a computer, especially those in information work. With the prevalence of workplace stress, it is sensible to expect that stress can be influenced by, and manifested in, workplace computer use.

## 1.1 Motivation

Workplace stress, which results from perceived job demands exceeding available resources [55], is a main factor for employee burnout, diminished productivity, and a number of health and wellbeing risks including cardiovascular disease and impaired immunity functions [55, 138, 194]. Some workplace computer tasks are known to be associated with stress, such as answering emails [145, 176] and presenting to a remote audience [306, 128]. Besides cognitively demanding tasks, workplace stressors include time pressure [196], social pressure [28], interruptions [172] and anticipatory stress from upcoming deadlines [8, 210]. Thus, capturing stress levels in the workplace is vital for improving our understanding of real-life stress and the factors surrounding it. Measuring stress unobtrusively and in real time at the workplace can enable affective computing applications that incorporate user's stress and new forms of context-aware interactions [226, 66]. Mental health professionals and organizational psychologists can also benefit from stress monitoring at the workplace, to better understand stress and associated factors, and to deliver interventions.

Quantification of affective states and work activities is becoming more widely adopted by individuals and organizations [72]. Tracking workplace practices can enable higher productivity [74, 134] and employees desire systems which help them pair their tracked activities with stress [195]. Tracking employees' stress in the workplace has also been widely used by researchers to understand what influences employees' stress. For example, prior work has leveraged stress monitoring to suggest an association between stress and email use at the workplace [176, 177]. Research has also suggested incorporating real-time stress data to provide just-in-time interventions to manage stress, including offering suggestions for users (e.g. playing games or guided breathing [219]) or changes in the computer interface (e.g. changes in the screen color or brightness [86] and managing stressful notifications [324]). Technology-supported stress tracking in the workplace can help individuals and organizations understand stress patterns and manage stressors.

However, measuring stress at the workplace is a non-trivial task. Workplace stress can be challenging to measure. Although several questionnaires have been developed to measure work-related stress (e.g. [126, 263]) or overall stress [52], these questionnaires are retrospective and not designed to measure real-time stress, as they mainly measure stress as a trait in the context of life events over weeks or months. In HCI, self-reports and wearable physiological sensors are commonly used for real-time stress tracking. Self-reports of stress represent a subjective evaluation based on cognitive appraisal of a given situation. Self-reports can provide frequent measures of stress when delivered through Ecological Momentary Assessments (EMAs) [262] where users are prompted with short questions about their stress multiple times throughout the day. Self-reports are subjective and are affected by memory and emotion expression biases [252, 94]. They can also be disruptive as they require the full cognitive attention of the user, and do not allow continuous stress measurement. Wearable sensors, on the other hand, do not require manual user input, but rather infer physiological stress from objective measures such as heart rate variability (HRV) [69, 161].

Advances in wearable sensors and the algorithms that filter and analyze their data enable objective continuous unobtrusive sensing of physiological measures directly associated with stress, such as HRV (See Ch.2.2.2). Measuring HRV throughout the day can give an objective and continuous measure of stress and relaxation, which can be used to identify events associated with stress in more granularity than is possible with self-reports.

Limited research has been done in real workplace environments using wearable sensors for stress tracking. This presents an opportunity to study factors associated with daily stress at the workplace in more depth and detail than has been previously done with self-reports. In previous work [7], we identified a gap in the literature as most studies modeling workplace stress with wearable sensors focus on specific high-stress short-duration computer tasks to induce stress in laboratory settings (e.g. [87, 166, 156]), which might not be representative of those in real workplace settings and can overlook issues and challenges related to stress measurement with physiological sensors during different computer activities. A limited number of recent in-situ studies used wearable sensors to assess the stress associated with specific computer tasks such as email [176], which shows promise for the deployment of computer activity logging and wearable sensors to investigate a wider range of computer work factors associated with daily stress.

## 1.2 Thesis Statement

The dissertation is structured around the following thesis statement:

*Continuous and unobtrusive tracking of workplace computer use and stress has several conceptual and methodological benefits. Computer use metrics can be quantified and can reveal information about daily stress in different work contexts. Individual factors are expected to affect the relationship between computer use and associated stress. Workplace tracking presents benefits and challenges to employees and organizations related to engagement with and understanding of the tracked data, and privacy concerns. Design and organizational recommendations can be made based on the observed associations between computer use metrics and stress, and employees' perceptions of tracking.*

To defend this statement, four studies were conducted in two workplace settings. First, information workers' daily computer use factors associated with stress are identified, by using computer logging and wearable devices. Second, Information workers' perceived benefits and challenges of workplace tracking are evaluated. Third, physicians' electronic health record use is analyzed to characterize patterns of work. Fourth, physicians' stress associated with electronic health record use is evaluated. For each study, I identify the design implications and recommendations for organizations.

## 1.3 Research objectives

The objective of this dissertation is to apply computational methods to quantify metrics about computer work, and to employ unobtrusive approaches for tracking these metrics, along with tracking stress, in real workplace settings. The specific aims of this dissertation are:

- To assess the extent to which tracking computer use patterns at the workplace predicts stress for different working populations
- To identify factors related to computer use that correlate with stress in the workplace
- To model and understand the underlying mechanisms through which individual stress is affected by different computer use factors by identifying individual differences that affect the relationship between computer use and stress
- To evaluate employees' perspectives on and experience with technology-supported stress tracking in the workplace
- To develop actionable guidelines for organizations and system designers to tackle stress associated with workplace computer use
- To provide recommendations for effectively deploying technology-supported stress tracking at the workplace

## 1.4 Dissertation outline

The outline of the remainder of this dissertation is as follows:

**Chapter 2** provides background information on stress and its measurement, and an overview of relevant research on stress tracking during computer work.

**Chapter 3** details the methods used in this dissertation work. The chapter reviews the data streams and data collection tools for physiological and perceived stress at the workplace, as well as computer use data.

**Chapter 4** investigates the relationship between information workers' stress and their computer interactions in the workplace. In particular, features extracted from computer activity logs of 51 employees are entered into a mixed model predicting daily stress duration as measured unobtrusively and continuously by wearable heart-rate sensors. In addition, the variability of daily time spent on work-related computer activities is correlated with several stress measures.

**Chapter 5** explores the benefits and challenges of two daily stress tracking modalities from the employees' perspective: automated tracking with wearable sensors and manual tracking through self-reports in EMAs. The chapter also provides actionable design guidelines for deploying daily stress tracking in the workplace.

**Chapter 6** presents a large-scale study of a potential stressor in medical computer work: inbox management. The chapter analyzes Electronic Health Records system use for 1,257 physicians. Specifically, we quantify the extent to which inbox work permeates physicians' time during and outside of workhours, describe daily patterns of inbox work, contrast temporal patterns of inbox work and other computer work, and identify factors associated with inbox work duration.

**Chapter 7** builds upon the findings from Chapter 6 with a study that analyzes objectively measured stress of physicians during their daily work and interaction with the EHR inbox. The study investigates physicians' EHR inbox work patterns by identifying clusters of distinct temporal inbox work patterns, measuring physicians' stress throughout the workday using wearable sensors, and identifying EHR inbox work factors associated with stress.

Finally, **Chapter 8** concludes the dissertation by discussing the major findings of the dissertation drawing on relevant previous work. The limitations of the dissertation are also discussed, as well as the implications of the presented findings and areas for future work in workplace stress tracking and system design.

# Chapter 2

# Background: Stress and Computer Use in the Workplace

This chapter provides an overview of relevant background information and research on stress and computer use at the workplace. In particular, the chapter explains: (1) the definition of stress in psychology and physiology, (2) stress measurement approaches, (3) theoretical frameworks for stress research, (4) literature review of stress tracking with sensors in the workplace and laboratory studies examining computer use, and (5) a review of workplace computer-use factors potentially associated with stress for information workers and physicians.

## 2.1 Defining stress

Stress is the perceived imbalance in demands and resources and is experienced when a situation is appraised as personally significant, taxing or exceeding resources for coping [79]. According to the American Psychological Association (APA), stress exists in two main forms:

acute (passing) and chronic (global, long-term) [1]. Acute stress is the result of interpreting a situation in the recent past or near future as requiring more resources (e.g. time, mental resources, money, etc.) than we have [92]. Most people occasionally experience acute stress. Besides its effect on performance, acute stress is associated with emotional distress, rapid heartbeat, shortness of breath, gut problems, and muscular problems such as tension, back pain and headaches [227, 205, 102]. These symptoms usually subside when short-term stress passes.

The second type of stress is chronic stress, which is due to long-standing pressures and demands including those experienced as a result of difficulties in socio-economic conditions, interpersonal relationships, or one's career [92]. Chronic stress depletes mental and physical resources and is associated with suicide, violence, heart attacks, immune dysregulation and an overall lower quality of life [40, 14, 190]. Both short and long term stress play a role in our daily functioning, especially for regulating important processes such as attention and memory acquisition.

In the psychology literature, stress is divided into two components: the stressor and the stress response [152]. Stressors can be thought of as stress antecedents, the triggers or environment in which stress occurs. Psychologists studied two types of stressors: daily hassles and life events. Lazarus (1984) argued that seemingly minor negative events of daily life are the most significant form of stress as they accumulate and affect health and wellbeing [151, 125]. Examples of events appraised by individuals as daily hassles include financial responsibilities, dislike of colleagues, feeling lonely and lacking sleep, among others [125]. Life events, on the other hand, are events that require significant change in one's accustomed pattern of life, and have been found to be related to stress and illness [113]. The Holmes-Rahe stress scale lists 43 life events including death of a spouse, foreclosure of mortgage, change in residence and change in habits [113].

[1] http://www.apa.org/helpcenter/stress-kinds.aspx

The second component of stress, the stress reaction or response, can be thought of as the immediate stress consequences. When the body experiences stress, a number of physiological events occur driven by two branches of the Autonomic Nervous System, which is a control system in the human body that regulates bodily functions. The first branch is the Sympathetic Nervous System, which drives the body's resources to respond to a challenge or a threat in ways such as quickening the pulse, deepening respiration, and tensing muscles, a reaction called the 'fight-or-flight' response [230]. During the fight-or-flight response, the systems that are not essential to immediate survival, such as the digestive system, the reproductive system and the immune system are suppressed, and more resources are allocated to the heart and brain. This process is controlled and complemented by the Parasympathetic Nervous System which regulates bodily functions at rest conditions. In non-stressful settings, these two systems work in coordination to achieve homeostasis, the condition where internal functions remain stable and balanced. In a stressful setting, the autonomic nervous systems are uncoordinated. Prolonged imbalance in these two systems leads to a physiological condition known as allostatic load [191, 190] in which the body fails to trigger appropriate responses for stress and rest condition, and would trigger a fight-or-flight response in non-stress conditions, making it difficult to return to a homeostasis condition and leading to serious health problems.

Beside the physiological response, psychological and emotional reactions include experiencing negative emotions such as annoyance, fear and anger [152]. Finally, behavioral consequences of stress include changes in health practices such as shortened and fragmented sleep [329] and less physical activity [278]. Behavioral consequences also include declining performance on complex tasks [315], and alterations in interpersonal behaviors such as insensitivity towards others [51] and social avoidance [53]. Figure 2.1 summarizes the two components of stress and their traditional measurement approaches, which are further detailed in the following section.

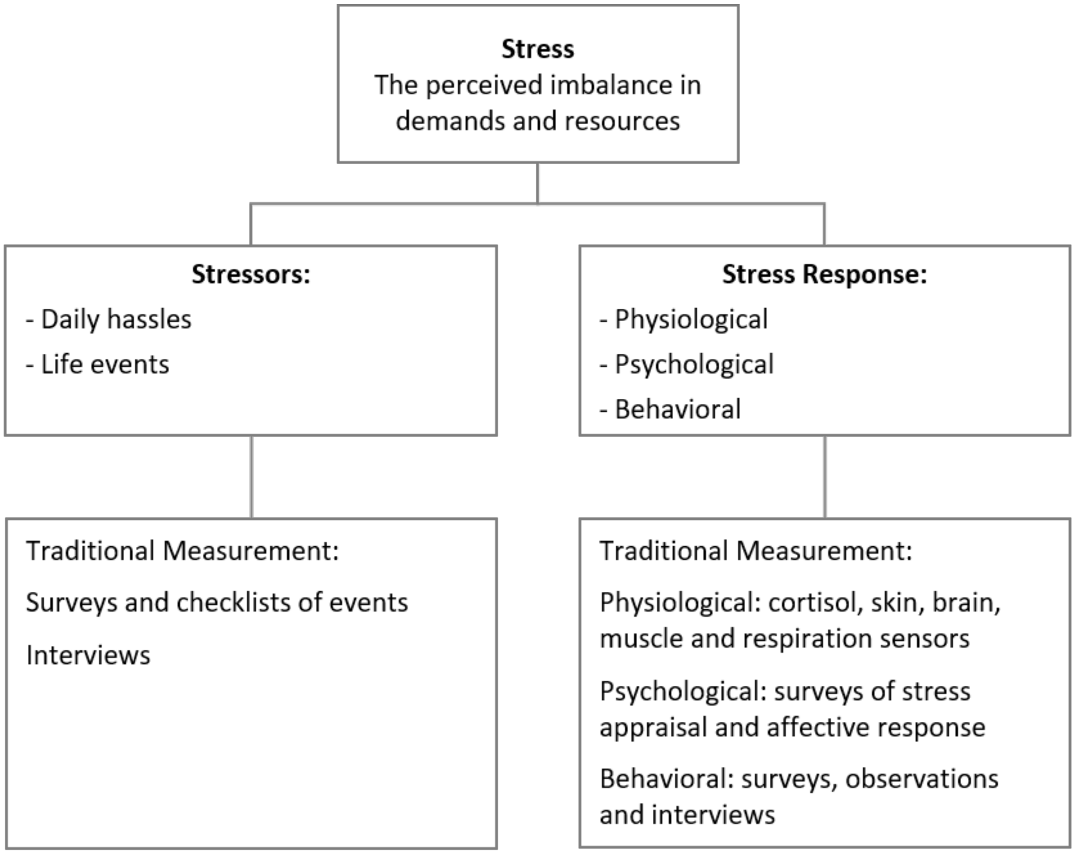



Figure 2.1: Components of stress and their traditional measurement approaches.

## 2.2 Stress measurement approaches

Cohen et al. (1997) outline three broad traditions of assessing stress: (1) the environmental tradition that focuses on the assessment of environmental events or experiences that are normatively associated with substantial adaptive demands, (2) the psychological tradition that focuses on individual's subjective appraisal of their ability to cope with demands, and (3) the biological tradition that focuses on activation of specific physiological systems that are triggered by demanding conditions. Data for these approaches are collected either subjectively through self-reports or objectively using sensors.

### 2.2.1 Stress measurement through self-reports

#### Retrospective Surveys

Numerous surveys have been developed to measure long and short-term stress experienced by individuals in different settings. Retrospective surveys vary in their retrospective period, as some measure stress in an immediate preceding event (e.g. NASA-TLX [98]) while others are based on life-long events (e.g. [272]). The Perceived Stress Scale (PSS) [52, 12] is a widely used instrument to measure appraised general stress. The questions are worded as "in the last month, how often have you felt...." to capture feelings of stress and loss of control. In HCI research, PSS is commonly used as a baseline for global stress (e.g. [176, 249]). Limitations of PSS include retrospective recall biases and the varying ability and willingness of people to accurately identify and express their emotions. Other surveys rely on reconstructing specific stressful events to quantify the stress experienced within a time period. For example, the Holmes and Rahe Stress Scale counts stressful events in the past year [113]. Similarly, but for a day rather than a year, the Daily Stress Inventory [37] quantifies the stress experienced during the previous day by counting the occurrence and intensity of relatively minor stressful events. To address recall bias in remembering events from the previous day, Kahneman et al. (2004) developed the Day Reconstruction Method (DRM) [124] that systematically guides individuals to reconstruct their activities and affective states from the previous day. Comparison between EMA and day reconstruction method showed that agreement between the two methods varies across different emotions [70]. The lowest agreement was on stress. This shows that even when people experience stress, they cannot recall it a day after. This is an important shortcoming in stress studies that rely on delayed self-reports.

### Ecological Momentary Assessments

A widely adopted method of collecting self-reports in daily life settings is Ecological Momentary Assessments [262] (a.k.a Experience Sampling Method, ESM [149]). Advancement in mobile computing enabled a wide adoption of this technique by prompting the user at specific times through their mobile phones (e.g. [244]) or smart-watches (e.g. [117]) for real-life stress monitoring. A commonly used mood assessment approach for self-reports through EMAs is based on the Russell's circumplex model [246], a two-dimensional approach for affect classification. With the circumplex model, affect is quantified on the dimensions of valence (positivity) and arousal (activeness). In EMAs delivered through mobile phones and wearables, users can be asked to rate their current mood on these two dimensions, which can then be used to infer the affective state (e.g. stress would be high arousal and low valence). One of the benefits of using EMAs on the everyday devices is that people are already carrying these devices and are used to sending and receiving information with them. Hernandez et al. (2016) compared EMAs delivered through multiple wearable devices to measure stress and found that differences in devices affect user responses [104]. Designing and delivering EMAs should take into account the effect of the delivery medium. Another limitation of EMAs is that receiving the prompts several times in a day can be disruptive or frustrating, which can affect answers to some questions about affective state and makes EMAs unsuitable for continuous long-term stress monitoring. Nevertheless, compared to retrospective surveys, EMAs minimize retrospective recall biases [25, 262] and allow for measuring the subjective experience of acute stress on-the-go in real-life settings.

Although self-report instruments are commonly used in the literature, they are subjective, require the full cognitive attention of the user, often do not capture context, and are affected by memory recall as well as emotion expression biases [252, 94]. While self-report instruments are not suitable for continuous unobtrusive measurement of stress, they remain important as perceived stress has been shown to have health and wellbeing implications [243, 146].

### 2.2.2 Physiological sensors

Stress produces a number of physiological responses in the human body. Most physiological reactions are momentary and do not carry over after the stressor has passed. Thus, while they are good at detecting stress when it happens, they are not very good at measuring carried-over experiences of stress as subjective surveys do. The physiological response to stress can be measured through measuring the level of cortisol, known as the stress hormone. While cortisol measurement for stress is used in clinical and lab studies (see [137] for a review), it requires collecting saliva or blood samples, which is intrusive and not suitable for continuous tracking. In addition, cortisol is not responsive to all types of stressors, as it is more responsive to social evaluative stress [67]. Therefore, measuring physiological changes related to the autonomic nervous system (e.g. changes in heart activity) rather than HPA activity (i.e. the cortisol hormone) are more appropriate for the minor and diverse stressors of the workplace environment and especially computer work which is not typically socially evaluative. Physiological sensors can provide continuous, unobtrusive stress tracking. This subsection focuses on skin and heart activity, which are commonly found in wearable devices and used in HCI research, and have been used in both lab and in-situ settings.

#### Skin Conductance

Electrodermal activity (EDA), also known as Galvanic Skin Response (GSR), is a physiological response to stress that captures the sympathetic nervous system's activation through skin reaction. EDA measures skin conductance, which is the skin's susceptibility to conduct electricity. When a person experiences stress, sweat glands are activated as a physiological response of sympathetic nervous system activation. When sweat glands are activated, perspiration decreases skin's resistance to electrical current. Thus, stress is related to higher skin conductance. Ideally, EDA should be measured at the palm or the fingers, where capturing even small EDA responses precisely is feasible [293]. In everyday settings, using sensors on the palms and fingers is not practical as it interferes with daily tasks. Thus, wearable wrist-bands equipped with electrodes are commonly used in HCI research to monitor EDA.

In a study using EDA to measure stress, Hernandez et al. (2011) developed personalized models to automatically recognize the stress levels for call center employees with an accuracy of 78%, compared to 58% accuracy when using generalized models (i.e. training and testing models on different people) [107]. The sample size was 9 subjects, and stress after the call was measured based on the subjects answer to 1 question ("how was the call") on a Likert scale with the endpoints labeled as "extremely good" indicating non-stressful and "extremely bad" indicating very stressful. In another study, Setz et al. (2010) built a classifier to classify EDA signals for cognitive load and stress with accuracy higher than 80% [258]. Healey and Picard (2005) also successfully detected stress form EDA with a 97% accuracy of detecting stressful driving conditions [99]. Besides building predictive models, EDA signals have been used to develop systems aimed at encouraging personal reflection (e.g. [181, 248]).

While EDA is widely used and is available in wrist-worn devices, it has several limitations from usability and signal validity perspectives. The electrodes used for EDA sensing can be uncomfortable for long-term wear, and require maintenance to change worn electrodes (as they degrade with time) or charge the device. Sensor readings in real-life settings are prone to different types of sensor artifacts. For example, sensor electrodes can move, detach from the skin, or change in pressure on the skin, all which can affect the sensor signals. More importantly, physical activity and humidity levels confound EDA readings, although some work has shown that physical activity effects can be modeled and removed from EDA signal (e.g. [10]). Furthermore, some people naturally do not produce adequate EDA responses [225].

**Heart activity**

Two parameters relating to heart activity are widely used for measuring stress: Heart Rate (HR) and Heart Rate Variability (HRV). Heart rate becomes elevated when a person is stressed. HRV provides more information, as it is a measure of the variation in the interbeat intervals (i.e. time between one beat and another, a.k.a R-R interval). Contrary to HR, HRV is inversely related to stress. A low HRV indicates that the body is under stress as the autonomic nervous system is trying to regulate the body. HRV is measured with electrocardiogram (ECG) devices that pick up electrical pulsing from the heart contractions through electrodes attached to the chest. Another way of measuring HRV is through blood activity. Changes in HR and HRV produce fluctuations in blood volume and blood pressure. Therefore, monitoring changes in blood activity is a means of stress measurement. Blood activity can be measured with photoplethysmography (PPG), a low-cost, noninvasive optical technique [92]. A PPG sensor sends an optical pulse through a light emitting diode and has a receiver to capture the reflected light. When light is emitted in an area where blood volume is high (a reaction to stress), more light is absorbed and less light is reflected.

Although heart activity is a momentary physiological reaction, studies have shown relationships between HRV and self-reported perceived acute and chronic stress in clinical settings [69], a relationship which has not been strongly supported for other physiological signals [115, 165]. HRV is widely used in research and clinical studies to measure clinical conditions related to the autonomic nervous system such as neuropathy, heart conditions (see [279] for a review), and stress [41, 1, 237]. In HCI research, HRV has been used as a measure of stress in studies investigating different factors influencing daily stress, such as ICT usage for college students [178] and email use for information workers [177] using a wearable chest-strap that accurately captures HR and HRV.

A shortcoming of heart activity from a signal validity perspective is sensitivity to respiratory influences. Choi et al. (2010) tried to address this issue by building a linear model to predict the effect of breathing on HRV, then subtracting that effect from HRV reading to get a better assessment of mental stress [47]. Another effort to address the confound of physical activity and missing data is by Sarker et al (2016), who estimated the recovery time for physiological signals after physical activity to remove it from the model [251]. Another limitation of signal validity in real-life settings is that body posture affects the signal [297]. For example, HR is usually higher when a person is standing as compared to sitting. From a usability perspective, sensors in chest straps provide the most accurate signals but could be uncomfortable to wear for long periods, although wrist-worn alternatives are commercially available.

## 2.3 Theoretical frameworks for stress research

This section presents theories and theoretical constructs that have been proposed in the literature to explain and understand stress in general, and workplace stress in particular.

### 2.3.1 Individual differences in the experience of stress

Stress is multifaceted, hence, beside the objective measures of physiology, there are subjective aspects relating to the personal experience of stress. People's experiences and manifestation of stress, whether acute or chronic, vary greatly based on how they evaluate, interpret and cope with stress [165, 287]. Personality factors, job-related factors, and contextual factors can affect whether and how a person might experience stress. For example, individual differences in the experiences of stress have been linked to perception of mastery (i.e. perceived ability to exercise control) [84], personality traits (i.e. extraversion and introversion) associated with responding more or less effectively to happiness strategies [256], and the personality traits of openness to experience and need for personal structure that are associated with whether a person experiences stress as a cost of interruption at work [172]. Another individual factor influencing the experience of stress is gender. For example, a study found end of day (an approximation of sleep time) influences stress differently for males and females, as males who end their day after 2 a.m. have the highest stress the next day, whereas females who ended end their day before midnight have the highest stress the next day [178].

The manifestation of stress as captured through everyday devices also varies with individual differences. For example, Vizer (2013) found that keystroke dynamics and linguistic features of typed text change differently for different individuals in stress and rest conditions, although the study did not report personality or demographic measures associated with this difference [299]. Thomée et al., (2005) found that prolonged use of mobile and computer, and the number of short text messages, was associated with stress for women, but not men [290].

### 2.3.2 Stress, motivation and performance

The biopsychosocial model of challenge and threat (BPS) by Blascovich and Tomaka [32] states that the ratio between resources and demands dictates a person's stress experience. When the ratio is close to balance, the stress response reflects perceiving the situation as challenging, which is associated with positive outcomes such as productivity and engagement. Otherwise, when the gap between resources and demands is greater (i.e. increased stress) the person experiencing stress will perceive the situation as threatening, which is associated with negative outcomes.

The BPS model supports the idea that some level of stress is desirable. The underlying *physiological* mechanism of why some stress is good is the fight-or-flight response which leads to higher arousal and focus in order to deal with life-threatening situations. The underlying *psychological* mechanism, on the other hand, can be linked to motivation. Since stress involves personal appraisal of the imbalance between demands and capacity [31], it has been argued that stress is linked to motivation to meet those demands, as motivation arises as an effort to improve conditions that are less than optimum [315]. However, when stress is too high, a person's perception of their ability to improve conditions diminishes, which affects motivation and performance [315]. This relationship between stress and performance is portrayed by the Yerkes-Dodson law (also known as the inverted-U hypothesis) (Figure 2.2), which shows that increased stress (more specifically, arousal, the physiological reaction to stress) is associated with improved performance up to a certain point, and when stress exceeds that point, performance declines [323, 50]. Therefore, over- or under-arousal reduces task performance. An example of this inverted-U relationship is performance on a driving task under different stress levels. When driving under no stress (little active control needed for driving) or very high stress, driving performance is impaired because of boredom and low alertness, or overload and distraction, but some stress yields safer driving [180, 222].

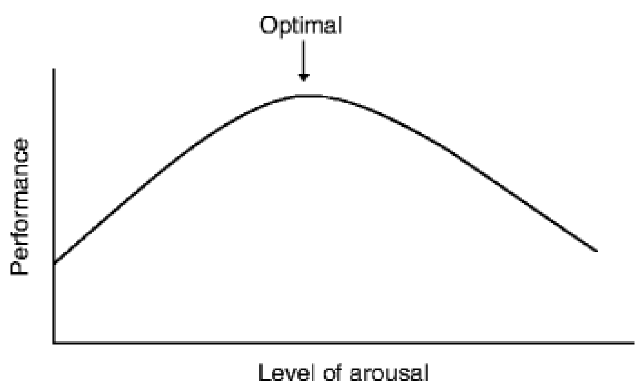



Figure 2.2: The Yerkes–Dodson Law showing the relationship between arousal and performance (Cohen, 2011).

### 2.3.3 Stress and affect

In the psychology literature, stress has been studied as a pathway connecting affect and health [62]. Pressman and Cohen (2005) posit that positive affect can decrease the negative consequences of stress on one's health, and psychological and physiological wellbeing [228].

Moreover, positive affect can reverse the negative effects of stress such as cardiovascular consequences [81]. For example, one study reported that heightened cardiovascular arousal levels returned towards baseline levels more rapidly for people who were shown positive films after fear-inducing films, compared to slower cardiovascular recovery for people who were shown neutral or sad follow-up films [81]. Another study showed similar cardiovascular recovery benefits for inducing positive affect through smiling during stress [144]. Positive affect during stress not only can decrease or reverse negative consequences of stress, but it can also help in building resilience and endurance during stress, which can aid in handling future stressors [144]. These insights into the role of affect in moderating stress consequences can inspire the affective computing and computational mental health communities to go beyond stress measurement and tap into stress moderation through affective technologies.

### 2.3.4 Workplace stress

Workplace stress received special attention over the years. Several models linking the psychology of stress to job-related factors have been proposed to explain stress at the workplace. Aligned with the general definition of stress in psychology, these models view workplace stress as an imbalance between two factors. The Person-Environment Fit (PEF) model suggests that job stress occurs when there is a mismatch between the person's abilities and the job's demands, or a persons' aspirations and the job's resources [71]. Another job stress model is the Job Demand-Control model [126], which explains job stress as an imbalance between job demands and the employees' control over those demands. According to the model, stress occurs when job demands are high and the employees' control (or decision latitude) is low. The demands-control model of job stress has the advantage of not relying solely on the subjective stress perception of the employee, but rather it assesses stress based on reported structural features of the employees' situation in their workplace (i.e. demand and control). Finally, the Effort-Reward Imbalance model postulates that job stress occurs when an employee feels that their job requires high effort and offers low rewards. Similar to the Demands-Control model, the Effort-Reward model assesses structural features of one's job situation, but it also allows for assessing the employees' perception of intrinsic effort (i.e. overcommitment) besides extrinsic effort (i.e. job demands).

## 2.4 Modeling stress in the workplace with unobtrusive sensors

In this section, we review studies of stress monitoring in workplace settings or computer use contexts. To narrow the scope of the review, we consider studies that use physiological signals to detect stress, and exclude studies focused on physical, facial and behavioral signals of stress (e.g., [162, 6, 34, 73, 108, 188, 250]). Studies approximating physiological measures with motion-based sensors such as accelerometers and gyroscopes (e.g., [307, 106, 105]) are also beyond the scope of this review, which is adapted from our previous work [7].

The keywords used to search Google Scholar were: stress tracking, workplace stress sensors, workplace stress tracking, computer stress tracking and computer stress sensors. Seventeen publications in areas spanning human–computer interaction, ubiquitous computing, biomedical informatics, user modeling, multimodal interaction, and affective computing from the years 2006–2017 were included in this review. Tables 2.1–2.3 summarize the reviewed studies based on the sensors and physiological signals, the computer task/stressor involved, the dependent variable (i.e., the stress measure), number of subjects, duration of physiological measurement and whether it is a lab or field study.

Most of the reviewed studies are controlled lab studies where subjects perform a task on the computer while wearing sensors to capture stress. The reviewed studies used computer tasks that simulate workplace computer use scenarios that might lead to stress. The tasks include

Table 2.1: Sensors and signals of the reviewed studies.

| **Publication** | **Sensor: Signal** |
|---|---|
| [87] | Wrist sensor: PPG, EDA, ST |
| [109] | Chest sensors: HR, HRV, BR; Wrist sensors: EDA, ST |
| [139] | Wrist sensor: EDA, ST, acceleration |
| [142] | Chest sensors: HR, HRV; Finger sensor: EDA; Cameras; Kinect 3D |
| [143] | Chest sensors: HR, HRV; Finger sensor: EDA; Cameras ; Kinect 3D |
| [156] | Thermal imaging of the corrugator |
| [166] | PPG: sVRI, blood pressure; ECG: HRV |
| [189] | Digital camera: HR, BR and HRV |
| [200] | Smartphones: audio, physical activity, social interaction; Chest belts: HRV |
| [206] | Pressure sensor; eye-tracker; fingertip sensor: EDA, BVP, HR |
| [257] | Hand sensor: EDA |
| [267] | Necklace sensor: ECG; Fingertip sensor: EDA and ST; Chest sensor: BR. |
| [275] | Chest sensors: HR, HRV; Finger sensor: EDA |
| [319] | Chest belt: ECG and respiration; Hand sensor: EDA; Shoulder electrodes: sEMG |
| [326, 327, 23] | Hand sensor: BVP, EDA, ST; Eye-tracker: PD. |
| [7] | Wristband: HR and EDA; chest-band: ECG (HR), BR; Thermal camera: PP |
| This work | Wristband: HRV |

PPG: Photoplethysmogram, EDA: Electrodermal Activity, ST: Skin Temperature, HR: Heart-Rate, HRV: Heart-Rate Variability, BR: Breathing Rate, sVRI: Stress-Induced Vascular Response Index, ECG: Electrocardiogram, BVP: Blood Volume Pulse, sEMG: Surface Electromyogram, PD: Pupil Diameter, PP: Perinasal Perspiration.

Table 2.2: Computer tasks/stressors of the reviewed studies.

| Publication | Computer Task/Stressor |
|---|---|
| [87] | MIST |
| [109] | Unconstrained work environment |
| [139] | Unconstrained work environment |
| [142] | Writing reports with email interruptions and time pressure |
| [143] | Writing reports with email interruptions and time pressure |
| [156] | CWT and mental arithmetic |
| [166] | Arithmetic problems |
| [189] | Cognitive tasks: ball control task and BCST |
| [200] | Unconstrained environment—in and outside of work |
| [206] | CWT and information pick up task |
| [257] | MIST |
| [267] | CWT; talking about stressful experiences; math test |
| [275] | Writing reports with email interruptions and time pressure |
| [319] | Problem solving, puzzle, and memory task, done under time pressure, social pressure, and distracting noise |
| [326, 327, 23] | CWT |
| [7] | CWT, relaxing video, multitasking, monotasking, essay writing, online presentation |
| This work | Unconstrained work environment |

MIST: The Montreal Imaging Stress Task (mental arithmetic under time and evaluation pressure),
CWT: Stroop Color-Word test, BCST: The Berg Card Sorting Task.

computerized versions of validated stress-inducing tasks such as problem solving, solving puzzles, memory tasks, cognitive tasks, and mental arithmetic. Some tasks are validated stressors (such as the Stroop Color-Word test) while other tasks had additional stressors introduced (such as time pressure or social stress) to create the desired effect. For most studies, sample size ranged from 10 to 35 subjects, but varied in terms of unit of analysis (i.e., hours, sessions). A direct comparison of the results of all the above studies is not possible due to their differences in stress definitions, study design, sensors used, features extracted, and analysis methods.

The most common experimental setting in the reviewed studies was comparing a condition where stress was induced (e.g., by performing a stressful task or introducing social stressors), against another condition where no stress was induced. This approach results in binary classification models where data points are classified into either stress or rest. This classification

Table 2.3: Summary of reviewed studies.

| Publication | Dependent/Output Variable | # Subjects | Duration of measurements | Controlled |
|---|---|---|---|---|
| [87] | STAI-Y | Lab: 21, Field: 5 | Total: 1564 min (lab), 1327 h (field) | Partially |
| [109] | Self-report | 15 | 5 days | No |
| [139] | EDA level | 10 | 4 weeks | No |
| [142] | Self-report | 25 | 3 h | Yes |
| [143] | Self-report | 25 | 3 h | Yes |
| [156] | Difference from baseline | 11 | 12 min | Yes |
| [166] | Physiological measures | 40 | 50 min | Yes |
| [189] | Stress condition | 10 | 10 min | Yes |
| [200] | Self-report | 35 | 4 months | No |
| [206] | Stress condition | 10 | 21 min | Yes |
| [257] | Stress condition | 33 | 4 h | Yes |
| [267] | Stress condition | 20 | 20 min | Yes |
| [275] | Stress condition | 25 | 3 h | Yes |
| [319] | Stress condition | 30 | 40 min | Yes |
| [326, 327, 23] | Stress condition | 32 | 10 min | Yes |
| [7] | Difference from baseline | 61 | 90 min | Yes |
| This work | Duration of difference from baseline | 47 and 51 | 1 week and 3 weeks | No |

STAI: State-Trait Anxiety Inventory. Controlled: Whether data is collected in a controlled lab experiment.

is an oversimplification of workplace stress, as employees are seldom at rest (i.e., doing nothing). Some studies tried to address this limitation by increasing the number of classes (e.g., 'relaxed', 'concentrated', and 'stressed' in [206]) or replacing the 'rest' condition with non-stressful computer work (i.e., 'low cognitive load' vs. 'stress' in [257]). Other than predicting the stress condition, studies have also considered self-reports as ground truth, and used physiological signals as predictive variables (e.g., [87, 200]). Finally, a stress measure that has been used, which captures more variation in stress, is departure from the baseline physiological measure, where stress is said to be detected if the physiological signal during the task is higher than the subject's baseline measure (e.g., [156, 166]).

While many studies measure stress during standardized tasks (such as the Stroop Color-Word test) as a proxy for workplace computer use, Koldijk et al., [142, 143] present a dataset of physiological measures during email interruption and time pressure as simulated workplace stressors, validated by self-reports of mental load. Using this dataset, Sriramprakash et al. [275] were able to build a model discriminating a neutral condition from the email interruption and time pressure condition using heart-rate and skin conductance measures. More work exploring workplace computer use scenarios beyond standardized computerized stressors is needed to account for the variation in workplace activities and the possible challenges for real-time stress monitoring during those activities.

While these studies help advance unobtrusive stress measurement in the workplace, deploying these systems in real-life work scenarios requires a more nuanced understanding of the costs and benefits involved. There is a lack of in-situ studies using these wearable sensors to track stress at the workplace and correlate it with potential stressors, and to understand employees' perspectives on tracking their stress with wearables.

## 2.5 Computer-use factors potentially associated with stress

Tracking stress in the workplace has gained increased attention in HCI, often with the goal of building predictive models from wearable and environmental sensors, or for designing workplace stress interventions. For example, studies have explored whether tracking keystrokes, keyboard pressure, and mouse clicks [108, 140, 300, 299, 328] or tracking posture [141] can predict mood and stress. Physiological sensors have mostly been tested in simulated workplace lab settings (see Section 2.4 for a review). A few studies incorporated stress tracking in real workplace settings to explore patterns and correlates of stress (e.g. [109, 165, 183]). In the below subsections, we present relevant literature that explored computer-use factors potentially associated with stress or stress proxies such as workload and burden.

### 2.5.1 Email

Email is an integral part of everyday work for many in knowledge work professions. The benefits of email at the workplace go beyond providing a means for communication, as we increasingly rely on email for information sharing and archiving, and for assigning and delegating tasks. Despite these benefits, previous work on patterns of email use in the workplace consistently noted the considerable time and attention that email management requires. Email has been found to be associated with stress and burnout due to the time it takes to go through an ever-increasing volume of emails, the task demands associated with emails, and the interruptions they create [239, 22, 177]. A 2016 study [176] found that during business hours, employees spent an average of almost one and a half hours on email per workday, and checked their emails on average 77 times. There was a wide variation among employees in how long and how often they checked email. No difference in average email duration was found between employees who checked email based on external notifications and those who checked email on their own. However, differences in email duration existed between employees who check their emails all at once or a few times (batching) or consistently throughout the day, where the latter group had longer average email duration.

Batching may decrease stress by avoiding disruptions of task activity and reducing cognitive load [292]. However, results are mixed as to whether individual strategies of checking email are related to stress [36, 145, 176]. Email is often managed while multitasking with other work tasks. One might be working on email while having a concurrent task at hand and also anticipating an upcoming deadline or important appointment. Little research has explored this complex workplace dynamic as it relates to stress and performance. In previous work, we modeled this complexity of a real-world work environment in a study that examined the interplay between email use patterns, stressors, and task performance [8]. We looked at stress reactions when participants worked on a task while receiving emails in one of two modes: high interruptions (intermittent email notifications) or low interruptions (getting a batch of emails to process all at once). Participants in the high email interruption mode spent more total time on email as their stress increased, perhaps due to the time it takes to re-focus after being interrupted. Higher stress in the high interruptions mode was associated with a higher use of anger-related words in email responses. We also found that in the low interruption mode (i.e. when participants received a batch of emails to process all at once), stress increased for people who scored high on the personality trait of neuroticism [8]. A potential explanation for this is that neurotics are more susceptible to stress in general, and since handling emails in a batch requires a more sustained focus duration than addressing emails intermittently with breaks in between, it could be that this sustained focus causes stress. These individual differences should not be overlooked when examining computer use and stress, as they might explain conflicting results from prior work on email stress.

### 2.5.2 Attention switching

Computer work often involves managing multiple tasks. Multitasking could increase workload and stress [318, 178, 236] as well as cause errors in the tasks performed [216, 193]. As multitasking requires frequent attention switching, it can causes cognitive burden and feelings of inefficiency and diminished productivity [170]. Attention switching can also result from interruptions such as receiving an email notification, a phone call or a face-to-face interaction.

Studies in HCI have repeatedly shown the prevalence of multitasking, interruptions, and attention switching for information workers during computer work [88, 172, 171]. While some interruptions can be beneficial for providing important information or social interaction [116, 209], they can also be detrimental by affecting productivity, lengthening the time to resume tasks, and causing errors [77, 154, 216, 292]. Interruptions also affect mood and stress as continual switching of attention increases cognitive workload and consequently stress [192].

### 2.5.3 After-hours work connectivity

Many working individuals continue to be connected with their workplace tasks outside of formal work hours, either by staying longer in the office to finish tasks, or by accessing workplace systems through portable devices such as mobile phones or laptops during commuting or at home, before or after formal work hours. The use of communications technologies after-hours to perform job-related functions has been widely investigated for many job roles (e.g. [35, 78, 185, 241]). Not surprisingly, after-hours work connectivity is associated with work-life conflict, distress, and sleep problems, but these relationships are moderated by perceived job autonomy and control [35, 253].

Inbox management is one of the top work activities that people engage in outside of work hours and on mobile devices [185, 241]. One study that examined email use after formal work hours found that time spent on email after hours, as well as organizational expectations regarding monitoring work emails after hours, led to emotional exhaustion and a negative perception of work-life balance [29]. Another study based on interviews with office employees reported that participants view incoming messages less than an hour after they are received during non-work hours through their mobile phones [185].

A more detailed analysis of work connectivity after hours can be achieved with log data analysis, rather than self-reports, to uncover daily patterns and to characterize users based on these patterns.

## 2.6 Physician stress related to EHR systems and EHR inbox

An important working population affected by computer-related stress at the workplace is physicians. Studies have noted the burden of EHR digital work for physicians [54, 259, 85]. EHR-related factors that could lead to physician stress and burnout include the extra time needed, often beyond work hours, to complete EHR-related work [16, 247, 17, 4], usability issues [101, 295, 129], risks associated with errors [218] and taking time from face to face interaction with patients [45].

With EHR systems being accessible through laptops and mobile phones, this creates the possibility of constant connectivity for physicians to their work tasks. Concerns have been raised regarding EHR use extending beyond usual work hours for tasks such as completing patient notes and placing work orders [16, 247]. In this dissertation, we extend those findings to investigate physicians' email use patterns after work hours and their association with stress.

Physicians have increasing inbox management demands. Email has been introduced into physicians' work relatively recently as part of advanced EHRs used in providing clinical care to patients. The EHR inbox is being increasingly integrated into physicians work [3, 63, 260]. Tasks that used to be done through other means like paper and face-to-face communication are now integrated into the EHR inbox. These tasks include communication with patients, receiving lab results and approving medication refills. Inbox integration has advantages making these tasks more streamlined and automated, as well as enhancing access to physicians and building relationships with patients and families [159, 234, 233]. Patients are also increasingly adopting secure messaging to communicate with physicians. In 2015, 64% of physicians had an EHR with the capability to exchange secure messages with patients, an over 50% increase from 2013 [288].

A 2017 study [16] using EHR logs found that time in the inbox accounted for 24% of total EHR time, and of the time spent in the inbox, a larger proportion was spent after-hours compared with the time spent on other EHR activities. Another study reported that 86% of surveyed physicians worked outside of work hours to respond to inbox messages [266]. Besides the time it takes within and outside of work hours, inbox-related burden has been attributed to the volume and source of EHR messages [284, 93], and information overload from notifications (aka asynchronous alerts) [265]. Although some studies quantified EHR inbox–related factors and measured self-reported workload, well-being, or burnout at a single time point [202, 203], they did not measure daily stress associated with EHR inbox use.

As with office employees, email has been associated with increased interruptions and burnout for physicians [159, 284]. The nature of physicians' work makes the patterns of interaction with email different from other types of office employees. For physicians, clinic time is dedicated mostly to patient appointments and using EHR functionality related to patient data/orders. This workday structure likely exacerbates the challenge of managing inbox work because of the lack of dedicated or flexible time to manage the inbox, so the patterns of physicians' email use might not be the same as those seen in other office environments where employees have more flexibility and control over when to check their inbox.

Finally, physicians' EHR work could be associated with many switches in attention when processing messages amid patient visits and other clinical responsibilities. Studies in healthcare show that attention switching, both due to endogenous and exogenous factors, is associated with lower performance and higher stress [167, 314], as well as an increase in the likelihood of errors during clinical tasks [317, 316]. We therefore investigate the frequency of window switching during EHR work and its association with stress.

## 2.7 Summary

This chapter has provided an overview of how stress is defined and measured, how workplace stress is conceptualized, and how previous research investigated modeling workplace stress with wearable sensors. The chapter also covered computer-work related factors that are potentially associated with stress for information workers and physicians. Stress is defined as the perceived imbalance in demands and resources, which triggers several psychological and physiological responses. These responses can be measured with self-reports or sensors, although each modality presents some limitations. Frequent self-report surveys might not be suitable for unobtrusive and continuous stress tracking at the workplace, so researchers have explored measuring physiological stress response with unobtrusive wearable sensors. Limited research has been done in real workplace environments using wearable sensors for stress tracking. This presents an opportunity to study factors associated with daily stress at the workplace in more depth and detail than has been previously done with self-reports.

# Chapter 3

# Methods: Unobtrusive Tracking of Stress and Computer Use

The methods used in this dissertation combine unobtrusive sensing of physiological stress through wearable sensors, collecting self-reported perceived stress through Ecological Momentary Assessments, and tracking computer activity through computer activity logging software. These methods were used in two real-world workplace contexts: information (i.e. office) work and medical work. This chapter describes the common aspects of methods used in these two study contexts. The details specific to each study are presented in their respective chapters.

## 3.1 Physiological stress through a wearable sensor

Participants were given a wrist-worn device (Garmin Vivosmart 3, Figure 3.1a) with an optical heart-rate sensor to measure HRV-based stress. The optical sensor (photoplethysmogram, PPG) works by emitting light onto the skin (an area where arteries are close to the skin) and measuring how much light is reflected back. As the heart contracts and pumps blood, the arteries of the body swell slightly and return to normal. When the arteries are slightly swollen, they will absorb more light reflecting less light back to the sensor. Continuously emitting and measuring reflected (or transmitted) light can measure heart activity such as heart rate, heart-rate variability and blood pressure. PPG sensors are commonly used in consumer-grade wrist-wearable devices given their usability and comfort compared to using electrodes.

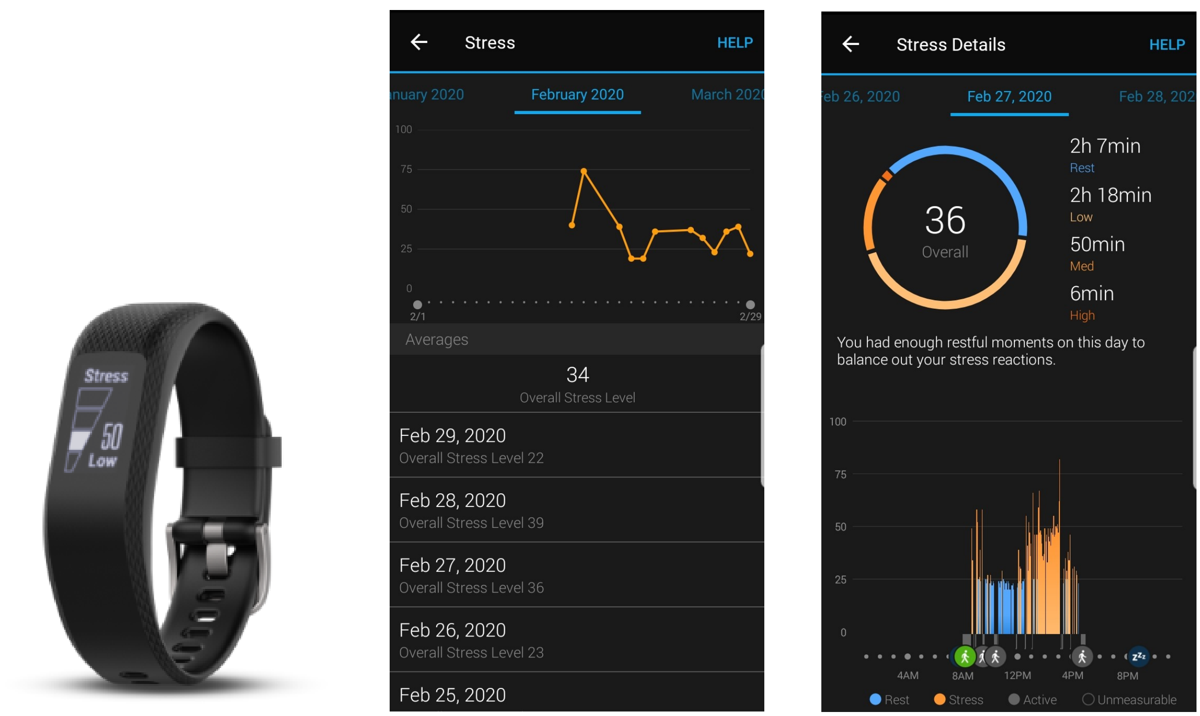
Compared to other physiological stress measures that can be obtained from wearable sensors in daily life, HRV is more reliable in real-world (outside the lab) settings. For example, skin conductance (i.e. electrodermal activity, EDA) can be hard to measure in dry indoor air-conditioned settings as the electrodes rely on sweat to measure conductance. In addition, some people naturally do not produce adequate EDA signal [225]. HRV sensors in wrist-wearable devices are light-based and are more commonly used in consumer-grade wearables.

### 3.1.1 Validation of the Garmin Stress Score

The wearable device we used produces a real-time "stress score" based on HRV in still moments (i.e. excluding times with physical activity that interfere with HRV readings) and accounts for the physiological norm of each user. The method used for HRV analysis in the Garmin wearable uses an algorithm by Firstbeat (Firstbeat Technologies Ltd., Jyväskylä, Finland) which builds a digital model to recognize different states of physiology and their intensity [161]. The intensity of stress calculated based on variables related to sympathetic dominance of the autonomic nervous system. These variables include, for example, high frequency power,low frequency power, respiration rate, and HR. The model takes into account the individual scale of physiological features within a person. Changes in HRV that are known to occur during postural changes, such as standing up, are also differentiated from other factors that influence cardiac activity to more accurately capture stress from HRV. The provided stress score ranges from 0 to 100, with 0-25 indicating rest, 26-50 indicating low stress, 51-75 indicating medium stress, and 75-100 indicating high stress. The Garmin API provides this stress score as 3-minute averages of the real-time stress scores generated on the device.

According to the developers (Firstbeat Technologies), their stress classification method is based on data from thousands of lab assessments and more than 100,000 field assessments. The method has been empirically tested and validated in several settings such as corporate wellness, work ergonomics, healthcare and sports [161]. Stress classification using this method has been shown to correlate with cortisol after awakening and indicators of stress and relaxation during sleep for 17 hospital workers, and significant differences in stress were observed between sleep and awake times, as expected [245]. The method's classification of stress and relaxation was also found to correlate with psychological work-stress-related variables such as work effort (r=-.66 for relaxation %) and daily self-assessments of stress and satisfaction at work (r between .67 and .88 for different HRV-based measures) [294], as well as self-assessment of occupational burnout (standardized beta =.3, p=.001 for the percentage of stress time in a workday) [289]. In another study involving 12 participants over the period of 10 weeks [13], days were divided into "good days" and "bad days" based on a factor analysis of a number of self-assessments of mood. Significant differences in the duration of stress (by the Firstbeat method) were found between good and bad days, where bad days had longer periods of stress (p=.001). Direct correlations between self-reports and HRV stress measures were weak, as only mental strive and busyness were associated with stress duration (r=.16 and r=.13) measured by the firstbeat algorithm [13]. Other self-reported emotions such as anger, anxiety and fatigue were not individually associated with HRV-based stress [13]. While these studies used the Firstbeat stress classification method with different HR monitoring devices, Garmin heart-rate sensors were compared to other devices and found to be among the most accurate in both lab and real-life settings based on a review of 42 studies from 2016 to 2019 [82]. Thus, using the Firstbeat HRV-based Stress Score generated on Garmin devices can give an accurate and continuous measure of stress for this study.

Besides its validity and reliability, we chose to use the Garmin wearable because it does not require daily charging like other validated wearables for HR monitoring (e.g. the Apple Watch). The Garmin device needs to be charged once every 4-5 days, thus significantly reducing participant burden. Participants were able to view their real-time stress data on the device, and view their previously recorded stress measures as timeline charts in the associated mobile app (Figure 3.1).




(a) (b) (c)

Figure 3.1: The display of tracked stress scores on the wearable device and its app.

In our analyses, the HRV-based stress measure is the number of minutes in medium and high stress (stress score >50) during the workday. We excluded low stress periods (scores from 25 to 50) because a certain amount of stress (or arousal) is expected and needed for performing daily tasks [32]. Low stress during the workday could reflect a healthy autonomic nervous system that adapts to the daily demands of the job. We chose the outcome measure of mid-high stress duration rather than stress intensity (i.e. the level of departure from the physiological norm, or how 'high' stress is) because stress patterns might be obscured when averaging stress and relaxation scores throughout the workday. An average daily stress score will not tell us how long the stressor lasted or show how stress fluctuated. Very different patterns of stress can produce the same average stress score.

There were some gaps in the continuous HRV-stress data. Missing HRV-stress data could be attributed to loose-fitting of the sensors on the wrist, removing the device for charging or forgetting to wear the device, or physical activity. We set a minimum of 20 minutes of HRV data per hour for hourly stress measures and two hours of data for daily measures to be included in the analyses. We further report the number of valid minutes of data each reported stress measures is based on.

HRV is affected by a number of factors other than stress, such as physical activity and overall health. Thus, HRV as a measure of stress is most reliable for healthy participants in sedentary settings. Previous studies used HRV from wearable devices as a measure of stress in office settings where participants were working on a computer [212, 176, 141, 166], making this method applicable to computer-based work by information workers and physicians.

### 3.1.2 Technical setup for data collection and storage

HRV data was unobtrusively and continuously collected and uploaded from the wearable device, to the participant's phone, to our servers. This mobile sensing system is a modified version of the StudentLife data collection system [309] which has been used in recent mobile sensing projects [183]. Participants were asked to install the Gamin Connect app, the official app of Garmin devices. After creating a Garmin account and logging in, participants were asked to enter their credentials in a portal in order for their Garmin app data to be transmitted through the Garmin API to a designated server at the University of California, Irvine, School of Information and Computer Sciences. The data from the Garmin app included processed data from the wearable device, such as stress scores calculated from HRV, heart-rate and steps. Participants were also asked to install a mobile app (Tesserae Phone Agent) which streams data from the wearable device to the participant's mobile phone via Bluetooth and uploads it to UCI servers (via Notre Dame servers, as the developers of the app) in JSON format whenever the phone is connected to WiFi. Participants downloaded the Tesserae app from a link sent to them by email along with their login credentials (a study username and password that we provide), and paired their wearable device with the app to start data collection. The data from the Tesserae app included more fine-grained data than the Garmin app data, such as continuous RR intervals and HR. However, since the data was raw and did not exclude physical activity or postural changes, and was not normalized per person, it was not used for this study and the Garmin processed HRV-based stress scores were used.

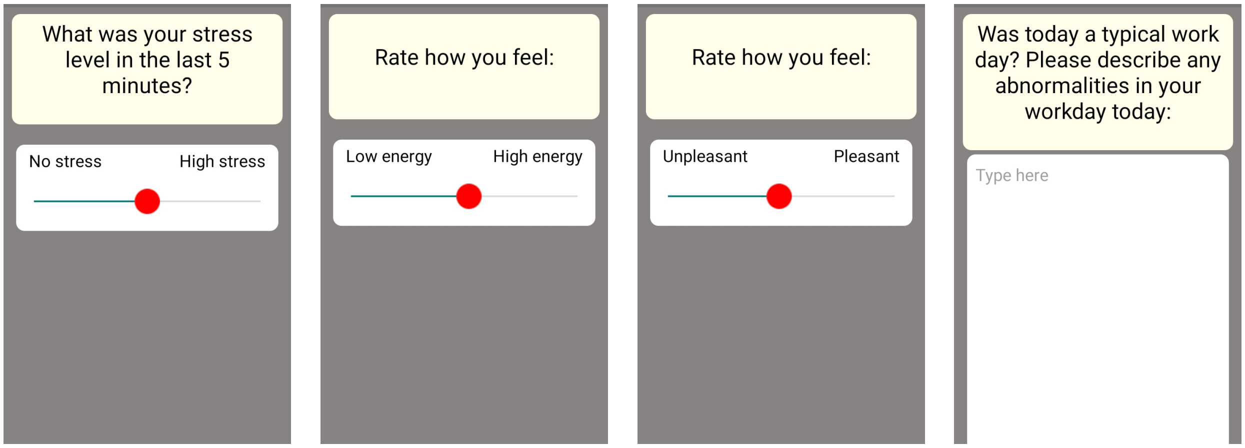
A designated server at the University of California, Irvine, School of Information and Computer Sciences was set up to receive and store the wearable device data. An automated script (cron job) periodically deleted extra files (e.g. data for troubleshooting WiFi connectivity) to avoid exceeding the server's storage space and to accelerate data backup and processing for the analyses. The server is password-protected with two layers of authentication needed to access it. Datasets are de-identified and do not contain personally identifying data such as names, phone ID, emails or phone numbers. Data is only linked to the wearable device ID and the provided participant ID. Datasets were backed up in an encrypted external hard drive stored in a locked cabinet in a locked office room.

## 3.2 Perceived stress through Ecological Momentary Assessments

Participants logged Ecological Momentary Assessment (EMA) of stress three times a day using an app (PIEL Survey [122]). The app was preconfigured to send notifications prompting participants to complete a short survey on their phones in the morning (at a random time between 9:30am and 10:30am), at mid-workday (between 1pm and 1:30pm), and in the afternoon (between 3pm and 4pm). If the notification was not opened within 45 minutes, the survey expired. The survey consisted of questions asking participants to rate their stress in the last 5 minutes on a sliding scale from no stress to high stress (Figure 3.2). To identify affect more precisely, we also asked participants to report their arousal level (from Low energy to High energy); and their valence (from Unpleasant to Pleasant) according to Russell's circumplex model (see Chapter 2) [246]. We used a sliding scale as previous studies indicated participants often want more granular options than a 5-point Likert scale [2]. The 5-minute window allows for correlating the reported stress with work activities that the participant was doing before answering the survey, rather than merely reflecting stress at the moment of taking the survey, which could be after a work activity has ended and its associated stress changed. In addition to these questions, the information workers' study included an additional free-response question on whether this is a typical day and asked participants to describe any abnormalities in their workday.

## 3.3 Computer activity logging

For each study context, we tracked computer activity to collect data including continuous timestamped logs of pages visited and actions performed. For office employees, we used a computer activity logging software (customized KidLogger, SafeJKA S.R.L.) and for physi-




(a) (b) (c) (d)

Figure 3.2: The mobile EMA questions and interface.

cians we used EHR system access logs provided by the medical group. From these logs, we created variables to quantify how time was attributed to different activities and characterize computer or EHR use patterns. Details about the logging process and data generated for each study is presented in the respective chapters.

We used computer activity logging instead of other methods like diaries or observations that have been used in previous studies. Computer activity logging is unobtrusive and provides granular data on the real use of the computer, with no recall bias or observer influence. Another major advantage of computer logging is the ability to automate the process of creating variables of interest to track computer use factors related to stress in real-time. These tracked variables can be fed into smart applications and real-time interventions for stress management.

## 3.4 Security and privacy

Security and privacy were a top priority throughout the entire data collection, storage, and data analysis cycle. The data is de-identified (i.e. no direct link between the data and the name of the subject) and stored on a secure UCI server and an encrypted hard drive. The data per participant is linked to a participant ID, and all names, emails and other personally identifiable information are unlinked. All analyses were conducted only on de-identified data. The information workers' study was approved by UCI's institutional review board, and the physicians' study was approved by the institutional review board of Kaiser Permanente Northern California.

In accordance with IRB (Institutional Review Board) requirements for research involving human subjects, the study protocol ensured that all participants were fully informed of the details of the study, had the right to withdraw from the study at any time, and if they choose, to have their data erased without any repercussions. Participation was fully voluntary. The study team was available to respond to any participant who had concerns or needed assistance at any point during the study.

# Chapter 4

# Information Workers' Stress and Computer Interaction

## 4.1 Introduction

Capturing stress levels in the workplace in real-time can help uncover information about related factors and behaviors that will advance and broaden our understanding of workplace stress. However, while many studies have tested methods for technology-supported stress tracking, only a few used these methods in real-life workplaces to identify everyday computer work factors associated with stress. In this chapter, we present a study where we unobtrusively monitored stress levels for office workers and identified everyday computer interactions related to daily stress. As described in Chapter 3: Methods, participants were asked to wear a wrist band (Garmin commercial activity tracker) and install an application on their phone to capture data from the band as well as an app that captured daily self-reported stress through short surveys multiple times a day. To capture work-related variables, participants were asked to install a computer-activity logging software and fill out one-time surveys.

Scholars have identified several ways in which computer-based work can lead to fatigue, burnout, overload, and stress [286, 49] (see Chapter 2.5). The use of information and communication technologies (ICTs) is linked with working more and working longer [20]. For example, given the ease and low cost, people report handling more communications and more sorting and filing of messages with ICTs [30].

Related to increasing the amount and duration of work, the feeling of constant connectivity is a frequently cited source of ICT-related stress [304, 184]. Having continuous access to the work email, documents, and systems removes temporal and spatial barriers to work [241], but also makes employees feel as they are always "on call" [286]. This constant connectivity has also created a sense of urgency to complete tasks and respond to work commutations as everything is "instant" [136]. Work-connectivity outside of work hours, whether by staying late in the office or taking work home has been associated with work-life conflict, distress, and sleep problems [35, 253]. It is unclear, however, how working outside of work hours affects *daily* stress. That is, do employees experience more stress on days when they work outside of their typical work hours? Previous research reported overall stress with overall reflective measures of employees' self-assessed overtime work. Capturing objective daily measures will extend and complement previous work by offering real-time assessments that can inform more timely decisions and interventions for stress management.

Computer work is characterized by frequent interruptions, which also extend work time by taking time away from the main task and creating attention residue that extends the time needed to refocus and resume the main task [154]. Interruptions or lack of focus within computer work (e.g. receiving an email notification or being distracted by social media) can be reflected in computer window switching [174]. Previous work has quantified interruptions in the workplace [171], and established a correlation with perceived stress in controlled experiments [172], but this correlation has not been empirically tested in real-workplace settings using objective measures.

Email has received increasing attention as a source of workplace stress [22]. As discussed in Chapter 2.5.1, email can be related to stress as it is associated with task demands and can build up in employees' inboxes creating a sense of overload [239, 22, 177]. Empirical studies supported the proposition that email is associated with stress [176]. Replicating these findings in other work settings with a larger sample and a longer observation period, as well as accounting for time spent on other work and non-work computer activities, can help to assess the generalizability of these results.

Overall, studies of computer-related stress at the workplace were mostly based on surveys where employees self-report their overall behaviors (e.g. how often they use email) and feelings (e.g. overload, fatigue). More recently, a few studies tried to link the day-to-day workplace computer use with employee productivity and mood [172, 176]. I extend previous work on the relationship between day-to-day computer use and stress by simultaneously examining a number of computer-use factors in a real workplace environment, accounting for individual factors, and using an objective measure of stress that addresses biases in self-reports (see section 2.2.1). I address the following research question:

RQ1: What computer use factors are associated with daily stress at the workplace? How do individual differences affect those factors?

With objective computer usage measures, novel metrics that are otherwise hard to quantify can be constructed. One such measure is regularity. Research in psychology has shown that regularity of daily routines is related to wellbeing and feelings of security [19]. Regularity of work routines relates to mental health and wellbeing at work, as routines help employees deal with demanding aspects of their jobs under stable circumstances [211, 21]. A study found that variability in work schedules is more strongly associated with psychological distress, poor sleep quality, and unhappiness than low wages are associated with these outcomes [255]. However, no study to date has quantified the variability of workload as reflected in computer work. In HCI and ubiquitous computing studies, wearable and mobile sensors have been used to construct measure of daily routines and regularity of location, physical activity, ambient sound and phone usage, and significant associations were found between these daily routine measures and users' mental health [321, 18, 147, 231, 309] and personality traits [310]. Variability in computer usage at the workplace could reflect changing, unpredictable or unstable work routines and demands, which could relate to feeling of instability or stress. I therefore construct regularity/variability measures from computer log data to quantify the regularity of computer workload and its association with stress, addressing the following question:

RQ2: How does variability in computer work patterns affect stress at the workplace?

While the first two RQs deal with computer use factors and their independent associations with stress, the next RQ investigates how much monitoring computer use altogether can tell us about employees' daily stress. In order to design systems that infer user stress from computer use, we need to know how much information a collection of computer use factors can reveal about stress. Research has explored ways to detect stress from computer peripherals such as keyboard and mouse [108, 300, 65], but predicting stress form computer activity and use patterns remain largely unexplored. Although scattered research has found associations between individual computer use factors (e.g. email, window switching) and stress (see section 2.5), no research has evaluated whether comprehensively monitoring computer use at the workplace can predict stress to an extent that allows building stress prediction systems from computer use data. If reliable predictions of stress can be made from computer activity tracking, future applications could eliminate the need to use wearables to monitor stress, and can rely on monitoring stress from "behavioral sensors" of computer use. I thus address the following question:

RQ3: To what extent does unobtrusive monitoring of workplace computer use help identify daily stress?

## 4.2 Methods

### 4.2.1 Recruitment

We recruited employees from the University of California, Irvine in January and February, 2020. An email was sent to all university employees via the all-employees mailing list. The email contained a flyer (Appendix A) asking for participants for a research study on workplace stress. A screening survey (Appendix B) was included in the email for interested subjects to fill out. The study was also advertised in the university's wellbeing newsletter. Within five days of sending the email, 663 responses were submitted through the screener survey. Eligibility criteria included being an office-based employee with access to a work computer with a Windows operating system and a smartphone with internet access, who self-reported that computer-based work constitute most or all of their workday. Participants also had to not be enrolled in another research study about managing stress. Employees taking cardiac medication, using pacemakers or implantable cardiac defibrillators, those previously diagnosed with atrial or ventricular arrhythmias, and those with a BMI over 30 were not eligible, as these factors have the potential to interfere with the HRV-based stress measures obtained from the wearable device. Due to privacy and security measures, employees from some departments were excluded from participation in the study.

We enrolled 51 eligible employees. Employees included junior and senior personnel in academic and non-academic departments such as human resources, information technology, finance, administration, student affairs and the housing office. Workplaces included closed offices as well as open spaces and cubicles. Table 4.1 shows participant demographics and job information. The majority (76%) of the sample were females, which is higher than the overall university's percentage of female staff (64%)[1]. The sample had a range of reported overall perceived stress. From the Perceived Stress Scale [52], which measures overall stress,

[1]https://www.oir.uci.edu/files/empl/VIA08NH-all-employees-by-gender.pdf

participant scores ranged from 4 to 28, making this sample suitable for assessing diverse perspectives on stress tracking. Most participants had moderate overall stress (30) and low stress (19), with one participant reporting high overall perceived stress. The Depression, Anxiety, and Stress Scale (DASS) [221] showed that 49 participants did not have severe scores on these measures which could have affected their daily stress experience.

| | |
|---|---|
| Sex | 39 Female, 12 male |
| Age | 22-63, mean 41.49, SD 12.78, median 40 |
| Education | Post-Graduate (28), 4-year college (18), Some college (3), 2-year college (1), high-school (1) |
| Job title | Analyst (13), Administrative staff (11), Director/Manager (13), Specialist/Advisor (7), Researcher (4), Software Developer (3) |
| Typical work hours | Start 7am-9am, end 4pm-6pm |
| PSS | low-stress (19), moderate stress (31), high stress (1) |
| DASS | Depression: normal (44), mild (4), moderate (2), severe (1)<br>Anxiety: normal (32), mild (7), moderate (9), severe (2), extremely severe (1)<br>Stress: normal (47), mild (2), severe (2) |

Table 4.1: Summary information about participants.

### 4.2.2 Study protocol

At the beginning of the study, I met with each participant at their office to go over study procedures, answer any questions about the study, have them sign the consent form, hand over the wearable device (Garmin Vivosmart 3) and install the computer activity logging application and the smartphone applications. I configured two mobile apps associated with the wearable device (Garmin Connect and Tesserae Phone Agent [183]) which streamed data from the wearable device via Bluetooth and uploaded the data to a server. A third app was used for EMAs [122], sending short questions at specified times (see Chapter 3 for details).

Participants were asked to wear the device during work hours and respond to the daily experience sampling prompts for 3 weeks. Participants who did not complete 3 consecutive weeks of data collection due to holidays or travel extended their participation in the study to compensate for missed workdays. Participants were told they could explore the wearable and its app however they like, but they were not given specific instructions on how to use the device to understand their stress. At the end of the data collection period, the lead researcher met with participants at their workplace for an exit interview and to uninstall the computer application and mobile apps. Participants were also asked to complete a battery of surveys on overall perceived stress, workplace stress and personality. Participants were each given a $50 Amazon gift card upon completion of all study procedures. The study protocol was approved by the institutional review board of the University of California-Irvine (UCI) and software installation was cleared by UCI's Office of Information Technology and approved by participants' respective IT departments.

### 4.2.3 Computer activity logging

As described in Chapter 3: Methods, computer activity was logged to collect data on pages visited and actions performed. An open-source software (Kidlogger for research, SafeJKA S.R.L.) was customized for computer activity data collection. The software logged timestamped window switches and the application name (e.g. Excel, Outlook, Firefox) as well as domain names of website visited (not the full URL; e.g. facebook.com, google.com) and inactive (i.e. idle) time. The Kidlogger software was chosen due to the granularity of logs, the possibility of modifying privacy settings (e.g. temporarily pausing logging, turning off logging full URLs, storing data locally), the availability of open-source code for research, and our past experience with the software for a previous research project.

I collaborated with the developers of KidLogger from February 2019 to June 2019 to modify and test the Kidlogger tool in order to enhance privacy options. Specifically, we have added an option to only capture domain names of visited websites (e.g. google.com) rather than full URLs (e.g. google.com/searchPhrase). We also disabled other software features such as tracking keystrokes and logging webpage/document window titles. The logs are stored in an html format which I then converted to csv for processing and analysis.

For additional privacy and security, I decided to store the data locally on the participant's computer during the data collection period, giving them the option to view their logs at any time. The data was collected from the participants' computer onto an encrypted USB drive after the data collection period has concluded. Thus, the computer logs of participants are never transmitted online and are only stored and processed locally.

### 4.2.4 Measures

#### Active computer time

The duration of each computer activity was calculated by subtracting consequent timestamps. After reviewing data samples, I removed any activity with duration longer than an hour as it is likely idle time. It is unlikely that a user will work on a single computer window for longer than an hour without switching even momentarily. Multiple approaches were tested for setting the threshold for inactive time. Standard deviation (SD) and mean absolute deviation (MAD) were not suitable approaches given the skewness of the data (excluding activities longer than 2 MADs away from the mean activity duration would have removed any activity over 6 minutes). Setting a threshold based on the 99.9th quantile also sets a low threshold of 19 minutes, which would exclude many valid logs and will affect the calculation of the total time on the the computer. I therefore set a threshold of 1 hour, which excluded less than 0.01% of the data. The first and last day of the study for each participant were removed from the analysis as they have partial data due to software installation/removal. Days with no time spent on the work computer were also removed as they could be days off or outlier days that do not represent a typical workday. Daily and hourly duration of time spent on the work computer were calculated by summing the durations of user actions and excluding idle times.

### Computer activity type

There were 8,915 unique computer applications and URLs visited. To classify user activities, a 2-step approach was followed. First, a keyword search was used to find variations of common applications and websites. For example, searching 'facebook' returned www.facebook.com, apps.facebook.com and business.facebook.com which were all classified as social media. Any application/URL with the word 'mail' or 'outlook' was classified as email (e.g. mail.google.com, hotmail.com, mailchimp.com, webmail.uci.edu, outlook.office.com, outlook.exe, outlook.com, outlook.office365.com). Any URL including '.edu' was classified as work-related. Second, the remainder of the top 500 most visited applications/URLs that were not classified in step 1 were manually classified. Websites related to news, music, shopping, social media and sports were classified as non-work. Applications and websites related to documents, spreadsheets, presentations, programming and communications (e.g. Teams, Zoom, Skype) were classified as productivity applications. This approach resulted in classifying 2,243 (25%) of all unique applications and URLs, which covered 94% of all application/URL visits. The categories of email, productivity and non-work covered, on average, 81% (SD 10%) of each employees' computer activity.

### Computer work strategies and patterns

To capture email work patterns, we classified daily email checking patterns into batching and consistent. Workdays with dedicated blocks of time for inbox work were classified as batching email, while workdays where employees consistently checked their inbox throughout the day were classified as consistent email checking. With consistent inbox checking, a uniform distribution of inbox duration over the day would typically be observed, while batching would show 2-3 daily peaks of high inbox work duration [176]. We defined days with inbox work batching as days where 70% or more of the total inbox work duration for that day occurred in three separate blocks of time or less. Figure 4.1 shows samples from the data that illustrate the distribution of time spent on email throughout the workday.

The pattern of window switching was measured by counting the number of switches from one computer activity (i.e. application or URL) to another. A minimum threshold of 2 seconds was set to consider an activity change as a window switch. For example, if an employee switched from viewing a document to viewing a spreadsheet for 1 second then switched to view a URL for a minute, this would be counted as 1 switch (from the document to the URL), as the short switch to the spreadsheet is unlikely to be significant enough to constitute a cognitive switch (the employee could have clicked on the spreadsheet by mistake, for example). The 2 seconds threshold is a heuristic based on the 25th percentile of activity duration.

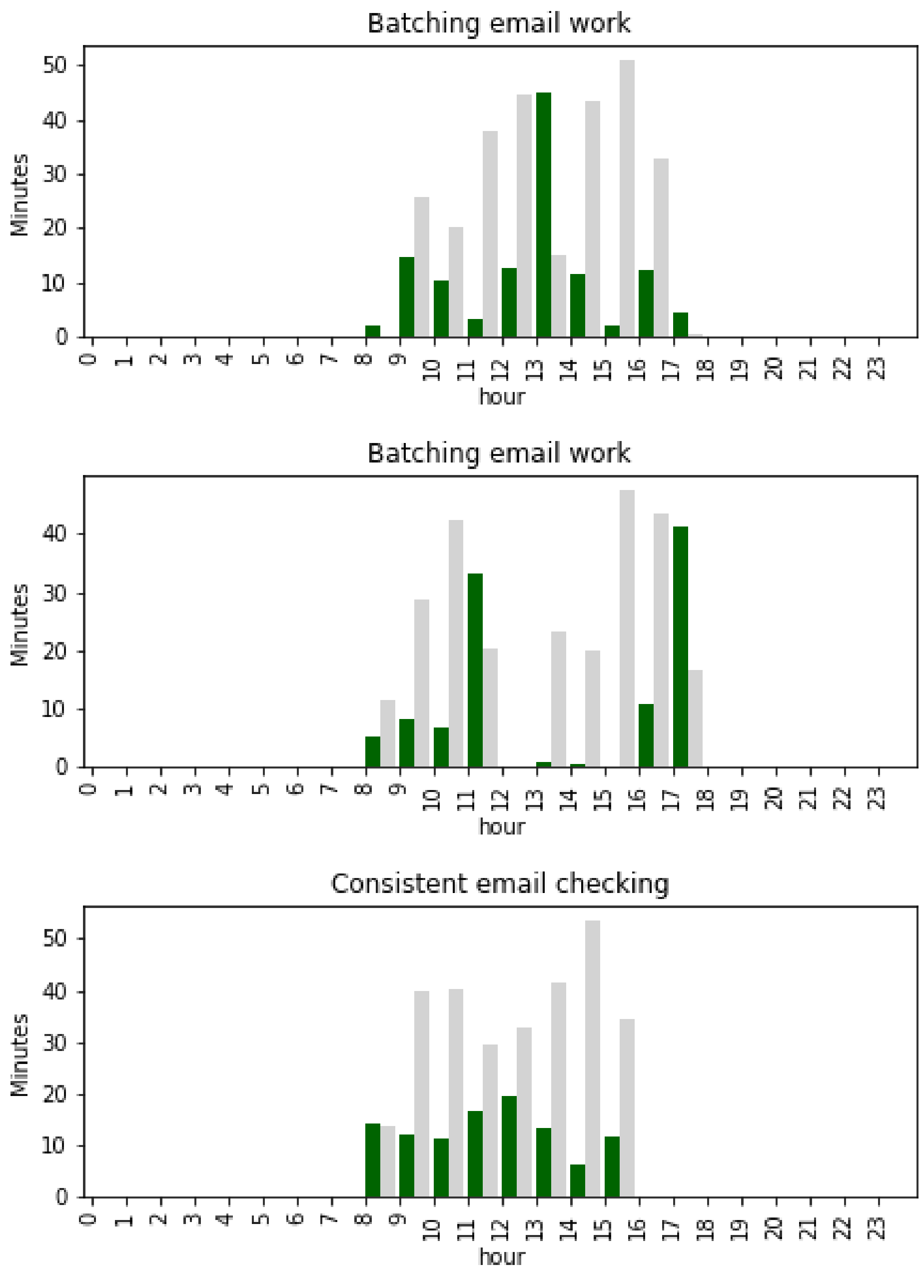



Figure 4.1: Samples of workdays where email work was batched or consistent. Green bars represent the duration on email work and the gray bars represent other computer work.

### Regularity metrics

We used three regularity/variability metrics to measure all-day and hour-by-hour differences in workload across workdays. To capture variability in overall daily computer work duration, we used the measure of standard deviation (SD) of all-day computer work duration. To capture hour-by-hour similarity between workdays, we used the regularity index (RI) and the flexible regularity index (FRI) [310]. All three measures were computed per person to capture within-person variability.

The regularity index (RI) measures the difference between the same hours across two different days. The hourly computer work duration is first rescaled for each participant to [-0.5,0.5]. If the original values of computer work duration across two days are close, the product of their rescaled values would be positive, and negative if the original values are further apart. Equation 4.1 defines the regularity between day a and day b:

$$\forall(a,b) \in S, \quad RI_{(a,b)} = \sum_{t-1}^{T} f(x_t^a) f(x_t^b)/T \tag{4.1}$$

For each pair of day $a$ and day $b$ in the participant's set of all two-day pairs $S$, the regularity between day $a$ and day $b$ is the mean of the product of their rescaled values f(x) for each hour $t$ in all workhours $T$. Workhours are hours of the day that had computer activity on any day during the study. For example, if a participant never worked from 1-4 AM during the study period, then 1-4 AM are removed from $T$ to avoid over-estimating regularity. A higher RI score indicates more similar days, hour-by-hour. The average and range (i.e. difference between the most similar days and the most different) of RI are computed per person. A higher RI mean indicates a higher regularity, while a higher RI range indicates lower regularity. A low RI range indicates that each day is approximately equally different from other days, which could indicate that the days are similar, or that it is typical for a person's workdays not to be similar, which indicates higher regularity.

The flexible regularity index (FRI) also measures hour-by-hour differences between days, but it allows more flexibility than RI by slightly shifting the hours being compared and evaluating how many changes are needed to make the two days similar, hour by hour. The FRI is based on the weighted Levenshtein distance [155], which measures the difference between two strings by counting the number and type of operations needed (i.e. add, remove, substitute characters in the string) to transform one string to the other. A lower distance between two strings indicates more similar strings. Following the approach of [310], we compute FRI for computer work duration by transforming a day's hourly data to a string. Each hour is labeled as 'a', 'b' or 'c', with 'a' indicating that computer work duration in this hour is under the 25 percentile of all data from this participant, 'c' for data over the 75th percentile, and 'b' in between. The weights for each operation to calculate the Levenshtein distance are as follows: 1 for insertion, 1 for removal, 0.5 for substitution if letter are adjacent (e.g. 'a' and 'b'), and 1.5 for substitution if letter are not adjacent (i.e. 'a' and 'c'). We compute the FRI for each pair of workdays per participant, then calculate the average and range of FRI. Both FRI average and range indicate variability (i.e. the higher these measures, the higher the variability of hour-by-hour computer work duration across workdays).

### 4.2.5 Analysis

We investigated the relationship between daily computer interactions and stress through a generalized linear mixed model with employees as random effects. In the first model, the dependent variable was the duration of HRV-based stress during workhours (hours with computer activity). Hours with less than 20 minutes of valid HRV data were excluded from the daily stress duration calculation and days with less than 2 hours of valid HRV data were excluded to avoid overestimating stress duration percentage of the workday. A Poisson distribution was used to represent stress minutes as events within the observation period (i.e. valid HRV minutes as an offset in the model). Adding an offset in the model accounts for the fact that the higher the observation period (i.e. more valid HRV minutes captured), the more likely that stress will be observed. The distribution of the dependent variable (stress duration) was right skewed, as expected in a Poisson distribution. One participant encountered a technical issue causing data loss for their wearable device data and was excluded from the analysis. Another participant was excluded as their wearable device continuously measured abnormally high stress levels and later stopped working, indicating that the device might have been faulty and the recorded data were inaccurate. In the second model, the dependent variable was the daily average of self-reported stress from EMAs. Independent variables were centered (i.e. mean subtracted) and scaled. The variance inflation factor (VIF) was under 5 for all independent variables, indicating multicollinearity was not a problem. Independent variables included the duration of computer activity, the percentage of computer work after-hours to all-day computer time, computer work patterns and time spent on different computer activities. Employee's age, sex and education were included as controls.

To evaluate the model, three models were compared with an ANOVA likelihood ratio test (Table 4.2). The base model included only the random effects, demographics, and the offset, and explained 10.9% of the variance in workday HRV-based stress duration. The second model added to the base model the day of the week, which only increased the marginal $R^2$ to 11%. The final model is the full model with computer-work variables, which resulted in a marginal $R^2$ of 14%. The ANOVA likelihood ratio test showed that the full model is a better fit and explains more of the variance in HRV-based stress than the other two models ($p < .001$).

## 4.3 Results

### 4.3.1 Daily stress

Employees experienced medium to high stress for 22.11% (SD 17.02%) of the workday (56.63 out of 254.11 minutes with valid HRV data, on average). Self-reported stress through EMAs indicated an average stress level of 41% (SD 14%). The average participant had 37% (SD 21%) of their submitted EMAs indicating a stress level over 50%, the midpoint of the slider scale. Participants classified 44% (SD 27%) of their workdays as relaxed, 47% (SD 22%) as moderately stressful, and 8% (SD 12%) as highly stressful.

The generalized linear mixed model (Table 4.2) showed that computer work factors explain 14% of the variation in daily HRV-based stress. The same factors could only explain 8% of self-reported stress through EMAs (Table 4.3). The following subsections explain the significant factors in these models.

### 4.3.2 Duration of computer work

As shown in Table 4.4, on workdays, the average active time spent on computer work was 4:37 hours (SD 1:12), of which a mean of 14:35 minutes (5%; SD 18:13 minutes; median 08:21) occurred outside of typical work hours (typical work hours were self-reported by participants). Average computer work duration was not correlated with demographics, self-reported job stress or overall perceived stress from the one-time surveys.

The generalized linear mixed model (Table 4.2) showed that daily time spent on the work computer and the percentage of computer work outside of work hours were inversely related to the duration of physiological stress during work hours ($p<.001$). The interaction between the proportion of computer work done outside of typical work hours and work-life imbalance

| Fixed effects | Model 1 Std $\beta$ | SE | P | Model 2 Std $\beta$ | SE | P | Model 3 Std $\beta$ | SE | P |
|---|---|---|---|---|---|---|---|---|---|
| ***Personal factors*** | | | | | | | | | |
| AGE | -0.265 | 0.146 | 0.069 | -0.263 | 0.146 | 0.072 | -0.212 | 0.156 | 0.174 |
| Female | -0.18 | 0.336 | 0.592 | -0.165 | 0.337 | 0.623 | -0.199 | 0.36 | 0.581 |
| Education | -0.207 | 0.146 | 0.156 | -0.205 | 0.146 | 0.162 | -0.193 | 0.155 | 0.214 |
| ***Day of the week*** | | | | | | | | | |
| dayMon | | | | -0.133 | 0.02 | 0 | -0.1 | 0.021 | <.001 |
| dayTue | | | | -0.057 | 0.018 | 0.001 | -0.066 | 0.019 | <.001 |
| dayWed | | | | 0 | 0.017 | 0.982 | 0.012 | 0.017 | 0.502 |
| dayThu | | | | 0.065 | 0.017 | 0 | 0.087 | 0.018 | <.001 |
| ***Active computer work time*** | | | | | | | | | |
| Computer work duration | | | | | | | -0.169 | 0.009 | <.001 |
| Non-workhours work | | | | | | | -0.065 | 0.008 | <.001 |
| Work-life imbalance | | | | | | | -0.031 | 0.156 | 0.844 |
| Non-workhours computer work * Work-life imbalance | | | | | | | 0.023 | 0.006 | <.001 |
| ***Computer work strategies and patterns*** | | | | | | | | | |
| Batching | | | | | | | -0.275 | 0.016 | <.001 |
| Neuroticism | | | | | | | -0.197 | 0.156 | 0.207 |
| Batching * Neuroticism | | | | | | | 0.07 | 0.014 | <.001 |
| Window switching rate | | | | | | | 0.011 | 0.011 | 0.323 |
| Window switching rate * Neuroticism | | | | | | | 0.135 | 0.011 | <.001 |
| ***Computer activities*** | | | | | | | | | |
| Email duration pct of all computer time | | | | | | | -0.09 | 0.013 | <.001 |
| Productivity apps pct of all computer time | | | | | | | 0.063 | 0.014 | <.001 |
| Non-work apps pct of all computer time | | | | | | | -0.062 | 0.015 | <.001 |
| Pseudo-R2 (fixed effects) | 0.109 | | | 0.11 | | | 0.14 | | |
| Pseudo-R2 (total) | 0.977 | | | 0.977 | | | 0.98 | | |

| | Df | AIC | BIC | logLik | deviance | Chisq | Chi Df | Pr(>Chisq) |
|---|---|---|---|---|---|---|---|---|
| mod1 | 5 | 30355 | 30376 | -15172 | 30345 | | | |
| mod2 | 9 | 30246 | 30285 | -15114 | 30228 | 116.65 | 4 | <2.2e-16 *** |
| mod3 | 21 | 29438 | 29529 | -14698 | 29396 | 831.67 | 12 | <2.2e-16 *** |

The dependent variable is workhours' HRV-based stress duration.
Friday is the reference category for the variable "day of week".
Std $\beta$ is the standardized coefficient.

Table 4.2: Nested model and Likelihood Ratio Test for information worker's model of daily HRV-based stress and computer use.

| Fixed effects | Std $\beta$ | SE $\beta$ | $P$ |
|---|---|---|---|
| **Personal factors** | | | |
| AGE | 0.046 | 0.021 | 0.031 |
| Female | 0.007 | 0.049 | 0.881 |
| Education | 0.031 | 0.02 | 0.133 |
| Neuroticism | 0.014 | 0.021 | 0.497 |
| Work-life imbalance | -0.007 | 0.02 | 0.73 |
| **Active computer work time** | | | |
| Computer work duration | 0.02 | 0.011 | 0.062 |
| Non-workhours work | -0.013 | 0.009 | 0.148 |
| Non-workhours computer work * Work-life imbalance | -0.011 | 0.009 | 0.248 |
| **Computer work strategies and patterns** | | | |
| Batching | 0.026 | 0.019 | 0.17 |
| Batching * Neuroticism | -0.016 | 0.018 | 0.384 |
| Window switching rate | 0.026 | 0.011 | 0.018 |
| Window switching rate * Neuroticism | 0.012 | 0.01 | 0.2 |
| **Computer activity type** | | | |
| Email duration pct of all computer time | 0.026 | 0.015 | 0.078 |
| Productivity apps pct of all computer time | 0.014 | 0.015 | 0.365 |
| Non-work apps pct of all computer time | 0.005 | 0.016 | 0.764 |
| **Day of the week** | | | |
| Monday | 0.018 | 0.025 | 0.464 |
| Tuesday | 0.015 | 0.022 | 0.493 |
| Wednesday | 0.045 | 0.022 | 0.042 |
| Thursday | 0.026 | 0.022 | 0.228 |
| Observations: 582, Groups: 50 | | | |
| Pseudo-R2 (fixed effects) = 0.08 | | | |
| Pseudo-R2 (total) = 0.43 | | | |

The dependent variable is the average self-reported stress from EMAs.

Friday is the reference category for the variable "day of week".

Std $\beta$ is the standardized coefficient.

Table 4.3: Generalized linear mixed model for information worker's daily self-reported stress and computer use.

| **Measure** | **Mean** | **SD** | **Median** | **Range** |
|---|---|---|---|---|
| Computer activity duration (hrs:mins) | 4:37 | 1:12 | 4:43 | 2:07-7:00 |
| Outside workhours computer work (mins:secs) | 14:35 | 18:13 | 08:21 | 0-103:21 |
| Outside workhours computer work (%) | 5.72 | 7.33 | 2.96 | 0-36.74 |
| Batching email work (% of workdays) | 39.02 | 23.46 | 40 | 0-1 |
| Window switches | 389.62 | 170.43 | 345.17 | 126.70-895.06 |
| Window Switching Rate (per min of computer use) | 1.39 | 0.43 | 1.22 | 0.84-2.65 |
| Time on email (mins:secs) | 88:02 | 40:30 | 82:40 | 17:44-189:17 |
| Email % of all computer time | 32.78 | 12.66 | 33.28 | 5.2-62.21 |
| Time on productivity applications (mins:secs) | 114:43 | 56:46 | 100:25 | 36:08-251:05 |
| Productivity % of computer time | 39.66 | 14.05 | 38.94 | 13.86-74.66 |
| Time on non-work applications (mins:secs) | 22:28 | 30:41 | 11:20 | 00:27-165:34 |
| Non-work % of computer time | 8.17 | 11.13 | 3.92 | .15-47.47 |

Table 4.4: Descriptive statistics of daily averages of computer use measures. N=51.

was significant ($p<.001$). Specifically, working outside of typical work hours was associated with more stress for employees who indicated problems with work-life balance (Figure 4.2). It is worth noting that working outside of typical work hours was not common in our sample, as shown in Table 4.4.

In comparison, the duration of computer work and work outside of workhours were not significant in the model of daily self-reported stress.

### 4.3.3 Computer work strategies and patterns

#### Email batching

The average employee batched their email on 39% (SD 23%, range 0-100%) of their workdays. On average, 68% of daily email work was done in 3 blocks of time or less. To investigate whether following a batching strategy for email work was associated with daily stress, we included batching in the generalized linear mixed model predicting daily stress. An interaction term of batching and neuroticism was also added based on our finding from a previous

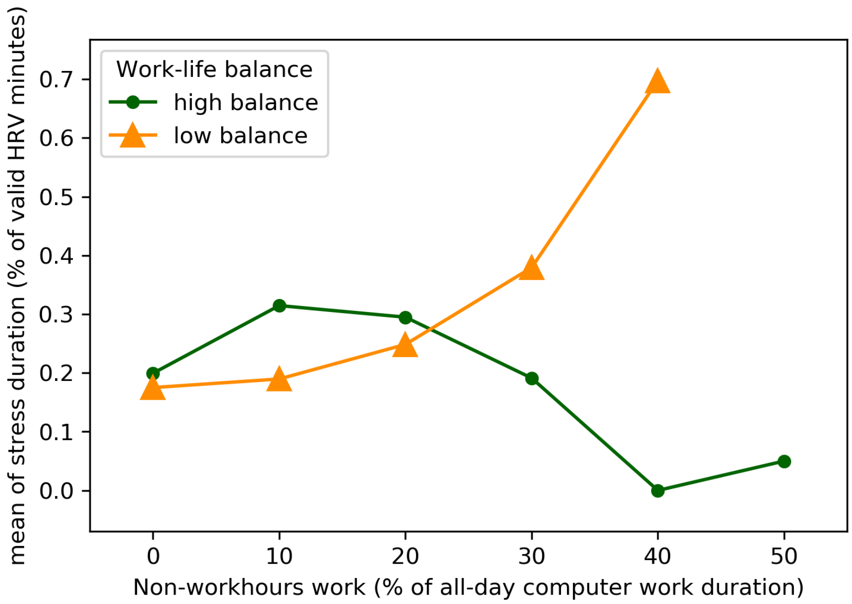



Figure 4.2: Interaction between non-workhours computer work duration and work-life balance on HRV-stress.

study in a lab setting [8]. The results in Table 4.2 show a significant main effect of batching and the batching x neuroticism interaction. Batching email was associated with less daily stress. Neuroticism, contrary to expectations, did not have an effect on daily stress, although it was associated with self-reported overall life stress (PSS) in a separate analysis (r=.63, p<.001). Confirming our previous findings from a lab study [8], employees who score high in the neuroticism trait are more stressed when they batch their email work than those low in neuroticism (Figure 4.3). Batching email work was not associated with self-reported stress.

### Window switching

Employees switched computer windows every 43.73 seconds, on average (SD 13). Window switching was not associated with HRV-based stress (Table 4.2), but was associated with perceived stress (Table 4.3). A previous study found that neuroticism was associated with shorter online focus duration [175] so we included an interaction term of neuroticism and window switching rate (the inverse of focus duration) in the generalized linear mixed model.

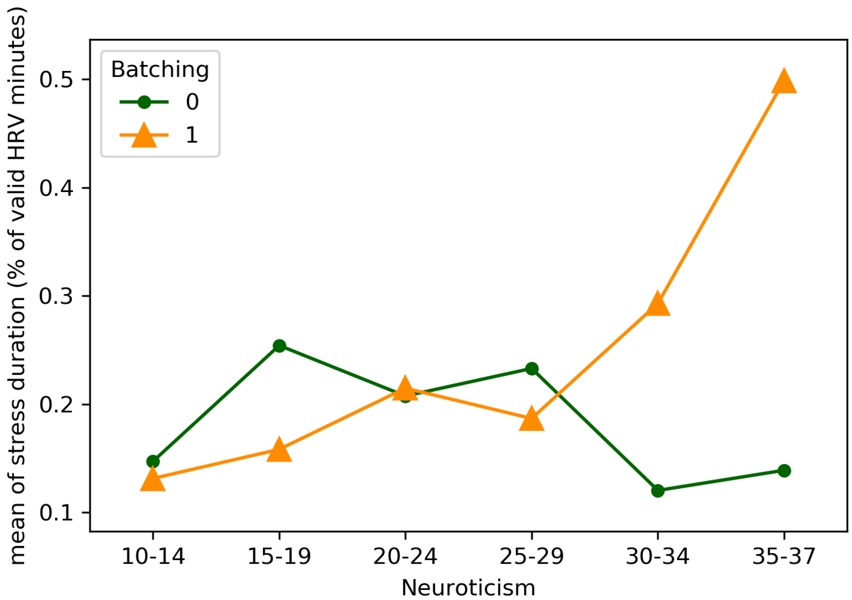



Figure 4.3: Interaction between batching and neuroticism on HRV-stress.

The interaction effect was significant. The relationship between window switching rate and daily HRV-based stress was moderated by the personality trait of neuroticism: window switching was associated with more physiological stress for employees with higher neuroticism than employees with lower neuroticism (Figure 4.4).

### 4.3.4 Computer activity types

Employees spent an average of 1:28 hours on email (SD 40 mins) and the average email duration as a percentage of computer work duration was 33% (SD 13%). The average time spent on productivity applications was 01:55 hours (SD 56:46 mins). For non-work applications, the average time spent was 22:28 minutes (SD 30:41). Productivity and non-work applications constituted 40% (SD 14%) and 8% (SD 11%) of all-day computer use duration at work, respectively.

The generalized linear mixed model (Table 4.2) showed that employees who spent more of their computer time on email were less stressed than those who spent less time on email. Time

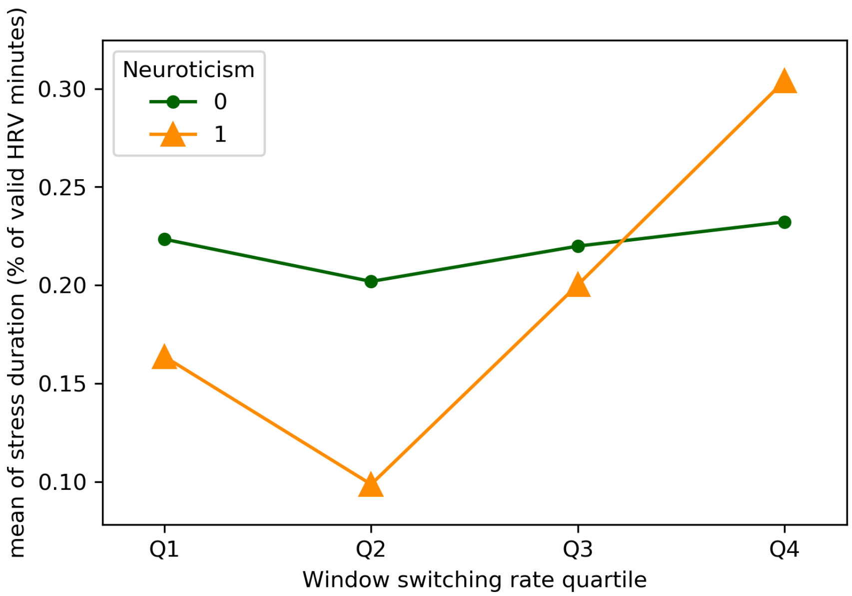



Figure 4.4: Interaction between window switching rate and neuroticism on HRV-stress.

on productivity applications (e.g. word processing, spreadsheets) was positively associated with daily stress, while time on non-work applications (e.g. social media, news, music, sports, shopping) was negatively associated with daily stress (Table 4.2). These factors were not associated with self-reported stress.

### 4.3.5 Variability of computer work duration

In a separate analysis, we investigated whether the within-person variability of computer work duration is associated with different measures of stress. To capture variability of daily workload, we excluded time spent on non-work related computer activities. Bivariate correlations between the within-person variability measures and measures of HRV-based stress, perceived stress, and job stress are shown in Table 4.5. We corrected for multiple comparison using the Benjamini-Hochberg method.

All three within-person variability measures (SD, RI, FRI) were not associated with the average HRV-based stress per person or perceived overall life stress (PSS). For self-reported work stress, the standard deviation of daily computer work duration was associated with perceived job demands (r=.53, corrected p=.003), but not control or support from the Job Content Questionnaire. Similarly, for the effort-reward imbalance measure, the standard deviation of daily computer work duration was associated with reported effort (r =.45, corrected p=.016) and overcommitment (r =.47, corrected p=.011), but was not associated with reward. The standard deviation of daily computer work duration was also positively associated with work-life balance problems (r =.42, corrected p=.033). The hourly variability measures (RI and FRI) were not associated with reported job stress measures.

The variability measures were not associated with daily self-reported stress (last 5 minutes or overall day). However, FRI measures were positively associated with arousal (Table 4.5). The higher the variability of workdays in terms of computer work duration, the higher the arousal (ie. energy). High arousal includes a range of moods and emotions from stress and frustration to excitement and enthusiasm. On the other hand, the more similar the days the lower the arousal. Low arousal includes a range of moods and emotions from bored to calm.

## 4.4 Discussion

### 4.4.1 What computer use factors are associated with daily stress at the workplace? How do individual differences affect those factors?

With computer activity logging, we aimed to create computer use measures that can be unobtrusively tracked and their computation easily automated. We aimed for these measures to be grounded in the literature on human-computer interaction, occupational stress, and ergonomics. A frequently used measure of computer use at the workplace is the duration

| | SD | | RI mean | | RI range | | FRI mean | | FRI range | |
|---|---|---|---|---|---|---|---|---|---|---|
| | r | p | r | p | r | p | r | p | r | p |
| **HRV-stress** | | | | | | | | | | |
| stress duration | .028 | .847 | -.030 | .840 | -.082 | .582 | .102 | .492 | .025 | .863 |
| **Survey: PSS** | | | | | | | | | | |
| overall perceived life stress | .165 | .261 | .024 | .868 | -.022 | .879 | .095 | .506 | .006 | .966 |
| **EMA** | | | | | | | | | | |
| last 5 min stress | .137 | .358 | -.138 | .338 | -.119 | .409 | .011 | .937 | -.021 | .882 |
| overall day stress | -.064 | .668 | .137 | .342 | .006 | .965 | -.057 | .692 | -.039 | .786 |
| arousal | .159 | .287 | .262 | .066 | .333 | .018 | .491 | <.001* | .538 | <.001* |
| valence | .001 | .992 | .146 | .311 | .167 | .245 | .244 | .088 | .325 | .021 |
| **Job stress survey: ERI** | | | | | | | | | | |
| effort | .451 | .001* | -.191 | .180 | -.042 | .769 | -.018 | .901 | .033 | .819 |
| reward | -.074 | .618 | .229 | .106 | -.150 | .294 | .074 | .607 | .044 | .757 |
| overcommitment | .472 | <.001* | -.070 | .627 | -.023 | .870 | .264 | .062 | .120 | .401 |
| **Job stress survey: JCQ** | | | | | | | | | | |
| demands | .539 | <.001* | .002 | .990 | .027 | .852 | .223 | .117 | .189 | .184 |
| control | .196 | .181 | .161 | .259 | .283 | .044 | 0.177 | .214 | .203 | .152 |
| support | -.093 | .531 | .154 | .280 | -.014 | .924 | -.010 | .946 | -.026 | .858 |
| **Survey: WLB** | | | | | | | | | | |
| Work-life imbalance | .418 | .003* | -.025 | .860 | .038 | .792 | .103 | .474 | .085 | .552 |

* Significant after multiple comparison correction
HRV: heart-Rate variability, PSS: perceived stress scale, EMA: ecological momentary assessments
ERI: effort-reward imbalance, JCQ: job content questionnaire, WLB: work-life imbalance

Table 4.5: Bivariate correlations of stress measures (HRV-based, PSS, job stress surveys) and within-person regularity measures (SD, RI mean, RI range, FRI mean, FRI range).

of time spent actively working on the computer. We surprisingly found that computer work duration was inversely related to daily stress duration. A plausible reason could be that employees in our dataset might experience more stress during non-computer based activities, such as meetings, presentations and conferences. It is possible that days with longer computer time are days that are more quiet, and employees get to finish their work tasks without much external interruptions.

Another surprising finding is that more time spent working outside of work hours (proportional to total computer work time) was associated with less stress, contradicting research that showed associations between after-hours work connectivity and several health and well-being problems such as mental health and cardiovascular disorders [274], and lack of sleep [44]. Several pointers from previous research could explain this contradiction. First, people with higher ambition and job involvement have been reported to work more outside of work hours [35]. A study of university faculty members reported that overtime work was experienced as less effortful and less stressful than regular workhours, and that workers reported positive work characteristics [27]. The observed lower stress during work hours in our sample could reflect the positive job perceptions. Second, previous research has highlighted several factors that moderate the relationship between after-hours work and the negative outcome, showing that the relationship is not direct. For example, perceived job autonomy, control and rewards moderate the relationship between working after hours and the negative outcomes [26, 253, 296]. Previous work has also suggested that the quality of overtime work and the work environment (i.e. rewards for overtime work and whether overtime work is executed voluntarily) dictate the relationship between overtime and fatigue [26]. These moderating factors could explain discrepant results in the literature. Third, previous research has focused on perceptions of overall long-term associations between outside-hours work and health outcomes, while our study investigated the daily duration of after-hours work and the association with day-to-day stress. Although limited past research has found associations between day-to-day short-term increase in workload (i.e. excessive overtime) and increased adrenaline excretion and elevated heart rate [80, 163], the long-term effects could be cumulative and therefore more pronounced [296, 229], especially given the relatively low average of overtime work in our sample. Lastly, the average duration of work outside of work hours in our study was low. A study found that moderate overtime (less than an hour a day) is common among employees and is not associated with fatigue, although it was associated with higher perceived job demands and surprisingly also associated with higher motivation [26]. We extend these previous findings by showing that days with moderate work outside of work hours are associated with less daily physiological stress.

The interaction between non-workhours computer work and work-life imbalance further clarifies the association with stress. For employees with high work-life balance, their stress decreased on days with more computer time outside work hours, while for people with low work-life balance (i.e. reporting not having time to socialize/relax with family/friends, taking work home, worrying about the effect of work stress on health, relationship issues with partner due to work), they experienced more stress on days when they worked more outside of work hours. Since our computer activity logging only tracked employees' desktop computers at their offices, this measure reflects time spent in the office outside of work hours. Those with work-life balance problems could view days when they have to work outside of their hours as adding more demands on their already highly demanding jobs, or days that could exacerbate already existing work-related issues with their partners or families. Research suggested that employees who might feel compelled by external circumstances to work for long hours tend to report problems with work-life balance [95] and thus might also experience more stress than those who are internally driven to work long hours and do not report worklife issues. On the other hand, those who work long hours and report a high work-life balance might be internally driven [95], which could also explain our finding that their stress decreases. Employees who reported that their job does not adversely affect their work-life balance might find occasionally working outside of work hours to be a quiet time to finish work tasks without external interruptions, and therefore experience less stress. Further research is needed to support these propositions. This finding builds on previous work that stressed the importance of considering moderating variables in the association between overtime work and its health outcomes [26, 253].

We found that on most days, employees continually check their emails throughout the day (rather than 'batch' their email work). On days when employees batched their email work, they experienced less stress. Previous work has indicated that email stress, or the feeling of email overload, is a result of continual interruptions to ongoing tasks and focus disruption that emails create [177, 271, 121]. Therefore, checking email a limited number of times throughout the day (ie batching) could decrease feelings of email overload and consequently stress, as observed in our study. Other studies suggest that batching might increase stress as email will pile up and create a sense of email overload [64]. Studies have therefore found mixed results for the association between batching and stress. Our previous work in a lab setting has suggested that the interaction between batching and neuroticism explains the past mixed results [8], which was confirmed in our current in-situ study. Stress increases with neuroticism on days when when email is batched, but not on days when email is continually checked. A possible explanation is that handling email in a batch requires a more sustained focus, which increases stress [97] and this might be more evident in neurotics as they are more susceptible to stress in general [186]. Also, seeing a pile of email all at once might create a sense of overload for people susceptible to stress. Future research can examine further factors moderating the the association between batching and stress. For example, as the volume of emails and the perceived importance of email to an employees' job affects their batching behavior [64], they could consequently affect stress.

The window switching rate in our study was similar to that reported in previous studies of focus duration for information workers [175], which also found that focus duration is shorter for employees who scored high on neuroticism. We extended those previous findings to show that for neurotics, increased window switching increases stress. High-interruption computer work conditions create a sense of more mental effort [142], which could increase stress for some employees, or create a sense of more positive valence and feelings of dominance for others [142] who might feel more connected and accomplished when multitasking. Our findings, along with previous findings on the effects of interruptions, suggest that dispositional factors dictate whether and how computer-use factors relate to stress.

Our analysis of computer activity types showed that more time on email was associated with less stress. This finding contradicts previous studies that linked email to stress [177, 176]. In a controlled experiment comparing completing tasks with and without email interruptions, researchers found that the email condition yielded reports of more mental effort, but also more positive valence and dominance, and no difference in perceived stress [142]. The authors explained that emails might have caused a feeling of being connected and glad to help. Another study based on self-reports reported that job autonomy, the perceived importance of email for work, email volume and spam volume increased feelings of email overload [64]. A study that logged work computer use and collected self-reports about mood and engagement reported that email can be rote or focused work, and that rote work was associated with feeling happy while focused work can involve stress [174]. Further research is needed to investigate email work and stress, taking into consideration job-related and dispositional factors, as well as email content, importance and urgency as potential moderating factors.

We found that increased time on productivity applications increased stress duration, which is expected given that these applications reflect job demands and require focus. On the other hand, time on non-work applications was associated with less stress. It is unclear whether the lower stress associated with non-work applications is due to employees taking breaks, which reduces their stress, or whether employees who spend more time on non-work applications have overall lower job demands which could be associated with lower overall stress.

Finally, daily self-reported stress was not associated with any computer work factors except window switching. Higher window switching rate was associated with a higher self-reported daily stress score. It is unclear why only window switching is associated with perceived stress, especially given the mixed results in previous work on the attentional and emotional states associated with window switching. In studies of information workers, researchers found that higher window switching is associated with boredom [174] and feelings of lower productivity [173]. Boredom is a low arousal state [246], which could explain why physiological stress was not directly associated with window switching. It is possible that feelings of low productivity at work might make employees feel stressed about completing work tasks, but this kind of perceived stress does not manifest physiologically. In Chapter 8, I discuss the distinction between perceived and physiological stress in more depth.

### 4.4.2 How does variability in computer use patterns affect stress at the workplace?

Our study introduced *variability* as a novel computer-use factor that could relate to stress at the workplace. Findings revealed interesting associations between variability in computer work duration and different measures of stress. Computer work duration variability was not associated with daily stress from self-reports or wearables. This can be explained by the biopsychosocial model of challenge and threat (BPS) by Blascovich and Tomaka [32], which states that the ratio between resources and demands dictates a person's stress experience. There are two sides of the scale that make up stress: it takes both the perception of high demands and the perception of low resources to experience stress. When the ratio is close to balance, a situation might be perceived as a challenge, but when demands exceedingly outweight resources, a situation is perceived as a threat and causes stress [32]. We found that day-to-day variability in computer work patterns affect only one side of the scale: the perception of higher demands, effort, overcommitment and arousal. These factors alone do not necessarily lead to stress. The perception of high work demands is associated with stress when it is paired with the perception of low control and support. Work effort and overcommitment are associated with stress when there are also low rewards. Arousal is associated with stress when valence is low. These findings suggest that variability in computer work is associated with stress for some people (i.e. those who have low job control/support, low job rewards, low valence). Those with more balanced demands and resources might not experience stress from variability of computer work patterns because their high job rewards, support and control offset the added demands and effort.

To our knowledge, this work is the first to compute variability measures of logged workplace computer use data, and the first to associate these variability measures with stress. These variability measures were computed per participant rather than per day like the previous analysis with the mixed model. HRV-based stress was averaged per participant for the duration of the study so the patterns of daily stress might have been obscured. Given our results overall, HRV-based stress might better reflect daily stressors (e.g. time on productivity apps, working outside workhours) while self-reports of overall job stress better reflect overall job-related patterns (e.g. how regular the workload is).

### 4.4.3 To what extent does unobtrusive monitoring of workplace computer use help identify stress levels?

One goal of tracking workplace activities is to build predictive models that can accurately identify when an employee is stressed. Building such models based on tracked computer usage eliminates the need for the employees to continuously wear heart-rate sensors for stress tracking, as these wearables would only be needed for building the model, and occasionally to update it. In our study, the fixed effects based on computer usage and personal factors explained 14% of the variance of daily stress, which is considerable given the myriad factors that affect daily stress. However, this is not enough to build systems that predict or identify daily stress. A higher prediction power will be needed to accurately capture employees' stress level from computer usage alone, and to utilize those predictions for stress management interventions and other applications. The high conditional $R^2$ (fixed+random effects) tells us that most of the unexplained variation is between individual employees rather than between observations within an employee's data. Therefore, we would not expect to significantly improve our model by collecting more data on measures that mainly vary within individuals, but instead should find measures that mainly vary between individuals. This stresses the importance of individual differences and the need for individual models of stress rather than generalizing models.

Previous studies vary greatly in their attempts to build models that unobtrusively infer stress at the workplace. In a controlled laboratory experiment, Hernandez et al found that keyboard pressure and mouse contact could detect stressful computer tasks for over 70% of participants [108]. Another study that tracked 15 researchers for 5 days in their real work environment found that head motions (indirectly capturing head gestures and facial expressions) yielded a predictive value of 59% for predicting daily self-reported stress (2 classes, high/low stress) [109]. Features extracted from sensors tracking contextual signals such as atmospheric pressure, humidity, light and temperature yielded predictive values between 53% and 58% [109]. Using smartphone features (audio from microphone recordings, physical activity and location from motion sensors and social interaction features from phone calls, calendar, address book and battery), Muaremi et al . reported an accuracy of 55 % for predicting three levels of daily retrospective self-reported stress (low, moderate, and high perceived stress) [200]. It is important to note that these studies used machine learning models with 2 or 3 class classifications of stress, which have a different interpretation from our linear mixed-effects model. While the stress classification model prediction power reflects the accuracy of the model (i.e. what percentage of the cases were correctly classified), our model shows what percentage of the variance in stress is explained by the independent variables and captures more fine-grained day-to-day variation of stress than 2 or 3 levels of stress. As machine learning approaches in precious work are mostly focused on obtaining the highest model performance, these studies do not report on the direction and strength of the association between the features (i.e. independent variables) of the model and stress. To our knowledge, this is the first study to investigate and quantify this range of features for computer work and how they relate to daily stress measured objectively and continuously with sensors.

Regardless of the overall prediction power, capturing daily activities alongside physiological data can provide insightful information for a wide variety of applications, such as visualizations to augment memory, reflect on daily activities and feelings, and share emotionally significant moments with others [103, 139]. Research has shown that users benefited from visualizations pairing their physiological stress and work activities that were tracked continuously throughout the workday [103, 139].

Daily EMAs did not correlate with computer use factors in the daily analysis nor in the overall aggregate analysis per participant. The extent to which EMAs can accurately capture daily stress associated with specific uncontrolled events needs further validation. I return to this point in more depth in Chapter 8.

### 4.4.4 Limitations

As our main goal is to investigate computer use factors that can be unobtrusively and automatically tracked, we did not incorporate other data sources that might provide further work context, such as calendars. Calendar data requires careful annotation and curation, which does not align with our approach. There would be large differences in how employees label their calendar events, and capturing that information might also introduce privacy issues. Future work can investigate ways to capture the context of computer work.

Our sample is not balanced in terms of the number of male and female participants. The results could be biased towards females, and employees with higher education degrees. We believe our results generalize to other information work contexts, but generalizing to other work contexts requires careful consideration.

The classification of work and non-work computer activity is not 100% accurate. Some participants said they use social media, shopping or food delivery websites for work purposes. On the other hand, productivity applications could be sometimes used for non-work purposes. We believe our classification reflects the overall most common purposes of the classified computer activities, but future individual-level analyses should account the differences in what is considered work and non-work computer activity as per the individual's job role.

## 4.5 Conclusion

Computer interactions at the workplace can reveal information about an employee's stress. Unlike previous efforts that relied on self-reported measures of computer use and stress, we used unobtrusive and continuous measures that can be automated and incorporated in real-time applications such as visualizations and interventions. The duration of computer work within and outside work hours, computer work strategies and patterns, and time spent on different computer activity types explained 14% of the variance in daily stress. Explaining 14% of the variance of a highly complex affective state is meaningful, and motivate further investigations to get a more complete picture of workplace stress. Our findings, along with previous findings on the effects of digital interruptions, suggest that dispositional and job-related factors dictate whether and how computer-use factors relate to stress. Organizational interventions to reduce stress should consider how policies around working outside of work hours, email checking norms, and non-work computer activities could affect different employees differently.

# Chapter 5

# Information Workers' Perspectives on Technology-Supported Stress Tracking

## 5.1 Introduction

Previous chapters have discussed two methods for in-situ stress measurement: wearable sensors and EMAs. These methods vary in the physical, time, and emotional demands they place on the employees tracking their stress. Although some studies in HCI have deployed stress tracking in real workplace settings using varying methods [177, 176], it is not well-understood how employees perceive the benefits and burdens that the varied options for tracking workplace stress present. Contrasting employee perspectives on different technologies for stress tracking can help inform design recommendations for stress-monitoring systems which employees can use to effectively monitor their stress without disrupting their workflow or introducing more stress. Given the increasing prevalence of personal tracking in workplace wellness programs [48, 238, 277], designing with employee perspectives in mind can result in more successful deployments of these systems.

From interview data of participants in the workplace stress study detailed in Chapters 3 and 4, we compared employee perspectives on automated (i.e. passive) and manual (i.e. participatory) stress tracking in the workplace. We link those perspectives with discussions of automated and manual self-tracking in the personal informatics literature to provide insights for the research and design of future workplace stress tracking systems. Specifically, we address the following research questions:

- What benefits and challenges do employees perceive overall with technology-supported stress tracking in the workplace?
- How do employees' perceived benefits and challenges differ between automatic (e.g., wearable sensor) and manual (e.g., EMA) stress tracking?
- What are employees' preferences for how technology can support workplace stress monitoring, and what design guidelines can be produced from these preferences?

Our work contributes the following: (1) An understanding of the different perceived benefits of wearables and EMAs. Almost half of the participants reported not engaging with their wearable's data enough to understand their stress. Participants reported brief interactions with the wearable when they were stressed to validate how they felt, while their interaction with the EMAs encouraged them to reflect more on how they were feeling; (2) A description of the distinct challenges that affect data usefulness and data reliability for each tracking modality. Stress data from wearables were more difficult for participants to interpret, but EMAs created challenges quantifying and articulating subjective stress. Both modalities were subject to missing data, but missing data for EMAs resulted in data which might not accurately reflect a person's overall stress because EMA prompts were not answered during high stress periods; (3) Evidence that employees have varying and sometimes conflicting preferences around how technology can support stress tracking at the workplace, and that organizational and individual goals for stress tracking can be challenging to align; (4) actionable design guidelines. Building on these contributions, we present recommendations for how designs can account for this variability and meet organizational and individual stress tracking goals.

## 5.2 Methods

We aimed to understand the benefits, challenges, and preferences information workers have when stress tracking is implemented in their natural workplace. Employees' perspectives were obtained through interviews after the 3-week workplace stress tracking intervention using both automated and manual tracking methods, as described in Chapters 3 and 4.

In a semi-structured interview at the end of the study, participants were asked open-ended questions about their experience monitoring their stress during the study, what they liked and disliked about stress tracking, and any challenges they encountered. To further elaborate on potential benefits and challenges of stress tracking, participants were asked whether they viewed their stress data on the wearable or its app, and if they did, how viewing their stress levels helped or did not help them understand and/or manage their stress or stressors. Questions also included whether they feel that they changed their behavior knowing that their stress and computer activity are being tracked. They were also asked to describe what they would change about how technology could support stress tracking at the workplace. Interviews typically lasted for 15-25 minutes. Our IRB approval did not cover audio-recording the interviews, so the interviewer took detailed notes on participant responses and recreated transcripts immediately following the interview. Note-taking was done as the participant was speaking, and verbatim quotes were noted, excluding vocalized pauses such as "hmm" or "uhh". Notes were not taken when participants were discussing matters unrelated to the questions (e.g. talking about the gift card compensation or their overall job role). When transcripts were reviewed following the interview, typos were fixed and missing words (e.g.

the, and) were filled.

We thematically analyzed our interview data according to Braun and Clarke [38]. Two researchers read the interview notes, generating initial codes in four categories: perceived benefits, perceived challenges, perceived behavior change, and suggestions for stress tracking at the workplace. After refining themes, the final codebook consisted of 26 codes, including codes for missing data, interpretation challenges, reflection, validation, and technical issues. One researcher coded all interviews, with another researcher reviewing the coding and discussing final themes. We did not calculate inter-rater reliability because interviews were semi-structured, and people frequently apply the same code to different parts of a conversation [15, 187].

Participants were divided in two groups based on their objectively measured daily stress (HRV-based stress). Perspectives of participants in the top quartile of daily stress duration (stress duration >= 30% of the workday) were compared to perspectives of participants in the bottom quartile (stress duration <= 7.9% of the workday). The same themes emerged in both groups in similar frequencies.

We report counts of participants who expressed certain perspectives under each theme. However, since the interviews were semi-structured with open-ended questions, participants were not explicitly prompted about each perspective, and were not always prompted to elaborate on their perspectives beyond what they expressed. Therefore, our reported counts of participants who expressed a perspective might under-represent the actual number of participants who agreed with it.

## 5.3 Results

Our findings revealed distinct benefits and challenges employees expressed about different methods of stress tracking. Participants also expressed different preferences for how technology can support stress tracking at the workplace.

### 5.3.1 Perceived benefits

**Overall: Unobtrusiveness and awareness**

When asked about their experience in the study, most participants (n=34) indicated that their experience was smooth and that the study was unobtrusive and did not interfere with their work. As one participant indicated "it is in the background you forget it is running you don't notice it." (P52) and another commented "it was very non-intrusive. Not really a lot to do." (P48). Besides unobtrusiveness, the familiarity of activity trackers was appreciated, especially for those who already use other activity trackers, as P24 indicated "I do wear a smartwatch usually, so I'm used to having that." P42 expressed a similar sentiment that it is not out of the ordinary for them to wear a smartwatch. For the EMAs, participants appreciated that it was very short and did not take much time away from their work.

As an overall benefit, two participants said that they tried to stress less, "at the beginning I was aware. I was checking how my mood was more, I would calm myself down at emails that irritate me. After a while I forgot it was there and I was not thinking about whether I was stressed." (P48) and "I am a worry kind. I tried to be more relaxed. Not that I intentionally did it but the watch was a reminder to not stress" (P33) while others said they did not change their stress-related behaviors (see subsection 5.3.2 Overall: Minimal impact on stress management practices).

### Wearables: Validation of perceived stress

Participants who viewed their stress data on the wearable device tended to do so to seek validation of how stressed they felt. Participants reported checking to see if the device reflected that they are stressed. P1 noted, "when I had some really stressful moments, I looked at the watch to see what it said and it was accurate! And I thought yep! It detected when I was stressed." Others tried to confirm their perceptions, such as P32, "sometimes I feel stressed then I check to see it if shows" and P25 "The device was nice to use because it had stress and I was able to look at it and see and it was reflecting my stress".

Participants rarely brought up instances where their measured stress did not align with how they felt. Only P38 indicated that they felt the recorded stress did not align with what they felt "I don't think it showed that I am stressed, sometimes [when] something very stressful [happens], I check the watch and it showed no stress", which also indicates that participants seek validation of their perceived stress by checking the wearable device, even if the recorded score does not always align with their perception. Other participants seemed to trust the stress level on the device over their own evaluation of their stress level. P42 trusted the device over their own judgments, saying "I think for me it confirmed. I wouldn't know if I am stressed or not so checking to see confirmed it showed me I am stressed." P50 learned, "I realized I'm not as relaxed as I thought I am."

### EMAs: Reflection on stress and stressors

Participants often felt they benefited from reflecting on their stress through the EMAs, reporting that the short surveys prompted them to think about their stress level in order to report it in the survey. For example, one participant said "I found myself having to check in with the survey. It was good to be able to do that [because it helped me] evaluate and force me to think about it and reflect." (P25), another said "[it was] great to reflect upon how stressed I was and having those check in points about how stressed I am." (P42). The check-in points often caused some employees to realize that they were stressed. For example, P48 indicated "It did try to make me think whether I am stressed or not especially when doing the survey. I appreciate that it let me look into what I was doing and how I was feeling." P29 particularly appreciated how the EMAs required them to try to measure their stress level, saying "I liked the survey it was self-reflection for me because I never quantify how stressed I was after stressful moments." P12 felt they benefitted more from the survey than the wearable because it encouraged reflection, describing "I most benefited from the survey because [...] I thought critically [about] what had been impacting me in an unusual way in a given time and I had to articulate it and write it in the survey so that was helpful".

To further investigate how participants reflected on their stress in the EMAs, we analyzed the free-text responses to evaluate whether participants provided sufficient information about potential stressors in their day to contextualize their self-reported stress score. The average participant left the free text empty in 32% (SD 33%) of the surveys they took, providing no contextual information about their reported stress level. The non-empty responses often did not explain their stress level in-depth, as sometimes participants left comments like "none", "nothing atypical" or a single character like a dot. The most frequent comments in the free text question were variants of "yes", "typical" and "none" (46%). The average participant's response length was 5.13 words (SD 2.98). One-word and two-word responses other than variants of "none" and "typical day" included activities such as "meeting", "driving", "budgeting", "writing", "student interviews", "team workshop", "engagement review", "doing webinar" and "conference working". Others highlighted feelings such as "stress", "frustrating", "headache", "good" and "busy", while some described events such as "big event", "deadline day", "extra activity". Others described the overall sentiment of their day as, "good day" or "quiet day."

We sampled 100 random responses that were over 5 words long, finding that participants mostly described events rather than feelings in longer descriptions. Participants reported daily events such as when they arrived at work, e.g. "Arrived to work later than usual", non-work activities during work hours, e.g. "took a 15 minute walk around 11" and "I went out for lunch today", planned work activities for the day, e.g. "no meetings in morning so hopefully will spend most of my time on computer writing" and "Today is a typical recruitment day, but also have to interview participant for one of the studies I coordinate.", describing work activities such as attending a meeting, e.g. "Had two video conference trading this morning and afternoon" and "Team meeting and meeting with school leadership", or being busier than usual, e.g. "Back to back day no breaks" and "Covering extra work duties for vacant position". A few responses described feelings, such as "I typically feel a little more relaxed after the two conference calls this morning. My stress level decreases and am also less anxious.", "I'm at staff social having a good time", "A lot of work to do in the office but feel more relaxed today" and "Stressed and tired from working late last night and having a lot of work to do today". Two responses also indicated coping strategies "Had to take a mental break in afternoon and bought some chocolate" and "eating a lot of sugar, deadlines!!". Some also described non-work events that could affect stress at work, e.g. "[a person] was in a car accident and I just found out" and "Yes [typical day] but I have a cold". Overall, when describing their day for stress tracking, participants tended to reflect on events and activities rather than sentiments or feelings. These events can provide context for stress by identifying potential stressors.

### 5.3.2 Perceived challenges

#### Overall: Minimal impact on stress management practices

Despite some participants reporting awareness of their stress as a benefit, almost all participants (n=48) said that they did not change their stress or computer use behavior during the study period. Most participants felt they were too busy to change their work routine. P1 said, "It didn't make me change my routine. My routine stayed the same." P15 felt that passively tracking was insufficient to impact their stress management practices, saying "No [it didn't change my behavior], because it wasn't something that was in your face like notifications, It didn't change anything about my stress." Participants also described forgetting that their stress and computer use were being monitored, such as P63 "The first day or two I felt conscious of it and then no I was not thinking about it." Other participants said they were more aware of their activities on the computer as they know they are being logged, but their normal computer use did not change. For example, P42 said "I was aware of it but maybe not really changed my behavior [...] soon after I resumed to work as usual" and P59 said "I was more aware but I don't know if I changed my behavior".

**Wearables: Missing data, lack of engagement and difficulty interpreting stress levels**

For the wearable device, reported challenges included an uncomfortable form factor, remembering to wear the device, and technical challenges. The device was uncomfortable to wear for some participants (n=6). P48 said "Personally, I don't like wearing accessories so I had to get used to wearing it I couldn't wait to get home and take it off. The way it is positioned bothered me when typing". Some participants (n=7) stated that remembering to wear the device was a challenge. Technical issues, reported by 10 participants, mainly concerned the frequency of needing to charge the device and connection issues between the phone and device.

When participants reflected on their experiences of stress tracking, many felt that it helped them understand or be more aware of their stress and stressors, as indicated in the benefits section above. However, about half of participants (21) indicated that they did not view their data enough to say that they understand their stress better. Other participants said it helped them “understand but not manage” (P45) their stress.

Participants often struggled to interpret the stress level generated by the wearable device, which showed as a number between 0 and 100. Some participants stated that although they were interested or curious to see their stress data, they did not understand what the numbers meant. For example, P63 stated “I was curious but I couldn’t make of the numbers. I didn’t know what that meant so I decided I don’t really care” and P28, “I was not sure what it meant. Just something to look at.” One participant attributed the difficulty of understanding the automatically-generated stress scores to the lack of comparison against a threshold, “it was just interesting to see but I didn’t have anything to compare to so it didn’t help [me to understand my stress]” (P48), which might indicate that users are more interested in relative scores rather than absolute scores displayed on the device. Similarly, another participant said “I don’t know what the values/range is for stress” (P15) even though they were occasionally interested in viewing their data, which also indicates the need to have a threshold to compare to.

**EMA: Missing data, difficulty articulating stress, and intervention causing stress**

On the other hand, challenges specific to EMAs included not paying attention to the phone while busy or not being able to answer surveys during meetings (n=28). P27 indicated “I missed some [surveys] because I’m busy or in meetings can’t stop what I’m doing.” P60 indicated that when their day was busy and stressful, they would occasionally miss surveys, “I missed a few especially if it a stressful or a busy day I miss all of them.” This indicates EMAs sometimes missed high-stress periods when participants felt they could not stop what they are doing to take the survey. On average, participants missed a third of the EMA prompts on weekdays, amounting to 15 (SD 10) missed EMAs out of 45 EMAs over 15 workdays.

Participants reported not being able to articulate an answer for the free-text question or being unsure of how to rate their stress on a scale (n=12). P19 said that it was "hard to describe how I feel on the survey." For the stress sliding scale, participants commented that it is "confusing because what does it compare to? How do I select a point on it?" (P52). Another participant also mentioned that it is hard to pick a point on the scale because they could not compare it with their other responses and suggested "if there were markers at least to compare to my previous response" (P12). P12, P20 and P52 indicated that they hesitated to pick the highest stress point, in order to preserve it for a potential more stressful time. They were not sure at any point whether the stress they are experiencing is their maximum stress, half of their maximum stress, or any other proportion corresponding to a point on the scale. Others commented on the subjectivity of the survey, questioning its rigor for evaluating stress, as P58 stated "Surveys [should] be less subjective. I am bad at evaluating my own emotions" and P30 also said "It's very subjective how you say pleasant or stressful".

For the free-text question about whether this is a typical day and asking participant to describe any abnormalities in their day, participants described the question as "confusing" (P44 and P49), "difficult to answer" (P47), "vague" (P52) and "repetitive" (P55). They explained that it is hard to describe their day. Sometimes it is typical for their days not to be typical. Several tasks and events are typical of their job but can be stressful. P28 said "I had trouble articulating for the last question" and P44 also said "was it typical day? I would say yes or no but don't know how to describe" which is what a few other participants also expressed. This indicates that for open ended questions, participants might be unsure of what to log and in how much detail. Completing EMAs also created stress for some participants, such as for P32: "[the] survey adds a bit of stress". One participant (P35) indicated that "having to have my phone near me for the survey made it a little bit harder" because they manage their stress by keeping their phone away during work and only checking it during breaks. It is possible that the additional stress from EMAs was caused by EMAs sending notifications in inappropriate busy times, as described earlier, or that EMAs reminded participants of their stress as they reflected on how they felt to report it in the survey.

### 5.3.3 Conflicting preferences for stress-tracking at the workplace

When asked about what they would change about stress tracking in the workplace, almost a third of participants (n=15/50) said more workplace factors should be tracked to provide context as potential stressors, such as external interruptions from colleagues and meetings. Participants expressed interest in seeing how those factors relate to their stress. Five participants suggested adding actionable insights to stress tracking, and expressed opinions on how their organizations should be involved. For example, P12 suggested "the organization should have clear outcomes that benefit the individual like workshops. [For example,] you can go for a walk in these windows in a day. Clear actionable recommendations and follow through. Have the HR identify when you had a spike either live or in next day report". P13 also suggested providing advice for stress management, but added that "it should be somewhat self-directed but give tips". Another suggestion was adding community support: "have the opportunity for people participating to meet if they want to talk about how it is going and what you're learning. It takes commitment so it is good to see others for accountability" (P13).

For EMAs, one participant suggested adding the option to manually log stress scores in the daily surveys in addition to logging at pre-scheduled times. Another participant suggested adding multiple alarms as reminders to take the daily surveys, while others suggested fewer daily prompts and said that "surveys multiple times a day forever could be annoying" (P024). Because it was common to miss phone EMAs, some participants (n=5) suggested computer-based EMAs such as having them sent by email. For automatic stress tracking, three participants preferred stress tracking without the need for a wearable device.

**Privacy**

Participants had conflicting views about privacy. A few participants explained that they are not generally concerned about privacy due to tracking their stress in the context of an IRB-approved research study, while also stating that an organizational implementation of stress tracking would be considered invasive. For example, P13 said "I know it is for research not like my performance review" and P30 said "I would be uncomfortable with my employer doing something like this, I would feel it is too intrusive". During the study, some participants felt "a little bit weird to know everything was being tracked" (P32), with P43 adding, "sometimes I feel like I was being watched".

While most participants did not comment on privacy concerns regarding stress tracking with EMAs or wearable sensors, some raised concerns about computer activity tracking as a part of stress tracking. Several participants said that at the beginning of the study, they tried to limit their non-work web browsing to avoid that being logged. However, they soon returned to their normal routine. For example, P38 said "at the beginning I was aware and then I forgot it was there" and P48 said "At the beginning, [. . . ] I probably did less non work stuff then I realized those are break times for me to get away from doing too much work at a time so I just step away from my desk. Then I forgot things were there". P14 also said "I was more aware of what I was looking at on the screen because I knew my computer activity is being monitored. I don't like having things logged. I tried to just behave normally after a little while". A common alternative to non-work-related browsing on the computer was browsing on the mobile phone, which was not logged: "Maybe 1-2 days I was aware that my computer was monitored so I would use phone instead of computer for social media breaks" (P17), "I didn't go on as much on [non-work website] so I would open it on my phone instead." (P19). Participants felt that if their organization were to implement such a stress tracking system, they would be potentially concerned about what data was being collected, despite desiring collecting more potential stressors for their own self-understanding. For example, P59 said “Some individuals will have an issue with activities being logged even if anonymous. People have privacy concerns” and P43 said “Watching what we are doing on the computer will create a lot of dissatisfaction”. However, other participants did not have such privacy concerns, “I don’t really care. I always operate under the assumption that it is visible what I’m doing on the work computer” (P49).

## 5.4 Discussion

The findings revealed that stress tracking with wearables and EMAs provided different benefits to employees, and that employee preferences might not always align with organizational goals of workplace stress tracking. While wearables were beneficial for validating how employees felt at times of stress, EMAs were helpful for pausing and reflecting on stress several times a day. These perceived benefits raise questions about trust in algorithmic output, confirmation bias, and balancing unobtrusiveness and engagement. We discuss these findings in light of literature in personal informatics and broader self-tracking applications to provide an understanding of how people perceive their interactions with stress-tracking devices and data. We also discuss implications for designing and deploying stress tracking in the workplace to support potentially conflicting personal and organizational tracking goals.

### 5.4.1 Design implications for the validation vs. reflection on stress

Participants perceived different potential value from tracking with each modality. Seeking validation from the wearable’s objective stress measure could indicate people’s trust in algorithmic output [112, 273, 312, 322] which could be due to assuming sophisticated system capabilities [112, 269]. Surprisingly, some participants even trusted the wearable’s stress level over their own appraisal of their stress, and others questioned the validity of self-reported stress in EMAs for being subjective and relying on one's potentially poor judgement of their stress. This might indicate that people view the system as surpassing their own abilities to evaluate stress. Deferring emotions to devices and trusting algorithms to tell us how we feel has been discussed in previous work [112]. Users have been found to trust and confirm even random outputs of systems framed as "intelligent", assuming that the system has better insight into the mood of users than users themselves [273]. This has important implications as people's understanding of their stress, as well as their experience of stress, could be influenced by system outputs. False system outputs about stress based on heart rate have been found to impact self-reported anxiety after a stressful task, reducing user anxiety when the system's output falsely indicated low heart-rate [60]. Additionally, the same physiological data could be framed positively as 'engagement' or negatively as 'stress' and can influence how users perceive their emotional states [112]. When users of emotion-feedback systems have little or no knowledge about how these systems work, they might adopt incorrect conceptual models to confirm system outputs. To address this issue, implementation of stress tracking in the workplace should improve users' knowledge of how the algorithm operates to measure stress, to avoid adopting incorrect conceptual models confirming and deferring to system outputs. For example, employees should be educated on physiological stress reactions that sensors measure, the confounding variables affecting those measures such as physical activity and overall health, and the distinction between physiological stress responses and emotion recognition through subjective evaluation.

Some employees reported checking the device in moments of high stress to validate perceived stress. This behavior could introduce confirmation bias, where people look for confirming evidence to prove their own hypothesis and ignore disconfirming evidence [313]. Specifically, people will only confirm their perceived stress in certain moments and miss other stress episodes that could present disconfirming evidence about their perceived stress. A study found potential confirmation bias with self-monitoring for patients with diabetes, where patients formed hypotheses about increases in their blood sugar and monitored their data to reinforce their assumptions [169]. Confirmation bias in stress tracking could similarly lead to people creating an incomplete understanding of their stress. Automated methods for tracking stress should therefore facilitate discoveries about one's stress, rather than merely reinforce presumptions. This could be done in real-time through notifications giving people updates on their measured stress at randomized times, or retrospectively by presenting daylong patterns of stress at the end of the workday. In systems that capture context, providing contextual information alongside measured stress could facilitate learning of pairing context with stress.

Overall, while participants appreciated that they could validate or confirm their perceived stress and reported it as a benefit, this brief interaction with a wearable is unlikely to generate deep insights that lead to awareness and action [157]. In contrast, our finding that EMAs encouraged self-reflection points to a typically desired outcome of self-tracking. The self-tracking literature frequently suggests that self-reports encourage participation and engagement with the data, and help people reflect on their own emotions [179, 242]. To encourage more reflection and avoid deferring emotions to system outputs, automated approaches could engage users by encouraging them to actively evaluate system output [33, 269], either in real time or retrospectively.

### 5.4.2 Trade-offs between unobtrusiveness, engagement, and value

Several design recommendations can be made to balance the tradeoffs between unobtrusiveness and value to benefit employees and organizations. We found that employees often did not engage with their wearable's data or could not interpret it. To encourage engagement with and understanding of objective stress measures, glanceable or ambient feedback for wearables' data can be designed [91, 201, 325]. Research has also shown that lowering access burden such as providing feedback on the computer's taskbar instead of having to open an app or a website to view data increases engagement and improves awareness [57, 134].

Designers of these tools could consider adding a display of stress information, although the effects on privacy and potentially increasing stress about the results warrant further study.

While automated modalities run in the background to continuously and unobtrusively collect data throughout the workday, manual stress tracking with EMAs runs the risk of not capturing the moments which are important to capture, as employees indicated not responding to EMAs when they are busy. Because being busy often correlates with stress, those moments with missed EMAs could be the highest stress events. While this limitation might not need to be addressed to support *employees'* stress tracking goals, it poses a challenge for *organizations* trying to obtain complete logs of stress throughout the workday and during different work activities. Interruptibility detection [213, 214], which integrates with other work systems, could be used to find opportunistic moments within busy periods to prompt employees to fill out EMAs. System developers could also consider EMAs on other devices such as wearables instead of phones, which have been found to increase the number of answered prompts and the speed to start answering [109] and are potentially less interruptive or distracting [119]. EMAs could also leverage voice assistants (e.g. Siri on iPhones and Apple Watches), although the appropriateness to work environments might vary based on office space (e.g. open or closed office space) and perceived sensitivity of stress logging.

Our participants reported difficulties quantifying and articulating their stress in EMAs, which has also been found in a previous study [2] where participants criticized the 5-point stress scale and suggested adding further granularity. A study on food journaling also found that participants had difficulty deciding what to report in EMAs and in how much detail [59]. Adding more guiding questions in the EMAs could help employees articulate details about stress in their day. For example, because participants typically attributed their increased or decreased stress to the presence or absence of events (e.g., meetings, deadlines), prompts could ask employees to consider whether that day's events influenced their perceived stress. EMAs could also use open-ended logging to avoid quantifying stress on a scale, allowing employees to describe their stress without the imposed structure of a scale corresponding to numbers. While flexible logging might help employees to express their stress in ways that make more sense to them, it creates significant challenges for organizations trying to quantify or summarize employees stress to feed into intervention systems or to systematically compare stress patterns across employees.

For both manual and automated methods, adding comparative measures or signals might improve reporting and understanding of one's stress. Participants reported wanting to compare their stress to a threshold or to their previous stress in order to better report their current stress in EMAs or better understand the sensor stress score on the wearable. Facilitating comparison with a previous reported stress (or multiple recent stress logs) or a benchmark might assist with the difficulties in quantifying and self-reporting stress, and also assist in understanding objective stress measures on the wearable device.

### 5.4.3 Designing for varying and conflicting preferences

Participants had various and sometimes conflicting suggestions and preferences for how technology can support stress tracking in the workplace. The most frequent suggestion was the need to track more workplace factors such as emails, meetings and social interactions in order to better capture stress and provide a more realistic and complete picture of workplace stress and stressors. At the same time, some participants raised privacy concerns about the computer logging component of the study. It is possible that participants' varying privacy concerns could relate to their perception of their employers' involvement in the process. Privacy concerns could also shift over time, as more concerns can surface as employees have more experience with tracking [90]. Therefore, we recommend the design of flexible stress tracking interventions, allowing employees to choose what they are comfortable with for a system to collect about them alongside what workplace factors might help them achieve their stress-tracking goals. Tracking systems can have privacy controls breaking down the kinds of computer activity or other automatically-logged measures to let employees specify what to log and whether to share data with the organization.

Some participants indicated interest in having community support to discuss what they are learning from tracking and issues they are facing. Including social features in tracking applications, as well as having offline social communities of trackers in the workplace could encourage participation and sustain motivation [5, 96]. However, other studies have found that social features can both support or discourage tracking [59] as some participants could have privacy concerns about sharing their data and their practices. Further research is needed to examine the desirability and effects of integrating social features in stress tracking at the workplace.

While some participants wanted additional manual (i.e. non-scheduled) EMAs, computer-based EMAs, and more notifications and reminders about EMAs, others wanted less frequent EMAs. For the wearable, while some participants appreciated its familiarity and thought it is easy to wear, others reported that it is uncomfortable and preferred tracking stress without wearing a watch. To accommodate these varying and conflicting preferences, multimodal systems can be implemented [46, 135, 158], supporting both EMA and automatic logging that can be used separately or simultaneously. The tracking system can then merge the data or use the data of one modality as a backup to the other while maintaining distinct interpretations of each modality (i.e. clarifying that automatic tracking measures physiological reactions while manual input measures personal appraisal).

In summary, workplace implementations of stress tracking should support customization for what and how to track in order to accommodate varying and conflicting employee preferences. The workplace presents unique challenges around privacy and personal data sharing with the organization, as well as the challenge of balancing the need for complete and continuous data for organizations to track patterns of stress, with employees' preferences for different tracking modalities.

### 5.4.4 Limitations

We do not report on motivations or goals of people who track their stress, since participants in this study were instructed to wear the watch and answer EMAs, and were not necessarily self-motivated to track their stress. Hence, our conclusions might not generalize to people who elect to self-track their stress. However, it could generalize to workplace wellness programs where people are given wearable devices or EMAs.

Giving participants both stress tracking modalities at once might have affected their views. On one hand, they could compare the two methods directly. On the other hand, opinions about one modality could have affected opinions about the other. For example, reflecting through EMA responses might have changed the interpretation of physiological signals.

Our participant sample was skewed towards people with higher-education degrees, including post-graduate degrees. While this is typical in information work, generalizing to other workplaces should consider demographics and the nature of work could influence perceptions of stress tracking.

Adherence, perceived benefits and challenges based on three weeks can differ from those of long-term tracking. However, in the three-week period some people already said they engaged more with their data in the beginning of the study and their interest started to fade away at the end, which could reflect trends of long-term tracking.

Participants were told that we are recruiting for "a research study about stress in the workplace. This study may help us to better understand events, interactions, and contexts that surround high-stress episodes." It is therefore possible that participants intentionally did not change their stress-related practices in order to provide a realistic picture of stress in their daily work for purposes of the research study. Future work can investigate whether providing guidance on how to leverage stress tracking for stress management influences employees' stress-related practices.

Finally, conducting thematic analysis without recording the interviews is a potential limitation. Although note-taking was as thorough and verbatim as possible, there still could be missed nuances that reviewing an audio recording after the interview can reveal.

## 5.5 Conclusion

This chapter addresses a gap between employees' experiences with workplace stress tracking, underlying benefits and challenges in self-tracking, and organizational implementation of stress tracking systems. While automated tracking with wearables can be comfortable and unobtrusive, employees might be less inclined to engage with their data. Participants used the devices to validate how stressed they felt. These brief interactions might be prone to confirmation bias where employees only confirm their perceived stress in certain moments and miss the opportunity for disconfirming evidence to learn deeply about their stress. Manual input methods encouraged reflecting on momentary stress and stressors, but challenges need to be addressed around quantifying subjective appraisal of stress and articulating stressors. Individual and organizational stress tracking goals and desired outcomes might be challenging to align. While objective and continuous data from wearables provides more detailed insights on employees stress patterns and can be correlated with other continuous data streams such as work activity tracking, individual employees might not gain much value from tracking with completely automated methods. On the other hand, manual methods generate stress data that could be harder to analyze by the organization, harder to compare across employees, and has limitations in terms of frequent and long-term implementation, but might engage employees more with their data and encourage reflection that is likely to lead to awareness and action.

# Chapter 6

# Physicians' Electronic Inbox Work Patterns and Factors Associated With High Inbox Work Duration

## 6.1 Introduction

The electronic inbox forms a crucial hub for physicians to communicate with other clinicians, staff, and patients via electronic health record (EHR)–based messages. The ability to communicate with patients and families via email-like secure messages helps physicians build relationships [159, 234, 235, 42, 199, 233]. Electronic inbox management has become a progressively more important component of physicians' work as EHR adoption and patient use of secure messages have increased [3, 63, 260].

In business and information work, inbox management has been associated with stress and burnout due to the time required to handle the ever-increasing volume of emails, the task demands associated with emails, and the interruptions they create [239, 22, 177]. In the same way, electronic inbox management has been purported to contribute to physician stress and burnout [159, 284, 202]. To understand the relationship between EHR adoption and use and stress, it is critical to examine how physicians spend time on the EHR inbox.

Scant research on physicians' electronic inbox use patterns currently exists. Previous studies of daily EHR work patterns have evaluated overall EHR use throughout the day [308, 217, 16] but inbox management has received limited attention beyond studies that have quantified the time spent on it. To facilitate more adept use of electronic inboxes, it is critical to gain a more detailed understanding of how physicians manage them. Understanding the temporal patterns of inbox work and identifying factors associated with high duration of inbox work are key steps toward learning how to potentially reduce stress associated with inbox management. To this end, the study presented in this chapter aimed to:
(1) quantify the amount of time physicians spend on electronic inbox management,
(2) describe daily patterns of inbox use,
(3) describe variation in inbox work by the type of message received, and
(4) identify factors associated with high duration of time on electronic inbox work.

## 6.2 Study setting

Data collection was conducted at the Permanente Medical Group, one of the largest medical groups in the United States. The medical group has 9,200 physicians and serves 4.4 million members via 21 hospital-based medical centers and more than 250 medical offices in Northern California. Physicians are salaried, and each primary care practitioner has a defined panel of affiliated patients to manage.

Since 2008, the participating medical group has used a comprehensive EHR (Epic Systems, Verona, WI) that integrates inpatient, emergency, and outpatient care, including primary care, specialty, laboratory, pharmacy, and imaging data. The EHR inbox is the primary message hub that physicians use for clinical care. The EHR inbox receives messages patients send via a portal website (also available through patient-facing mobile applications), as well as messages from other physicians, clinical staff, the pharmacy, laboratory, and other departments. Physicians can access the inbox on computers or mobile devices. Physicians are expected to reply to each patient message within 2 business days. Patients are encouraged to use the messaging functionality of the EHR to enhance access to their physicians and the care experience.

The practice environment was relatively consistent across physicians in this study. Clinical work hours when clinical settings are open and patient appointments are booked were Monday to Friday, 8:30 AM to 12:30 PM and 1:30 PM to 5:30 PM. Clinical work hours were scheduled predominantly with patient appointments conducted in person or via telephone or video telemedicine. Within clinical work hours, some departments earmarked periods of time without scheduled patient visits called "panel management time," to protect time for physicians to manage inboxes and do other types of work. Physicians sometimes worked clinically on weekends, with work hours that could differ from weekdays.

This setting's patient population was diverse, with 54% of adults being White, 21% being Asian, 16% being Latino, and 6% being Black. Approximately half of adult members had a college degree, more than three-fourths were employed or self-employed, and most lived in urban or suburban areas [89].

## 6.3 EHR system logs

We used system access logs, which contained granular timestamped data on the Epic system EHR use. These logs contain second-by-second records of the EHR components physicians have accessed and records of the associated actions performed. The logged data for the inbox includes the type of message being accessed, the time, user ID, and the access mode (desktop or mobile). We created hourly time bins and variables from the log data to quantify how time was attributed to different activities and different types of inbox messages per hour. These variables include, for every hour, the number of minutes spent in the EHR, the number of minutes spent in the inbox, the number of minutes spent working on each inbox message type, and the number of window switches (i.e. clicking on a new computer window/page). The logged time spent on the inbox only included time spent on the inbox page, and did not count time spent on other EHR components that physicians might use to gather information to respond to messages. Hence, the inbox time in this analysis was an underestimate of the total time physicians spent on inbox-related work.

EHR log data were extracted from March 1 to 31, 2018, for all internists and family practitioners who were doing at least 70% fulltime equivalent (FTE) outpatient clinical practice at the time of the analysis. We chose March because it is a representative month: neither extremely busy due to influenza season nor slow due to summer vacations. We also obtained physicians' age, sex, years in practice, and their FTE in clinical practice, a measure of their workload in scheduled clinical hours where 40 hours per week of scheduled work is 1.0 FTE. Internal analyses by this medical group reveal that FTE is strongly correlated with panel size. We thus include patient panel characteristics per physician (mean age, percentage over 65 years of age, percentage female).

After preliminary analyses of the 1275 physicians in the dataset, we excluded physicians with <10 working days in the month, and those with no time spent on the inbox on all workdays, leaving 1257 physicians in the study population. We defined a workday for a physician as a day with at least 1 scheduled in-person visit or more than 7 telephone and video visits. We chose this cut point based on the number of telephone encounters among physicians on weekends in March 2018 to properly classify days off when a physician may make a few telephone calls to relay lab results to patients while not scheduled to work. We removed weekend workdays (Saturdays and Sundays) from our analysis of workdays to better distinguish work and nonwork hours, as work hours can vary when physicians are working on weekends.

The system logs included a "message type description" field, with 59 different labels. We categorized these labels into higher-level groupings based on an analysis of frequencies and input from this study's clinical collaborators, who were familiar with the meanings and patterns of different message types. This approach resulted in four message types: (1) messages from patients, (2) results, such as lab test results, (3) requests, which ask the physician to perform an action such as approving a medication refill or signing clinical orders, and (4) Informational and administrative messages. No message content or metadata (ie, sender, receiver, id, or timestamp) were included in were included in this study, which was approved by the Kaiser Permanente Northern California Institutional Review Board.

## 6.4 Statistical analyses

To identify factors associated with duration of time spent on EHR inbox work, we compared physicians in the top and bottom quartiles of average inbox use duration, and created multiple regression models. To compare factors between these groups, we conducted a series of 2-group tests (t tests for normally distributed variables, Mann-Whitney U tests for nonparametrically distributed variables and chi-square tests for categorical variables). We corrected for multiple comparisons using the Bonferroni correction, testing each $P$ value against an alpha value of .05/18 tests = .003. For the regression models, the distributions of dependent variables were inspected to confirm the normality assumption. Multicollinearity was tested using the variance inflation factor, and independent variables with variance inflation factor >5 were removed. Analyses were conducted using the Scipy and Statsmodels packages in

Python.

## 6.5 Results

### 6.5.1 Participants

Of the 1257 physicians in the study, 57% were female. Their ages ranged from 29 to 72 years (mean 46.72, SD 8.43). Years since medical school graduation ranged from 4 to 45 years of age (mean 19.35, SD 9.04). Participants' workload ranged from 0.7 to 1.0 FTE (mean 0.85, SD 0.11). On average, physicians in the dataset had 20 (SD 3) workdays and 11 (SD 3) nonworkdays. The average physician in the study had a patient panel with an average age of 46.78 (SD 6.25) years, with 19 (SD 10%) of patients being older than 65 years of age, and 52 (SD 18%) being female patients. The average number of patient encounters per physician for the month was 335.26 (SD 95.1), of which 237.39 (SD 60.58) were face-to-face encounters and 97.87 (SD 57.05) were telephone and video encounters.

### 6.5.2 Time spent on inbox management

On workdays, the average time spent on inbox work was a total of 52 minutes: 33 minutes during work hours and 19 minutes outside of work hours (Table 6.1). Thus, 37% of total time on the inbox on a workday occurred outside of formal work hours. On nonworkdays, physicians spent an average of 12 minutes on the inbox, with a range of 0-93 minutes. On workdays, physicians spent most of their inbox time on desktop or laptop computers, whether within or outside of work hours. On nonworkdays, however, the inbox was mostly accessed through mobile devices. On average, physicians had 100 message views a day during work hours, and 53 views outside work hours (Table 6.1). On nonworkdays, physicians had an

| | Mode of inbox access | | |
|---|---|---|---|
| | Desktop/Laptop | Mobile | Total |
| Workdays | | | |
| Message views, mean (SD) | | | |
| During work hours | 94 (38) | 6 (11) | 100 (38) |
| Outside work hours | 36 (25) | 17 (20) | 53 (31) |
| Total | 131 (47) | 23 (25) | 153 (47) |
| Time spent (min:sec), mean (SD) | | | |
| During work hours | 30:55 (11:34) | 2:01 (3:34) | 32:56 (11:14) |
| Outside work hours | 12:45 (9:31) | 6:09 (7:27) | 18:49 (11:52) |
| Total | 43:37 (15:29) | 08:10 (08:59) | 51:47 (15:30) |
| Non-workdays | | | |
| Message views, mean (SD) | 4 (13) | 29 (31) | 33 (34) |
| Time spent (min:sec), mean (SD) | 1:12 (4:30) | 11:07 (12:33) | 12:19 (13:16) |

Table 6.1: Electronic inbox message views and time spent per day by primary care physicians in the Permanente Medical Group, March 2018.

average of 33 message views daily. It is important to note that we cannot ascertain if these were unique message counts, or if they were the number of times physicians started a period of inbox work; rather, they were the number of times physicians switched into an inbox page. For example, it would count as 2 message views if a physician viewed a message, then shifted to another EHR system page to retrieve information, then returned to reply to the same message.

Of all time spent on the EHR inbox, physicians spent 28%, 29%, 25%, and 11% on patient messages, results, requests, and administrative messages, respectively (Table 6.2).

| Message type | Workdays | Non-Workdays | Overall (%) |
|---|---|---|---|
| Patient-initiated messages | 14:56 (6:49) | 2:24 (3:14) | 28 (10) |
| Results | 14:01 (5:20) | 4:45 (5:45) | 29 (6) |
| Requests | 13:21 (5:44) | 2:00 (2:33) | 25 (8) |
| Admin | 5:30 (2:44) | 1:51 (2:32) | 11 (5) |
| Other | 2:48 (1:37) | 1:16 (2:03) | 6 (3) |

Table 6.2: Average (SD) time spent on each message type during workdays and non-workdays. The overall column indicates mean percentages of total inbox time over the month.

### 6.5.3 Daily patterns of electronic inbox work compared with other EHR work

Figure 6.1 contrasts the daily patterns of time spent on the inbox component of the EHR with time spent on other components of the EHR. For many physicians, the most time spent on the EHR inbox was before and after formal work hours, and during the lunch hour (Figure 6.1a). Averaging across all users showed small peaks at the beginning and end of the workday. In contrast, time on other EHR components followed a different pattern. As Figure 6.1b shows, time spent on EHR functionality other than the inbox (including chart review, order entry, and charting) increased during work hours compared with nonwork hours. There was a sharp decrease in EHR use during the lunch hour as well. On nonworkdays, both overall EHR and inbox use had flatter patterns than workdays. As can be expected, there is a large variance across individual users in the amount of time spent on the EHR and the inbox per hour.

### 6.5.4 Factors associated with high duration of time on inbox work

To further explore different patterns of inbox use, we first compared physicians in our dataset in the top and lowest quartiles of average duration of time spent on the inbox on workdays

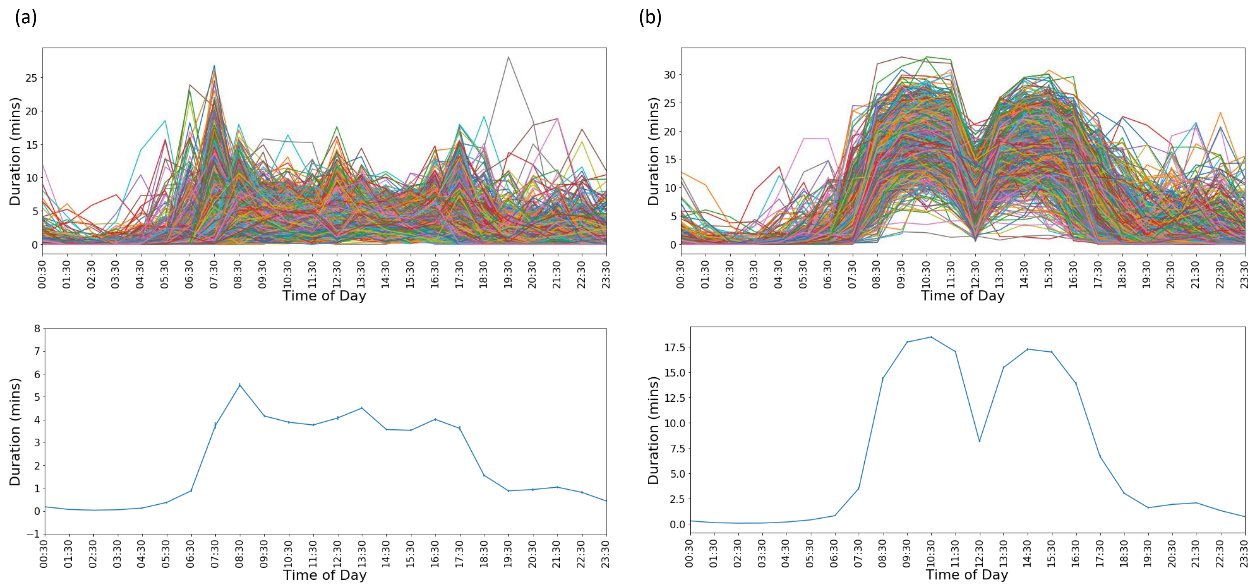



Figure 6.1: Time spent on (A) the EHR inbox and (B) EHR functionality other than the inbox. Top figures show daily averages for each user (1257 users) and the bottom figures show overall average across user averages.

(during and outside of work hours). This segmentation resulted in 314 users per group. The high-duration group (top quartile) had an average inbox duration of 72 (SD 9) minutes per workday, while the low-duration group (lowest quartile) had an average of 33 (SD 7) minutes.

Table 6.3 shows these comparisons. The high-duration user group had a higher proportion of women (62%) compared with lowduration user group (50%) ($P$=.003). There was no statistically significant difference between high- and low-duration users in the number of days worked or FTE. Both groups had an average of 19 (SD 3) workdays during the month and an average FTE of 0.85 (SD 0.1). However, high-duration users had more patient encounters, older patients, and a higher percentage of patients older than 65 years of age and female patients.

High-duration users spent more than twice the duration on messages after work hours on workdays (30 minutes vs 9.7 minutes; $P$ <.001), compared with users in the low-duration group (Table 6.3). Comparing the ratio of after-hours inbox use duration to all day inbox use duration, we found that a larger average proportion of time on the inbox occurred outside of work hours for high-duration users (41%) compared with low-duration users (29%) ($P$ <.001). We also found differences in the time spent on the inbox on nonworkdays. On nonworkdays, high-duration users spent almost double the time on inbox work relative to low-duration users. Users in the high-duration group averaged almost twice the message views on workdays (199 views vs 109 views; $P$ <.001), and spent slightly more time per message view (22.3 seconds vs 18.9 seconds; $P$ <.001). The duration of time spent on the inbox on mobile devices was also more than doubled for high-duration users, compared with low-duration users, although the ratio of inbox time on mobile to all inbox time was similar between the groups, as both groups spent 23%-25% of their total EHR inbox time on mobile devices.

To identify factors independently associated with inbox work duration, we created multiple linear regression models for 2 dependent variables: all-day inbox use duration and after-hours inbox use duration. Both measures were averages over each physician's workdays. Independent variables included physician characteristics (age and sex), patient panel characteristics (age), patient encounters (face-to-face and telephone and video visits), average percentage of inbox time spent on patient messages, average time per message view, and average percentage of inbox time using mobile devices. For predicting after-hours inbox time, work-hours inbox time was also added as an independent variable. Owing to multicollinearity, we excluded patient panel percentage of female patients, percentage of patients over 65 years of age, physician years of experience, and FTE. Model results are presented in Table 6.4.

For the model of all-day inbox duration all predictors, except physician age, had a positive relationship with inbox duration ($F = 63.71$, $P$ <.001, $adjR^2 = .29$). The average age of the patient panel, the number of face-to-face patient encounters, and time per message view had the largest effect sizes (standardized coefficients >0.3). For predicting after-hours inbox duration, the analysis showed a negative relationship of the more time physicians spent on

| | High-duration inbox users (n=314) | Low-duration inbox users (n=314) | $P$ value |
|---|---|---|---|
| **Physician characteristics** | | | |
| Age, y | 47.2 (8.5) | 46 (8.6) | .098 |
| Years since medical school graduation | 20.3 (9) | 18.2 (9.3) | .005 |
| Female, % | 62 | 50 | .003* |
| **Workload** | | | |
| Workdays | 18.9 (3) | 18.8 (3) | .286 |
| FTE | 0.85 (0.1) | 0.85 (0.1) | .291 |
| Total number of patient encounters | 347.1 (82.7) | 311.7 (112.7) | <.001* |
| Number of face-to-face patient encounters | 247 (57) | 216 (68.8) | <.001* |
| Number of phone/video patient encounters | 100.1(42.1) | 95.8 (66.8) | .005 |
| **Patient panel characteristics** | | | |
| Patient age, y | 48.3 (5.9) | 44.8 (6.4) | <.001* |
| Patients older than 65 y, % | 21 (10) | 16 (9) | <.001* |
| Female, % | 54 (18) | 48 (18) | <.001* |
| **Temporal patterns of inbox work** | | | |
| % of after-hours inbox duration to total inbox duration on workdays | 41 (15) | 29 (16) | <.001* |
| After-hours inbox duration on workdays, min | 30.1 (13.1) | 9.7 (5.6) | <.001* |
| Inbox duration on nonworkdays, min | 16.7 (14.9) | 8.5 (10.5) | <.001* |
| **Message view patterns** | | | |
| Message views per workday (including after hours) | 199.9 (42.1) | 108.8 (31.2) | <.001* |
| Time per message view, s | 22.3 (3.8) | 18.9 (3.7) | <.001* |
| **Use of mobile devices** | | | |
| Duration of inbox work on mobile devices per workday, min | 12 (11.5) | 5 (5.8) | <.001* |
| % of all inbox time spent in mobile device use | 25 (18) | 23 (22) | .015 |

Values are mean (SD), unless otherwise indicated.

FTE: full-time equivalent.

*Significant at $P <.003$ based on the Bonferroni correction.

Table 6.3: Comparisons of high-duration and low-duration users of the electronic health record–based inbox among primary care physicians in the Permanente Medical Group, 2018.

the inbox during work hours, the less time they spent on the inbox after hours. A negative relationship was also found between physician age and after-hours inbox use, although the effect size was small. Face-to-face patient encounters, time per message view, and percentage of inbox time that is on mobile devices had the largest effect sizes (standardized coefficients >0.3). Female physicians and physicians who had older patients spent more time on the inbox after hours. The model explains 33% of the variation in after-hours inbox time ($F = 68.36$, $P < .001$, $adjR^2$=.33).

| | All-day inbox work | | After-hours inbox work | |
|---|---|---|---|---|
| | Coeff (SE) | p | Coeff (SE) | p |
| Physician characteristics | | | | |
| Female | 306 (46) | <.001 | 159 (34.64) | <.001 |
| Age | -17 (3) | <.001 | -6 (2) | .018 |
| Patient panel characteristics | | | | |
| Mean age | 50 (4) | <.001 | 30 (3) | <.001 |
| Face-to-face appointments | 126 (9) | <.001 | 101 (7) | <.001 |
| Phone/video appointments | 33 (10) | .001 | 21 (7.72) | .007 |
| Inbox use patterns | | | | |
| % patient messages | 1297 (226) | <.001 | 485 (170) | .004 |
| Time per message view | 76 (5.76) | <.001 | 55 (4) | <.001 |
| % on mobile | 347 (145) | .017 | 1431 (113) | <.001 |
| Workhours inbox use | – | – | -0.14 (0.03) | <.001 |

Table 6.4: Regression models predicting duration of all-day inbox work (per 24-hour period) and after-hours inbox work.

## 6.6 Discussion

### 6.6.1 Major findings

To our knowledge, this study is the first to describe how electronic inbox work fits temporally into the workdays of PCPs. We found that, on average, PCPs spent 52 minutes on electronic inbox work on workdays, and that more than one-third of this time occurred outside formally scheduled work hours. This study also was unique in our ability to identify characteristics of physicians who spent higher than average amounts of time on inbox work. Those with the highest duration of inbox work were more likely to be female, have a higher percentage of female patients, have older patients, have more patient encounters, and be doing inbox work after hours, as well as spend more time per message view and spend a higher proportion of their inbox time on patient-initiated messages.

### 6.6.2 Interpretation and comparison with past studies

Previous studies of electronic inbox use have described the amount and timing of inbox work but have neither studied temporal patterns in depth nor described variation among physicians. A 2017 study found that PCPs spent an average of 5.9 hours per workday on EHR systems, including 1.4 hours outside clinic hours, with 24% of the total time spent in the inbox and a higher proportion of inbox time spent after hours compared with other EHR activities [16].

Our study went beyond previous studies [16, 308, 217] in comparing temporal patterns of electronic inbox use as distinct from use of other parts of the EHR (eg, chart review, order entry, creating notes). We found that inbox use patterns did not mirror work hours, but rather increased before and after work hours and during lunch hours. We also found that the more time physicians spent on the inbox during work hours, the less time they spent on the inbox after hours. These findings were not surprising because during workdays, PCPs usually are scheduled to see patients via in-person, telephone, or video visits for most or all of the available time. Electronic inbox work during formal work hours is typically fit in between patient visits, and work that cannot be finished during those hours is usually addressed just before work hours, just after work hours, or during lunch. Given that the average time spent on the inbox after hours is 19 minutes, which is small relative to the total time spent on EHR activities in other studies, it is possible that some physicians may be prioritizing inbox work during work hours and completing notes outside work hours. Future work analyzing the trade-off between different EHR actions within and outside of work hours could clarify this supposition.

This study found that physicians switch to view electronic messages 100 times a day. This number is higher than previous studies of information workers, who checked their emails an average of 77 times a day during work hours [176]. The high number of views may be due to physicians needing to switch screens between the inbox and other parts of the EHR to find information to reply to messages or to take other actions, indicating a high rate of task switching within inbox work itself. Given the associations between multitasking and stress [178, 236], physician patterns of multitasking with the electronic inbox warrant further study.

Our finding that physicians who spent more time on inbox work tended to be female is novel, and consistent with a recent finding that female physicians tend to spend more time on the EHR [17]. This contrasts with another recent study in which women reported less EHR-related stress and higher efficiency than men [130]. Studies of physicians have suggested that burnout symptoms may be associated with being female, and that work-home conflict may play a role [284, 114, 118]. Our observation that female physicians tend to spend more time on inbox work suggests that this deserves further exploration as a possible factor in sex differences in physician stress.

We also found that high inbox work duration was associated with more work outside of work hours. Not only was the absolute amount of time spent working after hours longer, but also the proportion of all time spent on the inbox that occurred after hours was higher. Research in office settings has found spending more time on email associated with greater work overload, frustration, and stress [22]. Our finding that inbox work tends to extend beyond work hours for high-duration users suggests that further study is warranted regarding how to best support this group.

In this study, inbox work duration was independently associated with clinician workload as measured by the number of appointments seen during the month studied. This is in accord with a previous finding that physicians with more clinical time were disproportionately burdened by after-hours EHR work [247], and another study that found that work relative value units (ie, work volume and complexity) were positively associated with EHR time within and outside of work hours [17]. In contrast, we did not find that lower FTE was associated with reduced inbox work. This finding may reflect that physicians who adopt reduced clinical FTE schedules may not experience a commensurate decrease in their amount of electronic inbox work, or conversely, that physicians who feel overburdened may adopt reduced FTE schedules to allow themselves more time to complete clinical work in general, including inbox work.

A recent study found that provider-to-provider variation was the largest source of variation in after-hours EHR usage [17]. Similarly, the wide variation we found in inbox work duration and use patterns among PCPs in this study suggests that individual preferences and approaches most likely play a role in inbox work patterns. Future research could attempt to identify different types of users based on the strategies they adopt (eg, batching inbox message work [176]), the temporal patterns of their work, and their use of mobile devices for inbox work. In addition, future work could investigate the effect of organizational efforts such as designating time for panel management within clinical work hours. It would be useful to study whether inbox use patterns are associated with physician stress or productivity, or patient satisfaction with electronic message communication with their PCPs.

### 6.6.3 System design and organizational implications

The results suggest practical implications for inbox system design. Given the high number of message views (i.e. counts of switching to the inbox page), a system design consideration would be to implement an interface that incorporates information that physicians need from sources outside the inbox page (e.g. patient data) to process an inbox message, thus reducing potential frequent switching between the inbox page and other windows.

Batching (i.e. attending to the inbox once or twice a day rather than consistently checking messages throughout the day) and checking the inbox fewer times has been found to be associated with less time on email and less stress [177, 176, 145]. While email batching may be desirable in some work settings, our study's results show that the temporal patterns of inbox use are different in medical settings, in which physicians spend smaller amounts of time on the inbox during periods when they are scheduled to see patients. It is possible that the messages being checked in between patient visits are those that are more clinically urgent, making batching infeasible. Thus, the feasibility and desirability of batching inbox time in medical contexts is yet to be evaluated. Another practical implication for inbox system design is to implement screening and categorization of patient-initiated messages, automatically or by assistants, which can help PCPs prioritize or delegate some messages.

### 6.6.4 Limitations

This study was conducted in a large medical group that encourages patients to use EHR portal messages to communicate with physicians. The group also tries to limit the amount of system-generated messages and administrative reminders sent to physicians' inboxes. Thus, this setting's volume of inbox messages from patients may be higher and the balance of different types of messages may differ from those in less integrated settings.

The dataset for this study had message views but not unique message counts. Thus, we did not analyze the volume of messages, time spent per message, or how many times a message was revisited. As noted in the Materials and Methods, inbox work duration counted time the inbox window was open, but did not count (for example) time when the inbox window was in the background while the user was accessing other parts of the EHR such as chart review or order entry in response to a message. Thus, the duration of inbox work in this study is likely an underestimate of the time physicians spent on inbox-related work.

Finally, our analysis did not control for panel management time (ie, time designated by departments specifically for tasks like inbox management). Thus, we cannot make assumptions about why inbox work patterns and peak use times differ among physicians. Our analysis shows an inverse relationship between time on the inbox during and after work hours. It is possible that physicians who have dedicated panel management time during work hours are those who spend less time on inbox work after hours. The effect of panel management time on inbox use patterns needs further study.

## 6.7 Conclusion

We conducted a large-scale study of physician's EHR inbox daily work patterns. Physicians in the largest medical group in the United States spend roughly an hour per workday on inbox management, and much of this work occurs outside scheduled work hours. In this setting, patient-initiated messages and results consume the highest proportion of inbox work time. Interventions to assist physicians in handling patient-initiated messages and results may help alleviate inbox workload.

# Chapter 7

# Physicians' Stress and EHR Inbox Work Patterns

## 7.1 Introduction

Inbox management is an important component of electronic health record (EHR) work for physicians and a key potential stressor [202]. Although several studies have addressed workload or burden related to EHR use, no study to date has directly measured stress related to EHR inbox use. There are two main limitations in prior work. First, scant research exists that focuses on the inbox component of the EHR (e.g. [202, 284, 93, 204]). Second, previous studies relied on self-reported proxies of stress (e.g. burden, burnout, wellbeing) measured at a single time-point (or a few time points) (e.g. [284]), which fails to capture the detailed and continual stress and EHR work patterns throughout the day and is prone to bias [94, 252].

This chapter investigates physicians' EHR inbox use patterns and associated stress, as measured unobtrusively and continuously by EHR system logs and wearable sensors. The objectives of this work were to:

(1) collect EHR use and stress data through unobtrusive means that provide objective and continuous measures;

(2) cluster and visualize distinct EHR inbox work patterns, and identify their associated characteristics;

(3) identify physicians' daily stress patterns; and

(4) evaluate the association between EHR inbox work characteristics and physician stress.

## 7.2 Recruitment and protocol

The study setting is detailed in Chapter 6. We recruited adult primary care physicians (PCPs) from five medical facilities within of the largest medical groups in the United States. To identify facilities for recruitment, one of the study's clinical collaborators sent an email to local adult and family medicine department chiefs in June 2019 describing the study and asking about interest in participating. In July through September 2019, a clinical collaborator gave a recruitment presentation in-person or remotely at a department meeting at each of five interested facilities, with a recruitment email sent immediately thereafter to all department physicians at the facility. Between 7 and 12 physicians were enrolled at each facility, with a total of 47 eligible physicians enrolled. Physicians were eligible if they did outpatient clinical work at least three and a half days a week. Physicians who were taking cardiac medications, had pacemakers or defibrillators, or had been diagnosed with cardiac arrhythmias were not eligible due to the interference of these factors with the HRV-based stress measure. Eligibility was confirmed via a recruitment email.

After obtaining written informed consent, the staff assigned a wearable device with heart-rate sensors (Garmin Vivosmart 3) and configured the associated mobile applications (Garmin Connect and Tesserae Phone Agent [183]) on the physician's work-issued mobile phone. The apps streamed data from the wearable device via Bluetooth and uploaded the data to a server. The research team also installed an experience sampling app [122] on the physician's mobile phone to send short questions at specified times (see Chapter 3: Methods). At enrollment, physicians were asked to complete a 5-question survey on their strategies for and feelings about Inbasket (their EHR inbox) management. Physicians were asked to indicate how distressful they found inbox management and whether they had responsibilities that restricted their ability to work before or after formal work hours.

Physicians were asked to wear the device and respond to the daily short survey prompts for seven consecutive days, and to keep their phones and the wearable device charged with Bluetooth enabled. Physicians were free to keep their wearable devices after the end of data collection. The study was approved by the institutional review board of Kaiser Permanente Northern California.

## 7.3 Analysis

Data for this study included EHR system logs and physician characteristics (see Chapter 6), as well as physiological and self-reported stress (see Chapter 3). We used the Gaussian Mixture Models clustering algorithm [240] to find distinct patterns of inbox work. Features in the model included the distribution of inbox time in work hours and outside of work hours, contiguous and non-contiguous to work hours. Multiple feature and cluster counts were tested and the clustering that yielded more balanced clusters and had a reasonable silhouette score (a score that indicates how distinct or overlapping the clusters are) [276] was selected.

To capture whether physicians dedicated certain blocks of time for inbox work or consistently checked their inbox throughout the day, we defined days with inbox work batching as days where 70% or more of the total inbox work duration occurred in three separate blocks of time or less. With consistent inbox checking, a uniform distribution of inbox duration over the day would typically be observed, while batching would show 2-3 daily peaks of high inbox duration [176]. We compared this measure across clusters and also used it as an independent variable in the mixed-effects model along with the other EHR inbox use characteristics.

To compare clusters (ie, groups of different inbox work patterns), each comparison variable was tested for normality and homogeneity of variances before conducting an analysis of variance for normal distributions with equal variances or the Kruskal-Wallis test otherwise. For pairwise comparisons, a posthoc analysis was conducted using the Tukey honestly significant difference test for normally distributed variables and Dunn test for nonparametric posthoc comparisons. Categorical variables were tested using the Chi-square test.

To plot hourly stress patterns, we removed hours with less than 20 minutes of valid HRV data to avoid overestimating the stress duration as a ratio of the measurement period (the measurement period being valid HRV measurement duration). From a total of 4245 hours, this filter removed 1177 hours (27.73%) of the workdays' HRV data. For daily stress measures, workdays with less than 2 hours of valid HRV data were removed from the analysis, as well as workdays that are Saturdays or Sundays, and those with no inbox activity. This filter removed 21 days in total, keeping 178 workdays for the daily stress analyses (cluster comparison and a regression model).

We investigated the relationship between daily EHR inbox use and stress through a generalized mixed effects model with physicians as random effects. A Poisson distribution was used to represent stress minutes as events within the observation period (ie, valid HRV minutes as an offset in the model). The distribution of the dependent variable (ie, stress duration) was right skewed, as expected in a Poisson distribution. The independent variables were centered (ie, mean subtracted). The variance inflation factor was under 5 for all independent variables, indicating that multicollinearity was not a problem. Several models were compared, starting with a base model and incrementally adding variables, to ensure that the improvement in the model justified the added complexity of adding variables. The model with the lowest Akaike information criterion and highest marginal (fixed effects) $R^2$ is presented.

## 7.4 Results

### 7.4.1 Participants

The 47 physicians (68% female) had an average age of 43.83 years (SD 9.51; range 31-68) with 15.17 (SD 9.93; range 4-42) years of experience in medicine and an average FTE of 81% (SD 14%). On average, physicians in the dataset had 5.26 workdays (SD 0.94) and 2.74 non-workdays (SD 0.94) over the 8 days of data collection (the day of enrollment plus 7 days in the study).

The HRV-based stress analyses included 42 physicians, because 5 physicians (1 male and 4 female) had technical issues, thereby causing loss of the wearable device data. The inbox strategies and stress survey was completed by 44 physicians.

### 7.4.2 Three distinct patterns of EHR inbox work

On workdays, physicians spent an average of 3.5 hours (SD 0.69) in the EHR, of which 1.08 hours (SD 0.38) were spent doing inbox work. On nonworkdays, physicians spent an average of 23.88 minutes (SD 36.3) in the EHR, including an average of 13.78 minutes in inbox (SD 23.78). The majority of time in the inbox was spent on patient messages (mean 37%, SD 11%), followed by laboratory results (mean 31%, SD 8%), requests (mean 20%, SD 6%), and administrative messages (mean 13%, SD 5%).

Using the Gaussian Mixture Models clustering algorithm, we found 3 temporal patterns of work, with a silhouette score of 0.41, indicating moderate separation between these clusters (ie, distinct groupings). Figure 7.1 shows the average hourly time spent in the inbox and other EHR work (such as charting and order entry) for physicians in each cluster. Group 1 (n=10) represented physicians who spent time in the inbox outside work hours, in the evenings and early mornings; group 2 (n=17) represented physicians who worked mostly within work hours; and group 3 (n=20) represented physicians who spent some time on inbox work after hours that were mostly contiguous to work hours.

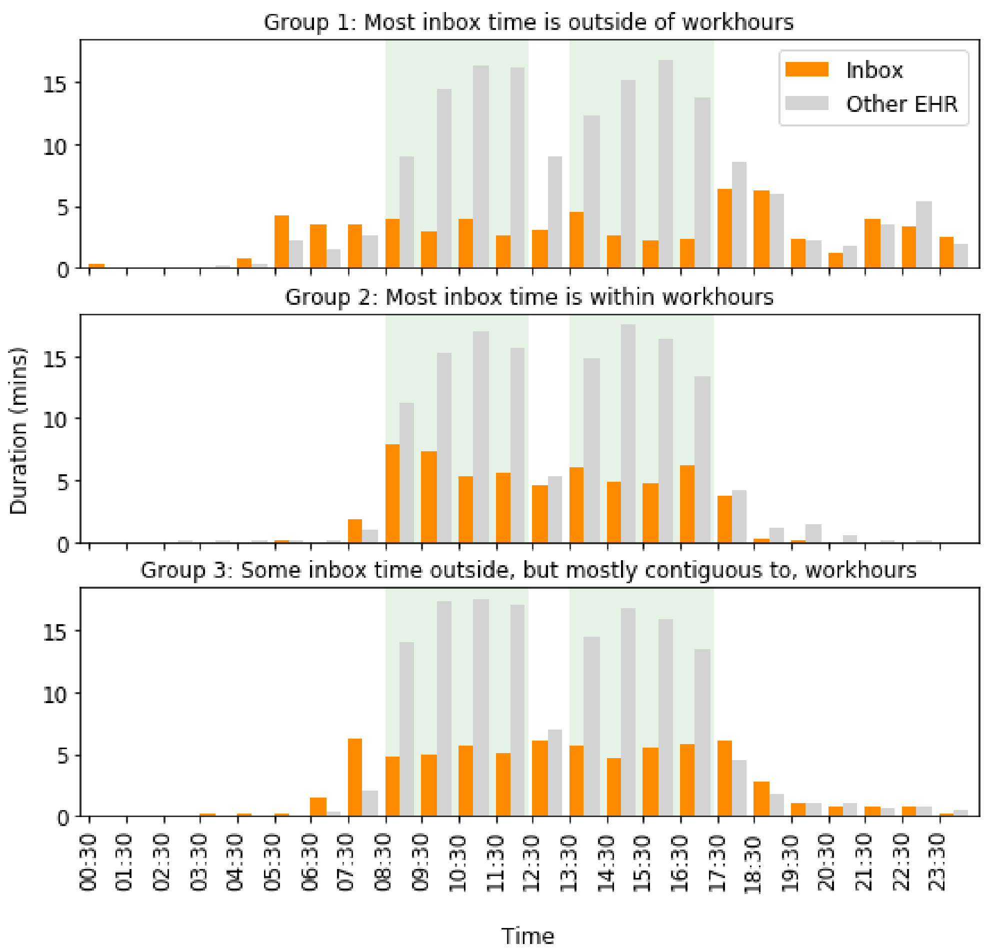



Figure 7.1: Temporal patterns of inbox and other EHR work. The green background indicates work hours.

Free-text responses from the survey on inbox management strategies supported these computationally generated inbox work patterns. Responses from physicians in group 1 indicated working beyond work hours, either by staying late in the office or taking work home. Some representative comments were as follows. A physician in group 1 reported, “I find when I sacrifice sleep to do more at home, I’m too tired during the day and I’m very inefficient at night,” indicating that they were working late at night. Physicians in group 2 indicated working mostly within work hours. For example, one physician in this group asserted, “I arrive around 8:30 and prefer to leave around 5:30.” Another stated: ”I just like to work and finish work during my allotted work time. I do not like to work at other times or at home.” Physicians in group 3 also indicated not taking work home but at the cost of staying late in the office to clear their inbox. For example, a physician in group 3 said, “I generally try not to take work home [...] so often stay very late to clean out inbasket.”

Physician characteristics (age, sex, years of experience, and FTE) did not show statistically significant differences across the 3 work patterns. In terms of EHR use, total daily time spent on inbox work and other EHR work on workdays (24-hour period) did not differ across groups ($P$ = .38 and $P$ = .15, respectively). However, as shown in Table 7.1, physicians in group 1 spent more time in the inbox after work hours compared with other groups, both in minutes and as a percentage of daily inbox time ($P < .001$). Posthoc comparisons showed that all the groups differed from each other. Group 1 also spent more time in the inbox work on nonworkdays ($P$ = .03).

Physicians in group 1 were more likely to batch their inbox work (ie, do most of their inbox work in a few chunks of time rather than consistently throughout the day) than group 2, as 50% (5/10) of physicians in group 1 batched their inbox work compared with 6% (1/17) in group 2 ($\chi^2$ = 4.03; $P$ = .045). The rate of switching windows within the EHR was not statistically different among the 3 groups ($P$ = .24), with all groups switching windows 4-4.5 times per minute of EHR use, on average. The groups spent different amounts of time per message ($P$ = .004). The time per message was higher for group 1 (mean 0.46 min, SD 0.11 min) than for group 2 (mean 0.35 min, SD 0.06 min) and group 3 (mean 0.38 min, SD 0.07 min). Groups 2 and 3 did not differ significantly ($P$ = .21). In terms of

| | Group-1 Mean (SD) | Group-2 Mean (SD) | Group-3 Mean (SD) | $P$ value |
|---|---|---|---|---|
| **Clustering factors (% of all-day inbox duration)** | | | | |
| Workhours inbox duration | 37 (12) | 82 (8) | 62 (9) | <.001 |
| Outside and non-contiguous to workhours | 42 (11) | 1 (2) | 12 (5) | <.001 |
| Contiguous to work hours | 21 (11) | 17 (7) | 26 (13) | .03 |
| **Duration of inbox work on workdays and non-workdays (mins)** | | | | |
| Workhours inbox duration | 25.36 (13.03) | 47.97 (13.35) | 42.13 (16.56) | .002 |
| Outside-workhours inbox duration | 41.37 (13.81) | 10.91 (5.63) | 26.97 (13.26) | <.001 |
| Inbox duration on non-workdays | 32.74 (37.46) | 11.13 (19.69) | 6.54 (11.3) | .03 |
| **Message types (% of all inbox time)** | | | | |
| Patients | 32 (10) | 35 (10) | 42 (10) | .02 |
| Results | 30 (9) | 32 (11) | 26 (10) | .1 |
| Requests | 24 (7) | 20 (6) | 21 (6) | .31 |
| Admin | 14 (5) | 13 (4) | 11 (4) | .14 |

Table 7.1: Comparing inbox use characteristics across three work patterns.

inbox message types, there were statistically significant differences among groups in patient-initiated messages ($P$ = .02), with group 3 spending a higher average percentage of their inbox time on patient-initiated messages than group 1, and no differences for other group pairs (Table 7.1).

### 7.4.3 Stress patterns

Visualizing stress patterns throughout the day showed that stress was high at the beginning of the workday. The first hour of work (8:30 AM to 9:30 AM) had an average stress duration of 35% of the hour (SD 26%; SE 4%). Stress then started to decrease until the lunch hour and increased again at the start of the afternoon clinic shift. Toward the end of the workday, the stress duration decreased. There was another increase in stress in the evening, followed by a decrease in stress at night and during typical sleep hours (Figure 7.2). This 3-wave pattern of daily stress was consistent across the 3 work patterns, although group 2 had their highest stress an hour earlier (ie, 7:30 AM to 8:30 AM) than the other groups (Figure 7.2).

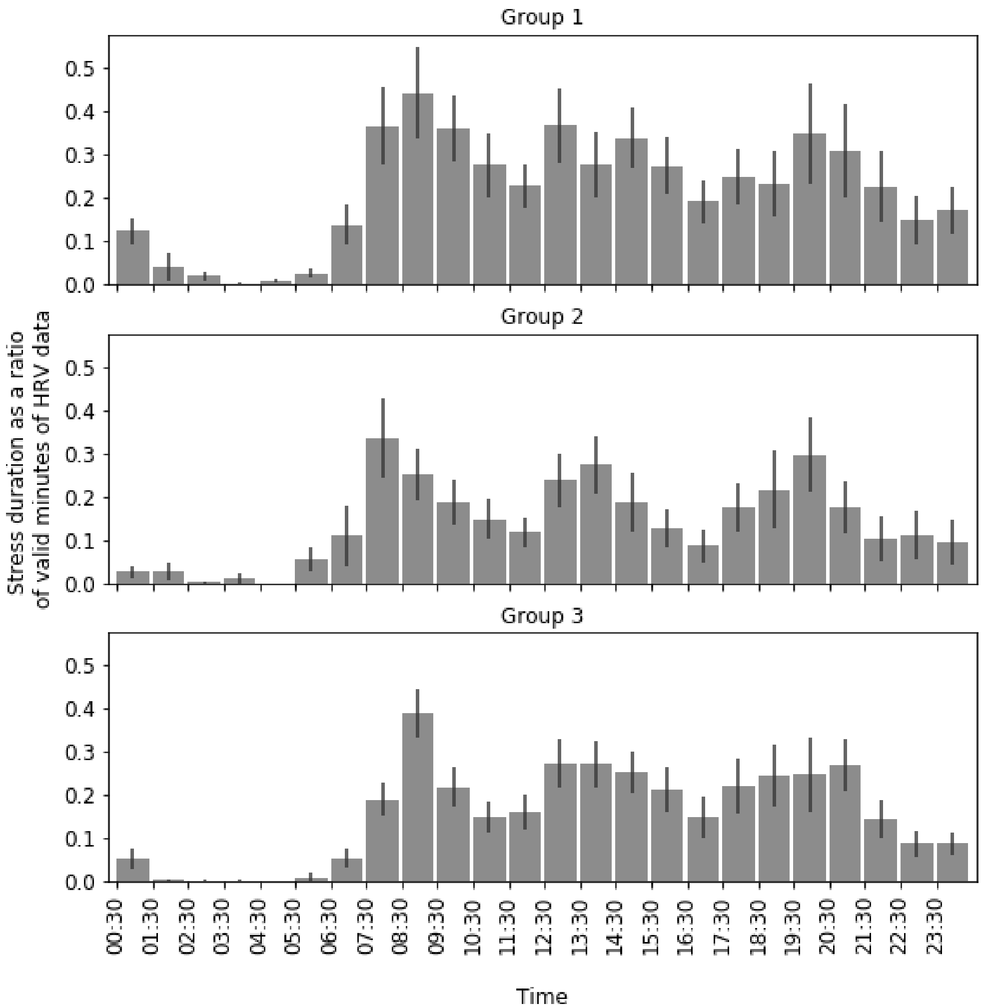



Figure 7.2: Workday stress pattern per group. Error bars represent the standard error of the mean (SE).

There was a difference in the average duration of stress during work hours among the groups (Kruskal-Wallis; $P = .02$). A posthoc comparison showed that group 1, the group with the highest after-hours inbox work duration, had a longer duration of stress during work hours than group 2 and group 3, with 33% (SD 27%) of work hours for group 1 being stressful (80 out of 243 min of valid HRV data indicated medium to high stress) compared with the 18% (SD 18%) for group 2 (47 out of 265 min of valid HRV data) and 22% (SD 24%) for group 3 (58 out of 265 min of valid HRV data). There was no significant difference between group 2 and 3 ($P = .73$). The number of valid minutes of HRV measurements was not significantly different across groups.

On average, physicians missed 45% (SD 20%; 9.4 out of 21) of the experience sampling prompts over the study period. Of the 485 submitted responses, 188 (38.8%) reported a stress level of over 50% (the midpoint of the slider). There was no significant difference in the average daily self-reported stress across the 3 inbox work patterns ($P = .99$).

Finally, in the survey on inbox management strategies and stress, physicians reported that 60% (SD 19%) of their work-related distress came from inbox management. Regarding the question of how distressful they find inbox management overall, of the 44 physicians, 19 (43%) said it was moderately stressful, 15 (34%) said it was very stressful, 6 (14%) said it was extremely stressful, and 4 (9%) said it was not very stressful. There were no statistically significant differences in survey responses across the 3 inbox work patterns.

### 7.4.4 EHR use characteristics associated with stress

We investigated detailed EHR use characteristics associated with stress using a mixed effects model, with workdays as the unit of analysis. The model showed that fixed effects accounted for 15% of the variation in duration of stress during work hours (Table 7.2). The physician's age, sex, and FTE worked were not associated with stress. The rate of switching windows when using the EHR was positively associated with stress ($P = .001$). Time spent on inbox work during work hours was positively associated with stress ($P < .001$), whereas time spent on other EHR activities during work hours was negatively (but very weakly) associated with stress ($P < .001$). Inbox work outside of work hours was positively associated with stress during work hours ($P < .001$). Interestingly, the proportion of inbox time spent on patient messages was not associated with stress. Surprisingly, batching inbox work for the day was also positively associated with stress ($P < .001$). Finally, days of the week were predictive of stress, with Mondays and Thursdays negatively associated with stress, whereas Tuesdays and Wednesdays positively associated with stress ($P < .001$ for each).

| Fixed effects | $\beta$ (SE) | Std $\beta$ | $P$ value |
|---|---|---|---|
| Full-time equivalent | 1.94 (1.39) | 0.27 | .16 |
| AGE | -0.01 (0.02) | -0.05 | .79 |
| Female | 0.45 (0.38) | 0.21 | .24 |
| Window switching rate | 0.1 (0.03) | 0.08 | .001 |
| Workhours inbox duration | 0.003 (0.001) | 0.08 | <.001 |
| Workhours non-inbox EHR duration | -0.002 (0) | -0.06 | <.001 |
| Non-workhours inbox duration proportion | 0.35 (0.07) | 0.09 | <.001 |
| Patient messages proportion | -0.09 (0.08) | -0.01 | 0.28 |
| Batching | 0.13 (0.03) | 0.06 | <.001 |
| Monday | -0.22 (0.04) | -0.1 | <.001 |
| Tuesday | 0.16 (0.03) | 0.06 | <.001 |
| Wednesday | 0.53 (0.03) | 0.2 | <.001 |
| Thursday | -0.13 (0.04) | -0.05 | <.001 |

The dependent variable is duration of stress during work hours.
Friday is the reference category for the variable day of week.

Std $\beta$ is the standardized coefficient.

EHR: electronic health record.

Table 7.2: Generalized linear mixed effects regression model.

## 7.5 Discussion

### 7.5.1 Principal findings

To our knowledge, this study is the first to measure physician stress using wearable sensors over several days of outpatient practice and the first to identify distinct EHR inbox work patterns and their associations with stress. Although the topic of EHR use and stress (specifically, self-reported burden, burnout, workload, and well-being) has been addressed in previous studies, this study is novel in that we measured stress unobtrusively and continuously through physiologic measures and used system logs to gain detailed insight about EHR use factors associated with stress. Higher rates of EHR window switching, longer inbox work duration, and a higher proportion of inbox work done outside of work hours were associated with higher stress. Daily stress patterns showed 3 waves of stress: in the first hour of work, at or after lunch hours, and in the evening.

In addition, we found that physicians fell into 3 groups with different patterns of inbox work. Some physicians tended to do most of their inbox work within work hours, whereas others did inbox work before or after but contiguous to work hours. The third group did inbox work in late evenings. These groups differed in characteristics such as inbox work batching, time per message, and the proportion of inbox time spent on patient messages. Physicians who did most of their inbox work outside of work hours were more likely to batch email and spend more time per message, whereas physicians who mostly do their inbox work within work hours were more likely to continually check their inbox throughout the workday, potentially in the short periods of time between patient appointments, and spent less time per message. The group that did most of their inbox work outside of work hours had the longest stress duration during work hours.

A strength of this study is that we measured stress using 3 different methods. The HRV-based stress provided a continuous timestamped stress measure that could be correlated with inbox use patterns throughout the day, the experience sampling measure provided momentary self-assessment of stress 3 times a day, and the survey provided a reflective measure on perceived overall stress related to inbox work. HRV-based stress differed across groups but self-report measures did not. It is well established in the literature that short-term self-reported (ie, perceived) stress and acute physiological stress do not always align linearly in daily life settings [123, 109, 200]; however, both are important to monitor as they both have health and well-being implications [243, 146, 298, 301].

### 7.5.2 Comparison with previous work

Previous studies on EHR use patterns have quantified the time spent on different EHR activities within and outside of work hours [16, 9]. However, variation among physicians is not well studied, and no previous study has attempted to characterize physicians based on their patterns of daily inbox use. One study [17] found that physician-to-physician variation explains most of the variability in EHR use time. We extend the findings on the variation in EHR use, focusing on inbox use and comparing physician characteristics across work patterns based on work hours and after-hours EHR inbox use. Aligned with previous findings [17], we did not find differences in physicians' sex distributions between the group with the longest after-hours inbox time and the group with the shortest after-hours inbox time. We also did not find differences based on FTE, contrary to previous findings [17] that more work relative value units generated by physicians (another measure of workload) were associated with more EHR time after work hours.

Most studies use basic measures to characterize EHR usage, such as the duration of time [16, 247, 11]. In one study, researchers used more complex measures to characterize mobile EHR usage, such as the number of log-ins and features used and usage paths (ie, the frequency and complexity of consecutive actions) [270]. They compared doctors across medical specialties and found that physicians other than surgeons had more diverse mobile EHR usage patterns with higher complexity and repetitive loops compared to surgeons [270]. In this study, we also used detailed EHR and inbox usage characteristics such as window switching, inbox work batching, the time per message, message types, and the time distribution between work and nonwork hours. Our finding that the window switching rate was positively associated with stress could reflect the complexity and repetitiveness of physicians' EHR interactions, as indicated in prior work [270], and the efficiency issues often associated with physicians' satisfaction with EHRs [320]. Another study on EHR inbox burden [204] also reported that excessive steps were needed to process messages and that physicians recommended reducing the number of mouse clicks necessary to process messages.

A recent study suggested a relationship between patient call messages and clinician burnout [110]. Their category of patient messages included all messages related to patient care tasks, such as phone calls, refill requests, and patient care forms. In our study, the category of patient messages included only patient-initiated messages and was not found to be associated with stress, although it comprised most of the inbox time for physicians.

It is not surprising that the differences among groups in HRV-based stress did not align with self-reported perceived stress. Previous studies have noted several issues in the interrelationship between perceived and physiological stress [111]. For example, the timing of the perceived stress prompt (before, during, or after a stressor event) could determine whether and how perceived stress correlates with physiological stress measured during the stressor event [254, 83, 215]. This has important implications for real-time stress monitoring for physicians, as it suggests that daily prompts to measure perceived stress in situ could fail to capture physiological stress. Increased and prolonged physiological stress reactions are associated with several health and well-being risks [160].

The results also suggest practical implications for organizational changes and system design. Previous studies have recommended a fundamental redesign of the EHR to improve data entry and retrieval [54]. On the basis of our finding that window switching is associated with stress, a redesign that minimizes the need to navigate to different windows to record or obtain information may be beneficial. For example, contextual information for inbox messages can be made visible from the inbox [204]. Our findings lend support to recommendations from a previous study to automate frequently performed actions such as message routing and leverage team support for inbox management [204]. Allocating time for inbox management within work hours, also recommended in a previous study, may also help reduce stress [204].

### 7.5.3 Limitations

In this study, the regression model with EHR use characteristics explained 15% of the variation in duration of stress during work hours, which is a considerable proportion given the myriad factors that can potentially influence stress. However, stress was likely to have also been influenced by other variables that were beyond the scope of this study. In addition, the associations we observed between stress and window switching, inbox work duration, and inbox work outside work hours do not necessarily prove that the latter factors cause stress. It is possible that physicians who are busier during work hours have more stress and also make more window switches, have more inbox work, and have to do more inbox work outside work hours.

HRV-based measures are affected by several factors, such as health and physical activities. Although we tried to control these effects with our participant inclusion criteria and by removing periods that had physical activity registered by the wearable device, it is possible that carry-over effects of physical activity are still present in the HRV data of sedentary moments. Moreover, removing periods with physical activities could have removed periods when psychological stress was experienced. For example, walking to an important meeting could be mentally stressful but it will not be captured in our data because of the elimination of periods when walking is detected.

HRV data were excluded during periods of physical activities and were occasionally missing because of sensors losing contact with the skin. We set a minimum threshold (measurement period) of 20 minutes of valid data per hour for hourly stress measures and 2 hours for daily stress measures. Although not complete, we do feel that this is a reasonable proxy for the stress experience of that hour and day and a reasonable mitigation method for missing data.

Inbox use patterns might differ from one setting to another based on the organization's policies and norms. For example, the medical group where this study was conducted encouraged patients to use EHR portal messages to communicate with physicians. Simultaneously, system-generated messages and administrative reminders are kept to a minimum whenever possible. Thus, the distribution of different message types may differ from that in other settings. These factors must be considered when generalizing our findings.

Finally, some physicians might have had panel management time (ie, time designated by departments specifically for tasks such as inbox management) incorporated within their work hours. In this study, we did not have access to data on panel management time. Thus, we cannot make assumptions about why inbox work patterns differed among physicians. We can only report the relationship of these different work patterns with stress.

## 7.6 Conclusion

This study is the first to use continuous and unobtrusive measures of stress to evaluate associations between EHR inbox use and stress among physicians. A total of 3 potentially modifiable factors were associated with stress: window switching, inbox work duration, and inbox work outside work hours. These findings have implications for research and organizational policies on stress measurement and EHR inbox management time and EHR system design.

# Chapter 8

# Discussion and Conclusion

## 8.1 Summary of findings

The studies presented in this dissertation investigated computer use factors associated with stress in two working populations (information workers and physicians) by using unobtrusive computer logging and stress tracking methods in their real work environments. With information workers, we aimed to uncover everyday computer interactions (active computer time within and outside work hours, computer work patterns, time on work and non-work computer activities) that are associated with daily objective measures of stress, and to identify individual factors that influence this association. We also aimed to introduce a novel measure related to computer use, which is the regularity of work-related computer use that reflects day-to-day fluctuations in workload, and we assessed the association between this measure and various perceived and objective stress measures. Employee perspectives on technology-supported stress tracking were evaluated with the aim of discovering the perceived benefits and challenges for different stress-tracking modalities in real workplace environments, and providing actionable recommendations for organizations and designers. With physicians, we focused on an emerging stressor in their Electronic Health Record (EHR) systems work, which is increased inbox management demands. We aimed to quantify the time primary care physicians spend managing inboxes, describe daily patterns of their inbox use, investigate which types of messages consume the most time and identify factors associated with inbox work duration. We also aimed to cluster distinct patterns of EHR inbox work, identify physicians' daily stress patterns, and evaluate the association between EHR inbox work patterns and physicians' physiological stress.

Information workers experienced stress for 22% of the workday (duration of medium and high stress to the duration of HRV recording) compared to 33%, 22% and 18% for physicians across the three different work patterns. The three work patterns for physicians, identified by clustering physicians based on their temporal inbox work patterns, were: (1) most inbox work is done outside of work hours, (2) most inbox work is done within work hours, and (3) some inbox time outside, but mostly contiguous to, work hours. There was no difference in EMA reported stress across the three work patterns for physicians, unlike HRV-based stress which showed a difference, where the group with most inbox work done outside of workhours experienced more physiological stress. Information workers reported high perceived stress in 37% of their EMAs, and physicians reported high perceived stress in 39% of their EMAs. Information workers missed 33% of their EMAs while physicians missed 45%.

Although information workers spent longer time working on the computer compared to physicians (4.6 vs. 3.5 hours), the proportion of computer duration that was spent doing inbox work was similar (33% vs 31%). For physicians, 37% of email work was done outside of work hours and the patterns of email work differed from the patterns of other EHR work. Physicians also worked more outside of workhours compared to information workers, although this might be explained by how work hours are defined differently for the two populations. For physicians, workhours were fixed, defined by the clinic hours (8:30 AM to 12:30 PM and 1:30 PM to 5:30 PM). Any work outside of these hours was counted towards the non-workhours work duration. For information workers, work hours were defined by the participants' self-reported "typical" hours as their hours are somewhat flexible, so if it were typical for an employee to arrive to work late and leave later than their formal hours, any work outside these typical hours would count towards the non-workhours work duration. Typical hours reflected the hours usually needed to finish their work tasks, therefore observing computer work activity outside of those hours was not common in our data. Another factor contributing to the observed difference in outside-hours work between information workers and physicians is the access mode to work systems. As we did not track information workers' phones or personal laptops, any work outside of work hours from their personal devices is unaccounted for, unless participants remotely accessed their work computers from their personal devices over a private network, which was captured by our logging software. Participants were asked whether they use personal devices for work, and most participants said they only occasionally check work emails on their personal devices. For physicians, access logs to the EHR system were recorded whether they used their work computers or personal devices, thus fully capturing their EHR work within and outside work hours.

Physicians had a higher rate of window switching, with 4-4.5 switches per minute, compared to 1.36 switches per minute for information workers. The study of information workers also quantified time spent on non-work computer activities (e.g. social media, music, shopping) and the day-to-day variability of computer work duration.

To our knowledge, this dissertation is the first to quantify this range of computer use factors and identify their independent association with physiological stress in two working populations. Among the interesting findings is that time spent on email work was associated with more stress for physicians, but not for information workers. The reason for this is unclear, but it is possible that the importance of email to the job could be a factor [64]. The medical group in this study minimizes informational and administrative messages, and primarily uses the EHR inbox for clinical care, as shown by our message type analysis (chapter 6). Information workers on the other hand might use email for a wider set of purposes, including personal communications. The perceived importance of email to the job depends on the job role, with higher management responsibilities being related to higher email importance. Working on many projects and collaborating with others such that one's work depends on the activities of others also increased the perceived importance of email to one's job, which in turn leads to the feeling of email overload [64]. These factors were not accounted for in our sample of information workers, and we expect that they would vary across the job roles included in our sample. For physicians, we expect less variation in these variables (management responsibilities, number of projects, collaboration) as the work environment, as well as the purpose of using the EHR email, was consistent across the sample. The EHR inbox is primarily used for clinical care rather than managing projects and collaborations as information workers might do.

Past research has shown several paths through which email can affect emotional states. In studies of college students, email use was linked to decreased depression as email provided social support [148, 198], but was not found to relate to daily physiological stress [178] or self-reported stress in a prospective study [290]. In studies of office employees, email has been found to be associated with daily physiological stress [177, 176], but employees also reported feeling "cut off" when their email access was restricted for the purpose of the study [177], which aligns with research highlighting the social aspect of emails. In another study, despite reporting higher mental effort with email interruptions, participants did not report more stress with email interruptions than without email [142] but surprisingly reported more positive valence and dominance. Emails during work might cause a feeling of being connected and happy to help, as the authors explained [142]. Furthermore, the nature of email work can determine whether it relates to positive or negative outcomes. Specifically, email can be rote or focused work [174]. Rote work is associated with feeling happy while focused work can involve stress [174]. Rote email work might relate to using email as a sanctioned way of procrastinating [223], or "workcrastination" as refered to in popular media, which refers to handling less important or less challenging tasks such as emails in avoidance of working on more important and challenging tasks. Overall, while some studies indicated a positive association between email and stress, other studies showed potential moderating factors and proposed ways that email could be related to positive states, which could explain the difference in email stress between our two study populations. This dissertation provided evidence that the relationship between stress and email use at the workplace can vary across work contexts, and even within the same work contexts (i.e. our findings with information workers contradicting previous findings). Further research is needed to investigate email work and stress, taking into consideration job-related and dispositional factors, as well as email content, importance and urgency as potential moderating factors.

Batching email work was associated with less stress for information workers (moderated by neuroticism) while batching was associated with more stress for physicians. Clinical hours for physicians are mostly dedicated to seeing patients, therefore, on days when physicians batch emails instead of checking in-between patient visits, they might be having busier clinics with no time to check email and therefore have a higher stress duration on those days. Additionally, our findings indicated that physicians who spent more time doing inbox work outside of work hours tended to batch their emails, which could indicate that physicians cluster their inbox work to be done after work hours when they cannot tend to it during work hours. For information workers, their schedules might be more flexible and they can choose when to check email. Batching email might be their preference rather than being dictated by their work schedule, and therefore it could be associated with less stress. More research is needed to confirm these propositions.

Window switching rate was associated with higher stress for physicians, as well as information workers who scored high on neuroticism. Physicians had almost 3 times higher window switching rate than information workers. For physicians, anecdotal evidence has been reported about the burden and inconvenience of having to collect information from several pages to accomplish tasks. The number of clicks needed to accomplish tasks and the number of screens visited have also been reported as measures of inefficiency that were associated with EHR fatigue [131, 54]. On the other hand, for information workers, more contextual data is needed to understand whether window switching is associated with boredom and lack of focus (less stress) [174] or inefficient design and information overload from visiting multiple pages to accomplish tasks (more stress).

In both study populations, we found that computer use factors explain 14-15% of the variability in daily stress. This is a considerable proportion given that stress is a complex affective and physiological state affected by many individual and situational factors that are not yet fully understood [120]. Our research contributed to the understanding of stress in the workplace by identifying computer use and individual factors related to objectively measured stress. However, for the purpose of building intelligent systems that predict stress from computer-use "behavioral sensors" (i.e. sensing stress based on behaviors exhibited through computer interactions), a higher prediction accuracy is needed. HRV data from wearable sensors can be used as ground truth to build models that predict stress from computer use patterns. Ultimately, when these models prove a high accuracy matching what a physiological sensor would detect, they can be deployed to track stress without the need for wearable sensors. Based on our findings, considering building a system that tracks computer use factors (time spent withing and outside work hours, work patterns and time spent on different activities) and makes a prediction about user stress at a given time point based on these tracked factors, the system would likely not be able to provide accurate stress predictions to match what a physiological sensor would detect. Concerns have been raised about the accuracy and biases of behavioral sensors such as cameras capturing facial expressions for automated emotion recognition [127]. Other computer-interaction-based "behavioral sensors" to measure stress from mouse and keyboard interaction have also had limited accuracy [328, 108, 168, 140]. Similarly, relying on our computers to detect our stress based on our activity patterns is still a challenging goal to reach. While a fully automated stress detection system from computer activity tracking is yet to be achieved, research in this area advances the understanding of stress, its correlates and its measurement challenges. For now, wearable devices provide accurate and unobtrusive method to measure stress in daily workplace settings, and replacing them with behavioral sensors is a challenging yet promising area of research.

The analysis of employee perspectives on stress tracking at the workplace revealed distinct benefits and challenges associated with each tracking modality: wearables and EMAs. While most users found the wearables comfortable and unobtrusive, users did not engage with their data enough to better understand their stress. Their brief interactions with their wearable device data pointed to concerns regarding understanding the data, trust in algorithmic output, and confirmation bias which made participants selectively validate their stress with wearable devices. These issues have been reported in limited self-tracking literature. We contributed empirical evidence showing the extent to which these perceived benefits of unobtrusive tracking could be problematic given the biases they might introduce. For EMAs, we showed that although they encouraged reflection of momentary affective states, participants reported difficulties assigning a "stress score" or articulating how they felt in the EMAs. Actionable design guidelines for organizational implementation of stress tracking systems include encouraging discoveries about one's stress to combat the effect of selectively engaging with the data (confirmation bias), encouraging active evaluation of system outputs to combat the effect of undue trust in algorithmic output, and providing comparative measures to help users rate their stress more accurately.

### Perceived and physiological stress

An important contribution of this work is demonstrating how perceived and physiological stress do not always align in real-life settings. Our results showed that daily physiological stress, assessed continuously through a heart-rate wearable sensor, was associated with daily computer-related work patterns. However, perceived stress measured three times a day through EMAs did not correlate with any work-related measure, except window switching. For physicians, different work patterns (i.e. working mostly within or mostly outside of workhours) showed differences in physiological stress, but no difference in EMAs of stress.

Theories on stress posit that cognitive appraisal of stressors as threatening or demanding lead to physiological stress responses such as changes in the level of the cortisol hormone, faster respiration, and elevated heart rate [153, 32, 207, 291]. This relationship between perceived stress and physiological responses has been confirmed in laboratory studies where stress is induced through a stressful task. For example, in a study that administered three different tasks as stressors [39], researchers found that all three tasks increased perceived and physiological stress. However, results from studies in the wild are inconclusive with regards to the association between perceived and physiological stress. In a review of field studies measuring self-reported stress and cortisol [111] , researcher found that some studies reported a positive association, while others reported a negative or no association between self-reported mental stress and the cortisol response. Therefore, there is no sufficient evidence from past studies for an association between self-reported stress and the cortisol response in real life settings. The authors attributed the mixed results to the diversity in study designs and the perceived stress instruments [111]. For HRV in real life settings, results on the association with perceived stress are also mixed. For example, no correlation was found between perceived job stress (i.e. difficulty and amount of work in the last month) and HRV [123]. Daily perceived negative affect and self-reported daily stressors were correlated with HRV in one study [264]. Another study collected physiological and self-reported momentary stress (high versus low) for 15 participants during five regular work days and found inconsistencies across participants in how accurately HRV can predict their EMA responses [109]. On average across participants, HRV yielded an accuracy of only 56% [109]. Another study [200] achieved a classification accuracy of 59% in a three-level prediction task of perceived stress (low, moderate, high). In a simpler classification task of high versus low stress, a study [197] found that HRV features achieved a classification accuracy of 78%.

The lack of consensus in previous studies on the relationship between HRV and perceived stress can be attributed to methodological issues. In laboratory settings where stress is induced through validated stressors such as mentally demanding tasks, stress is more salient and is reflected in both physiological responses and self-reports of the stress level experienced in the task. In everyday stress tracking in the wild, and without prior knowledge of specific stressful periods during the day, the correlation between continuous physiological data and a few EMAs during the day might not be strong.

The timing of the EMA of stress (before, during, or after a stressor event) could determine whether and how perceived stress correlates with physiological stress measured during the stressor event [254, 83, 215]. For example, a study in a controlled setting found that the physiological response to a stressor was related to subjective measures of stress during but not before or after the stress-inducing test [100].

Memory biases could also play a role in the lack of agreement between continuous physiological measures and EMAs in uncontrolled settings. When EMAs ask participants to evaluate their stress at a given moment, participants might be reflecting on a stressor that has ended. Thus, EMAs would reflect stress while physiological response would not. Conversely, one might also forget or not notice a stressor when reflecting on their stress, while physiological sensors would capture such stress. Moreover, self-reported momentary stress can be influenced by chronic stressors. Chronic stressors do not evoke a physiological response in the autonomic nervous system that produces the HRV stress response for acute stressors (i.e. the fight or flight response, see chapter 2.1), which might create a misalignment between EMAs of stress and physiological measurements.

Lastly, EMAs might be ignored during busy and stressful times, as our findings (chapter 5) and previous research has indicated. Thus, while the continuous HRV measure would capture stress during those times, the EMAs would not, which will lead to a lack of correlation for all-day stress measures, either due to missing data points at those time, or due to delayed EMA response until stressors have passed (i.e. tending to the EMA after finishing a stressful task). Given the high percentage of missed EMAs in our studies (39% for information workers and 45% for physicians), it is possible that several high-stress moments were not captured through EMAs but were still captured with the HRV sensor.

To build on our findings, future work can systematically evaluate the association between HRV-based stress from wearable sensors and perceived stress through EMAs in real-world settings to better understand when and how these different measures of stress agree and when they differ.

## 8.2 Research contributions

Through works presented in this dissertation, we made the following research contributions:

- We showed that tracking computer use metrics can reveal information about physiological stress for different working populations, and that much of the variance in daily stress duration is due to other (likely individual) factors.
- We identified factors related to computer use that correlate with daily stress duration in the workplace, and showed how these factors' association with stress can differ between different work contexts. Specifically, we showed that batching email affects different working populations differently, potentially due to the daily work schedule and whether employees choose to batch email or whether their daily work schedule allows for intermittently checking email. We also provided evidence that the relationship between stress and email use at the workplace can vary across work contexts (i.e. information work and medical work), and even within the same work contexts (i.e. our

findings with information workers contradicting previous findings). The association between window switching and stress also depends on the work context, potentially relating to the triggers of window switching.

- We modeled how individual factors affect the relationship between computer use and stress. Specifically, the relationship of batching and window switching with stress is moderated by neuroticism, and the relationship of non-workhours work duration and stress is moderated by perceived work-life imbalance.

- We provided evidence that unobtrusive sensing of stress via wearable sensors in real-life settings provides a more continuous measure that correlates with more computer work factors and captures more variation in daily stress among different work patterns compared to perceived stress.

- We evaluated employees' perspectives of automated and manual stress tracking in the workplace, and identified potential problematic behaviors such as confirmation bias and undue trust in algorithmic output of wearables. We also showed that employees desire systems that track their work and stress, but have concerns about privacy, challenges interpreting the data, and a lack of engagement with their data.

- We developed actionable guidelines for organizations and system designers in each work context to address computer use factors associated with stress, such as increased window switching and working after hours.

- We provided recommendations for effectively deploying technology-supported stress tracking at the workplace, taking into consideration both personal and organizational stress-tracking goals and concerns.

## 8.3 Implications

Previous chapters provided design and organizational implications specific to each study context. This section discusses stakeholders and potential applications more broadly. The results of research in the area of sensing and understanding stress will be of interest to several user, researcher and practitioner communities:

### 8.3.1 Users

With effective deployment of stress tracking at the workplace, users would be able to monitor their stress and understand its antecedents and consequences. With increased interest in self-tracking (e.g. the quantified self movement), the opportunity for the results of this line of research to reach and benefit users is promising. People with preexisting health conditions that can have more serious consequences of stress might be especially interested, or even instructed by their physicians, to monitor their stress [164] as personal tracking data is envisioned to contribute to the future of health management focused on personalized preventive health maintenance [282]. For example, people with existing heart conditions have to monitor their stress since high levels of stress lead to overarousal, which can cause a heart attack or sudden deaths in people with heart conditions [68]. Examples of stress tracking for people with pre-existing conditions in clinical settings (e.g. [132, 268]) assert the value of stress tracking and the need for unobtrusive continuous measurement of stress. Thus, implications of this research area will be of interest to multiple user groups.

### 8.3.2 System designers

Designers of stress-tracking applications (e.g. wearables and EMAs) would benefit from perspectives of employees reported in this work. Computer systems' designers would benefit from the identified computer interactions that increase stress. For example, we identified window switching as a contributor to workplace stress. Many HCI design studies have investigated the issue of multitasking and window switching (e.g. [311, 220, 261, 133]). Suggestions for increased efficiency include different design approaches such as ordering windows by importance, frequency of use or recency, in a grid, map, list or tile styles [305, 43, 261, 285]. Another approach suggested in the literature is semantic content extraction, which displays only the most relevant content in a window [182], and has been found to enhance task flow. These approaches can be extended to the office work applications. For the EHR inbox context, smart applications can be developed that interpret email contents and extract only the relevant information needed from other windows. Further studies are required to test these approaches in different work settings.

### 8.3.3 Personal Informatics community

The personal informatics community includes both users and researchers. Users in the personal informatics community are people who actively seek logging and tracking several aspects of their health and wellbeing, and sharing their "quantified-self" data on community forums [281]. This user base is motivated to track for behavior change, curiosity, or social engagement [76, 281]. Although the approach implemented in this dissertation, through unobtrusive tracking, targets regular employees who do not necessarily actively self-track, we believe the personal informatics community of users would be especially interested to track their stress and its related daily events. Researchers in personal informatics would also be interested in the results and methods used in this dissertation. Insights on what encourages people to track, how to keep the system unobtrusive while also ensuring data quality, how to incorporate contextual information, and how to provide value for users, will be of interest to researchers in the space of personal informatics.

### 8.3.4 Affective Computing and Context-Aware Computing

In context-aware computing [66], researchers and developers try to study and develop interactive systems that adapt to users' context, including their affect and mental state such as stress [226]. Besides interest in the methods, affective computing and context-aware computing can be an area of application for stress sensing, where the study and understanding of user stress state and related behaviors inform the design of interactive context-aware systems, and more complex forms of human-computer interaction [226].

### 8.3.5 Mental health research community

The mental health research community is concerned with monitoring mental health states and understanding markers, triggers and consequences of stress. Tracking stress and computer use at the workplace will uncover information about surrounding factors and behaviors that will advance our understanding of stress in the everyday life. The individual factors moderating the relationship between stress and computer work patterns could be investigated further in future work to advance the understanding of how dispositional factors affect mental health at the workplace.

### 8.3.6 Organizations

Stress is related to and manifested in workload and work performance, as well as professional relationships and ties at the workplace [55]. Organization would be interested in quantifying the stress of their employees and its related workplace implications to solve workplace problems and improve the productivity and quality of life for employees. We discuss the challenges and ethical considerations of such organizational programs in section 8.4. The dissertation provided actionable guidelines for deploying stress tracking at the workplace, balancing employees personal tracking goals and concerns with the organizational goals (chapter 5). The dissertation also highlighted how computer work contributes to employees' stress. Policies could be introduced to minimize after-hours work, reduce email workload and allow for breaks with non-work computer use. Employees individual differences and personal-life commitments could be taken into consideration to ensure the effectiveness of any policies or interventions.

## 8.4 Limitations

It is possible that participants' behavior may have changed through their knowledge of the computer or the stress data collection. However, given the length of our data collection period and its unobtrusive nature, as well as the participants' busy work routines, we expect that participants became habituated to the data collection very quickly and continued their regular work routines without behavior change.

Our HRV-based stress measure is the duration of medium and high stress according to the Garmin Stress Score. While studies have shown that the duration of stress is associated with negative health outcomes as the body experiences allostatic load [191] with prolonged exposure to stress, using the duration of stress as an outcome measure is unusual. Studies often use stress intensity, which reflects the level of departure from a person's normal or average HRV level. The decision to use duration of stress based on the Garmin stress score (score of 50-100) might make it difficult to compare stress levels found in our study to the findings of other studies that used other HRV measures. For example, it is unclear how our measure of stress maps to normative values of HRV [208], therefore, the health implications associated with our stress scores are also unclear. When we aggregate the duration of stress across different stress levels, we lose information about the variability of stress intensity. Future research can combine measures of intensity and duration to capture stress experiences in more detail.

The physiological response to positive arousal (e.g. feeling energized or excited) is the same as the physiological response to negative stress (i.e. feeling overwhelmed or worried about demands exceeding resources). While our measure of stress could theoretically include positive physiological arousal, we expect that typical workdays do not have a medium or high level of it. We exclude low stress, which could be the "positive stress" needed to get daily tasks done. Thus, we reduce the chances of our stress measure to reflect high energy and excitement. Medium and high stress, whether positive or negative, place substantial adaptive demands on the body and should be monitored.

On the other hand, psychological stress in daily life does not always cause physiological arousal. Controlled studies in laboratory settings inducing stress support the association between physiolgoical and perceived stress [283, 280, 39]. However, studies in daily life settings report mixed results, and have used a wide variation of instruments to measure daily, long-term, or job-related perceived stress [123, 264, 165]. Our measure of HRV-based stress might miss some events that a person might perceive as stressful, but that did not trigger a physiological response. We track physiological stress given its unobtrusiveness of measurement and significant importance for health, but we also recognize the importance of perceived stress and how it relates to wellbeing.

Participant self-selection might have unintentionally introduced bias. Only a portion of those who received the recruitment email expressed interest in participation. Stress level, attitude towards tracking, job responsibilities or other factors could have influenced participants decision to participate in the study. Our surveys showed that participants varied in their job responsibilities and their overall perceived stress, so we expect that the extent to which self-selection bias affects the sample is limited.

## 8.5 Challenges and ethical considerations

There are many challenges for real-life long-term application of stress and computer tracking at the workplace. This dissertation's approach addresses several limitations of using self-reports but also introduces new challenges. For any kind of personal sensors, the challenges associated with user adoption and abandonment of these technologies are widely recognized. Research has not yet demonstrated the feasibility of convincing "healthy" individuals to monitor their mental health [24], unlike monitoring physical health, which is becoming more common for healthy individuals. For wearables, Lazar et al. [150] examined why users adopt and abandon a wide range of wearables over a time period of two months. They enumerate several reasons why users take on and give up using wearables, but most importantly assert that unless the wearables align with people's daily routines, adoption remains short-lived. Given that adherence and abandonment are the main challenges of personal tracking, addressing these issues can enhance the opportunities for collecting and using long-term longitudinal data. Advances in sensors embedded in everyday devices enable continuous unobtrusive sensing without requiring extra effort from the user to input data, thereby addressing the major shortcoming of adherence and abandonment. As we highlighted in Chapter 5, the unobtrusiveness might come with the cost of forgetting about tracking and not engaging with the data.

A technical challenge of modeling stress for long-term applications is that models need to be continually improved and evolved. A personalized model may be less accurate a few years (or even months) later with lifestyle changes, changes in daily work responsibilities, and overall mental health changes. Because of the lack of empirical research on long-term stress monitoring, solutions to this challenge can only be hypothesized. A potential solution is an approach called "active learning" in machine learning, where the model can query a source to label new data. For example, the system can prompt the user to indicate their stress level, or wear the sensors for some time, whenever the model encounters a new pattern that has not previously been seen in the model training set. With long-term tracking, the models built will need to be continuously updated taking into account new patterns in tracked data, which might pose a technical burden.

With any user tracking system, privacy concerns are raised. Some privacy concerns associated with collecting data from wearables and computer logs were reported by our participants. Raij et al. [232] conducted a user study to explore privacy concerns associated with the disclosure of data collected by wearable sensors in the everyday life and found that disclosure of stress is among the highest concerns participants had about using wearables. Similarly, Esptein et al (2016) suggested that perceived privacy risks negatively influenced individual intention to track personal health with commercial fitness trackers [75]. The findings of Raij et al. (2011) further show that restricting or abstracting collected data had a significant effect on reducing privacy concerns about data exposure. Sharing personal tracking data with third parties such as physicians, insurance companies, and employers can have benefits for providing personalized services. However, privacy research shows that privacy calculus (ie the evaluation of risks and benefits of sharing personal data) affects users' willingness to share self-tracking data with third parties [303]. Specifically, perceived privacy risks always deter users from disclosing their data, while sharing data due to perceived benefits depend on data sensitivity [303]. Capturing sensitive data about users and their daily work activities must consider privacy challenges to avoid ethical and legal consequences, and to reach tracking goals and benefits for both users and organizations. A proposed solution is to store and process the data locally on the user's device, rather than sending it to a shared server. With data never leaving the user's device, the issue becomes more of a security issue, keeping the data from unauthorized access and use, than a privacy issue of exposing personal data.

Related to privacy, another ethical challenge for stress and computer tracking is user agency. Users should have agency over their data, making informed decisions on what to track and share [302, 56, 58, 61]. Voida et al. [302] suggest several steps to help users make informed privacy decision for systems collecting sensor data streams: telling participants what will be collected, how long it will be stored, where it is going to be stored and sharing a sample of the data with participants. Further, to ensure user agency, Voida et al. suggest allowing users to revoke participation and delete their data. The problem with this approach of user agency is the resulting incomplete and missing data. Missing data can be compensated for, either by building models without the features containing missing data, or by imputing missing values based on historical data, which might be a feasible solution if longitudinal data was acquired. The balance between user agency and data quality remains to be fully addressed in user tracking and sensing studies.

Another challenge in user tracking with sensors is making high-level inferences about observed phenomena. Voida et al. [302] explain that the primary weakness of using sensors for HCI research is their poor ability to answer questions of *why*. If used inappropriately, sensors can even poorly detect the *what*, especially for higher level extrapolations. For instance, as explained earlier, we observed window switching from computer log data, but it is unclear whether it is triggered by employees' interruptability, internal or external distractions, or system affordances. Higher level interpretations (e.g. triggers, intentions, goals) from log and sensor data need additional investigations.

## 8.6 Conclusion

This dissertation has presented novel findings on the association between computer use factors and stress in two working populations: information workers and physicians. Unobtrusive tracking methods were employed in real-workplace environments. The dissertation presented a holistic view of the process of deploying stress tracking in the workplace: from quantifying relevant computer use metrics, identifying patterns and clusters, to assessing the prediction power, identifying the association between computer use factors and stress (and the individual factors affecting the association), to reporting employee perspectives and providing actionable guidelines for organizations and designers. The findings advance the understanding of computer use factors related to stress, and advance our understanding of the benefits and limitations of unobtrusive stress tracking in real-life workplace environments. I believe workplace analytics is the future of monitoring employee wellbeing. Unobtrusively and continuously tracking workplace factors related to stress will help in directing efforts towards improving the productivity and quality of life for employees.

# Bibliography

[1] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri. Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051, 2006.

[2] P. Adams, M. Rabbi, T. Rahman, M. Matthews, A. Voida, G. Gay, T. Choudhury, and S. Voida. Towards personal stress informatics: Comparing minimally invasive techniques for measuring daily stress in the wild. pages 72–79. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[3] J. Adler-Milstein, A. J. Holmgren, P. Kralovec, C. Worzala, T. Searcy, and V. Patel. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. *Journal of the American Medical Informatics Association*, 24(6):1142–1148, Nov. 2017.

[4] J. Adler-Milstein, W. Zhao, R. Willard-Grace, M. Knox, and K. Grumbach. Electronic health records and burnout: Time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *Journal of the American Medical Informatics Association*, 27(4):531–538, Apr. 2020.

[5] A. Ahtinen, M. Isomursu, M. Mukhtar, J. Mäntyjärvi, J. Häkkilä, and J. Blom. Designing social features for mobile and ubiquitous wellness applications. In *Proceedings of the 8th International Conference on Mobile and Ubiquitous Multimedia - MUM '09*, pages 1–10, Cambridge, United Kingdom, 2009. ACM Press.

[6] J. Aigrain, S. Dubuisson, M. Detyniecki, and M. Chetouani. Person-specific behavioural features for automatic stress detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, volume 3, pages 1–6. IEEE, 2015.

[7] Akbar, Mark, Pavlidis, and Gutierrez-Osuna. An Empirical Study Comparing Unobtrusive Physiological Sensors for Stress Detection in Computer Work. *Sensors*, 19(17):3766, Aug. 2019.

[8] F. Akbar, A. E. Bayraktaroglu, P. Buddharaju, D. R. Da Cunha Silva, G. Gao, T. Grover, R. Gutierrez-Osuna, N. C. Jones, G. Mark, I. Pavlidis, and others. Email Makes You Sweat: Examining Email Interruptions and Stress Using Thermal Imaging.

In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 668. ACM, 2019.

[9] F. Akbar, G. Mark, E. M. Warton, M. E. Reed, S. Prausnitz, J. A. East, M. F. Moeller, and T. A. Lieu. Physicians' electronic inbox work patterns and factors associated with high inbox work duration. *Journal of the American Medical Informatics Association*, page ocaa229, Oct. 2020.

[10] F. Alamudun, J. Choi, R. Gutierrez-Osuna, H. Khan, and B. Ahmed. Removal of subject-dependent and activity-dependent variation in physiological measures of stress. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 115–122. IEEE, 2012.

[11] J. Anderson, J. Leubner, and S. R. Brown. EHR Overtime: An Analysis of Time Spent After Hours by Family Physicians. *Family Medicine*, 52(2):135–137, Feb. 2020.

[12] E. Andreou, E. C. Alexopoulos, C. Lionis, L. Varvogli, C. Gnardellis, G. P. Chrousos, and C. Darviri. Perceived stress scale: reliability and validity study in Greece. *International journal of environmental research and public health*, 8(8):3287–3298, 2011.

[13] K. Antila, M. van Gils, J. Merilahti, and I. Korhonen. Associations of psychological self-assessments and HRV in long-term measurements at home. In *Proceedings of the 3rd European Medical and Biological Engineering Conference, IFMBE Proceedings, 2005*, 2005.

[14] G. N. Armaiz-Pena, S. K. Lutgendorf, S. W. Cole, and A. K. Sood. Neuroendocrine modulation of cancer progression. *Brain, behavior, and immunity*, 23(1):10–15, 2009.

[15] D. Armstrong, A. Gosling, J. Weinman, and T. Marteau. The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study. *Sociology*, 31(3):597–606, Aug. 1997.

[16] B. G. Arndt, J. W. Beasley, M. D. Watkinson, J. L. Temte, W.-J. Tuan, C. A. Sinsky, and V. J. Gilchrist. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *The Annals of Family Medicine*, 15(5):419–426, 2017.

[17] S. Attipoe, Y. Huang, S. Schweikhart, S. Rust, J. Hoffman, and S. Lin. Factors Associated With Electronic Health Record Usage Among Primary Care Physicians After Hours: Retrospective Cohort Study. *JMIR Human Factors*, 6(3):e13779, Sept. 2019.

[18] M. S. H. Aung, F. Alquaddoomi, C.-K. Hsieh, M. Rabbi, L. Yang, J. P. Pollak, D. Estrin, and T. Choudhury. Leveraging Multi-Modal Sensing for Mobile Health: A Case Review in Chronic Pain. *IEEE Journal of Selected Topics in Signal Processing*, 10(5):962–974, Aug. 2016.

[19] D. Avni-Babad. Routine and feelings of safety, confidence, and well-being: Routine, confidence, and well-being. *British Journal of Psychology*, 102(2):223–244, May 2011.

[20] R. Ayyagari, V. Grover, and R. Purvis. Technostress: Technological antecedents and implications. *MIS quarterly*, pages 831–858, 2011. Publisher: JSTOR.

[21] V. V. Baba and M. Jamal. Routinization of job context and job content as related to employees' quality of working life: A study of Canadian nurses. *Journal of Organizational Behavior*, 12(5):379–386, Sept. 1991.

[22] S. R. Barley, D. E. Meyerson, and S. Grodal. E-mail as a source and symbol of stress. *Organization Science*, 22(4):887–906, 2011.

[23] A. Barreto, J. Zhai, and M. Adjouadi. Non-intrusive physiological monitoring for automated stress detection in human-computer interaction. In *International Workshop on Human-Computer Interaction*, pages 29–38. Springer, 2007.

[24] A. Baumel, J. Baker, M. L. Birnbaum, H. Christensen, M. De Choudhury, D. C. Mohr, F. Muench, D. Schlosser, N. Titov, and J. M. Kane. Summary of Key Issues Raised in the Technology for Early Awareness of Addiction and Mental Illness (TEAAM-I) Meeting. *Psychiatric Services*, pages appi–ps, 2018.

[25] D. J. Beal. ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annu. Rev. Organ. Psychol. Organ. Behav.*, 2(1):383–407, 2015.

[26] D. Beckers. *Overtime work and well-being: opening up the black box*. PhD thesis, s.n.], S.l., 2008. ISBN: 9789090235639 OCLC: 778264119.

[27] D. G. Beckers, M. L. van Hooff, D. van der Linden, M. A. Kompier, T. W. Taris, and S. A. Geurts. A diary study to open up the black box of overtime work among university faculty members. *Scandinavian Journal of Work, Environment & Health*, 34(3):213–223, June 2008.

[28] T. A. Beehr. *Psychological stress in the workplace (psychology revivals)*. Routledge, 2014.

[29] L. Y. Belkin, W. J. Becker, and S. A. Conroy. Exhausted, but unable to disconnect: After-hours email, work-family balance and identification. In *Academy of Management Proceedings*, volume 2016, page 10353. Academy of Management Briarcliff Manor, NY 10510, 2016.

[30] V. Bellotti, N. Ducheneaut, M. Howard, I. Smith, and R. E. Grinter. Quality versus quantity: E-mail-centric task management and its relation with overload. *Human–Computer Interaction*, 20(1-2):89–138, 2005. Publisher: Taylor & Francis.

[31] J. Blascovich, W. B. Mendes, S. B. Hunter, and K. Salomon. Social" facilitation" as challenge and threat. *Journal of personality and social psychology*, 77(1):68, 1999.

[32] J. Blascovich and J. Tomaka. The Biopsychosocial Model of Arousal Regulation. In *Advances in Experimental Social Psychology*, volume 28, pages 1–51. Elsevier, 1996.

[33] K. Boehner, R. DePaula, P. Dourish, and P. Sengers. How emotion is made and measured. *International Journal of Human-Computer Studies*, 65(4):275–291, Apr. 2007.

[34] A. Bogomolov, B. Lepri, M. Ferron, F. Pianesi, and A. S. Pentland. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. In *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, pages 477–486, New York, NY, USA, 2014. ACM. event-place: Orlando, Florida, USA.

[35] W. R. Boswell and J. B. Olson-Buchanan. The use of communication technologies after hours: The role of work attitudes and work-life conflict. *Journal of Management*, 33(4):592–610, 2007.

[36] A. Bradley, D. P. Brumby, A. L. Cox, and J. Bird. How to manage your inbox: is a once a day strategy best? In *Proceedings of the 27th International BCS Human Computer Interaction Conference*, page 20. British Computer Society, 2013.

[37] P. J. Brantley, C. D. Waggoner, G. N. Jones, and N. B. Rappaport. A daily stress inventory: Development, reliability, and validity. *Journal of behavioral medicine*, 10(1):61–73, 1987. Publisher: Springer.

[38] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, Jan. 2006.

[39] A. Brugnera, C. Zarbo, M. P. Tarvainen, P. Marchettini, R. Adorni, and A. Compare. Heart rate variability during acute psychosocial stress: A randomized cross-over trial of verbal and non-verbal laboratory stressors. *International Journal of Psychophysiology*, 127:17–25, May 2018.

[40] J. T. Cacioppo, G. G. Berntson, J. T. Larsen, K. M. Poehlmann, and T. A. Ito. The psychophysiology of emotion. *Handbook of emotions*, 2:173–191, 2000.

[41] A. J. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, and R. Kleiger. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *European heart journal*, 17(3):354–381, 1996.

[42] J. Car and A. Sheikh. Email consultations in health care: 1—scope and effectiveness. *Bmj*, 329(7463):435–438, 2004.

[43] O. Chapuis and N. Roussel. Copy-and-paste between overlapping windows. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 201–210. ACM, 2007.

[44] S. Chatzitheochari and S. Arber. Lack of sleep, work and the long hours culture: evidence from the UK Time Use Survey. *Work, Employment and Society*, 23(1):30–48, Mar. 2009.

[45] Y. Chen, V. Ngo, S. Harrison, and V. Duong. Unpacking exam-room computing: negotiating computer-use in patient-physician interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3343–3352. ACM, 2011.

[46] E. K. Choe, S. Abdullah, M. Rabbi, E. Thomaz, D. A. Epstein, F. Cordeiro, M. Kay, G. D. Abowd, T. Choudhury, J. Fogarty, B. Lee, M. Matthews, and J. A. Kientz. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing*, 16(1):74–84, Jan. 2017.

[47] J. Choi and R. Gutierrez-Osuna. Estimating mental stress using a wearable cardio-respiratory sensor. pages 150–154. IEEE, 2010.

[48] C.-F. Chung, N. Gorm, I. A. Shklovski, and S. Munson. Finding the Right Fit: Understanding Health Tracking in Workplace Wellness Programs. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4875–4886, Denver Colorado USA, May 2017. ACM.

[49] K. Clark and S. Kalin. Technostressed Out? How to Cope in the Digital Age. *Library Journal*, 121(13):30–32, 1996. Publisher: ERIC.

[50] R. A. Cohen. Yerkes–Dodson Law. *Encyclopedia of clinical neuropsychology*, pages 2737–2738, 2011. Publisher: Springer.

[51] S. Cohen. Aftereffects of stress on human performance and social behavior: a review of research and theory. *Psychological bulletin*, 88(1):82, 1980.

[52] S. Cohen, T. Kamarck, R. Mermelstein, and others. Perceived stress scale. *Measuring stress: A guide for health and social scientists*, pages 235–283, 1994.

[53] S. Cohen, R. C. Kessler, and L. U. Gordon. *Measuring stress: A guide for health and social scientists*. Oxford University Press on Demand, 1997.

[54] T. K. Colicchio, J. J. Cimino, and G. Del Fiol. Unintended Consequences of Nationwide Electronic Health Record Adoption: Challenges and Opportunities in the Post-Meaningful Use Era. *Journal of Medical Internet Research*, 21(6):e13313, June 2019.

[55] T. W. Colligan and E. M. Higgins. Workplace stress: Etiology and consequences. *Journal of workplace behavioral health*, 21(2):89–97, 2006. Publisher: Taylor & Francis.

[56] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay. Design requirements for technologies that encourage physical activity. pages 457–466. ACM, 2006.

[57] S. Consolvo, J. Jung, B. Greenstein, P. Powledge, G. Maganis, and D. Avrahami. The Wi-Fi privacy ticker: improving awareness & control of personal information exposure on Wi-Fi. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 321–330, Copenhagen Denmark, Sept. 2010. ACM.

[58] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, and R. Libby. Activity sensing in the wild: a field trial of ubifit garden. pages 1797–1806. ACM, 2008.

[59] F. Cordeiro, D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd, and J. Fogarty. Barriers and Negative Nudges: Exploring Challenges in Food Journaling. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 1159–1162, Seoul, Republic of Korea, 2015. ACM Press.

[60] J. Costa, A. T. Adams, M. F. Jung, F. Guimbretière, and T. Choudhury. EmotionCheck: leveraging bodily signals and false feedback to regulate our emotions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 758–769, Heidelberg Germany, Sept. 2016. ACM.

[61] K. M. Cresswell, D. W. Bates, and A. Sheikh. Ten key considerations for the successful implementation and adoption of large-scale health information technology. *Journal of the American Medical Informatics Association*, 20(e1):e9–e13, 2013.

[62] M. P. Cross and S. D. Pressman. Understanding the Connections between Positive Affect and Health. *The Handbook of Stress and Health: A Guide to Research and Practice*, pages 75–95, 2017.

[63] B. H. Crotty, Y. Tamrat, A. Mostaghimi, C. Safran, and B. E. Landon. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. *Health Affairs*, 33(10):1817–1822, 2014.

[64] L. A. Dabbish and R. E. Kraut. Email overload at work: an analysis of factors associated with email strain. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 431–440. ACM, 2006.

[65] D. R. Dacunhasilva, Z. Wang, and R. Gutierrez-Osuna. Towards Participant-Independent Stress Detection Using Instrumented Peripherals. *IEEE Transactions on Affective Computing*, pages 1–1, 2021.

[66] A. K. Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001. Publisher: Springer-Verlag.

[67] S. S. Dickerson and M. E. Kemeny. Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. *Psychological Bulletin*, 130(3):355–391, 2004.

[68] J. E. Dimsdale. Psychological stress and cardiovascular disease. *Journal of the American College of Cardiology*, 51(13):1237–1246, 2008.

[69] R. K. Dishman, Y. Nakamura, M. E. Garcia, R. W. Thompson, A. L. Dunn, and S. N. Blair. Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. *International Journal of Psychophysiology*, 37(2):121–133, 2000.

[70] S. Dockray, N. Grant, A. A. Stone, D. Kahneman, J. Wardle, and A. Steptoe. A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Social Indicators Research*, 99(2):269–283, 2010.

[71] J. R. Edwards and C. L. Cooper. The person-environment fit approach to stress: recurring problems and some suggested solutions. *Journal of organizational behavior*, 11(4):293–307, 1990. Publisher: Wiley Online Library.

[72] D. Enthoven. Quantified Self? How About a Quantified Workplace? *WIRED*, 2013.

[73] C. Epp, M. Lippold, and R. L. Mandryk. Identifying Emotional States Using Keystroke Dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 715–724, New York, NY, USA, 2011. ACM. event-place: Vancouver, BC, Canada.

[74] D. A. Epstein, D. Avrahami, and J. T. Biehl. Taking 5: Work-Breaks, Productivity, and Opportunities for Personal Informatics for Knowledge Workers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 673–684, San Jose California USA, May 2016. ACM.

[75] D. A. Epstein, M. Caraway, C. Johnston, A. Ping, J. Fogarty, and S. A. Munson. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1109–1113, San Jose California USA, May 2016. ACM.

[76] D. A. Epstein, A. Ping, J. Fogarty, and S. A. Munson. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pages 731–742, Osaka, Japan, 2015. ACM Press.

[77] H. Eyrolle and J.-M. Cellier. The effects of interruptions in work activity: Field and laboratory results. *Applied ergonomics*, 31(5):537–543, 2000.

[78] G. H. Fenner and R. W. Renn. Technology-assisted supplemental work: Construct definition and a research framework. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, 43(2-3):179–200, 2004.

[79] S. Folkman, R. S. Lazarus, C. Dunkel-Schetter, A. DeLongis, and R. J. Gruen. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *Journal of personality and social psychology*, 50(5):992, 1986.

[80] M. Frankenhaeuser. Coping with Stress at Work. *International Journal of Health Services*, 11(4):491–510, Oct. 1981.

[81] B. L. Fredrickson and R. W. Levenson. Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition & emotion*, 12(2):191, 1998.

[82] D. Fuller, E. Colwell, J. Low, K. Orychock, M. A. Tobin, B. Simango, R. Buote, D. Van Heerden, H. Luan, K. Cullen, L. Slade, and N. G. A. Taylor. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. *JMIR mHealth and uHealth*, 8(9):e18694, Sept. 2020.

[83] J. Gaab, N. Rohleder, U. Nater, and U. Ehlert. Psychological determinants of the cortisol stress response: the role of anticipatory cognitive appraisal. *Psychoneuroendocrinology*, 30(6):599–610, July 2005.

[84] M. W. Gallagher, A. M. Schoemann, and S. D. Pressman. Mastery beliefs and intraindividual variability of anxiety. *Cognitive therapy and research*, 35(3):227–231, 2011.

[85] R. L. Gardner, E. Cooper, J. Haskell, D. A. Harris, S. Poplau, P. J. Kroth, and M. Linzer. Physician stress and burnout: the impact of health information technology. *Journal of the American Medical Informatics Association*, 26(2):106–114, Feb. 2019.

[86] A. Ghandeharioun and R. Picard. BrightBeat: Effortlessly Influencing Breathing for Cultivating Calmness and Focus. pages 1624–1631. ACM, 2017.

[87] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski. Monitoring stress with a wrist device using context. *Journal of biomedical informatics*, 73:159–170, 2017. Publisher: Elsevier.

[88] V. M. González and G. Mark. "Constant, constant, multi-tasking craziness": managing multiple working spheres. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04*, pages 113–120, Vienna, Austria, 2004. ACM Press.

[89] N. Gordon. Highlights of results of the Kaiser Permanente Northern California 2014/2015 Member Health Survey. Technical report, Oakland, CA: Division of Research, Kaiser Permanente Medical Care Program, Oct. 2017.

[90] N. Gorm and I. Shklovski. Sharing Steps in the Workplace: Changing Privacy Concerns Over Time. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4315–4319, San Jose California USA, May 2016. ACM.

[91] R. Gouveia, F. Pereira, E. Karapanos, S. A. Munson, and M. Hassenzahl. Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 144–155, Heidelberg Germany, Sept. 2016. ACM.

[92] S. Greene, H. Thapliyal, and A. Caban-Holt. A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, 5(4):44–56, 2016. Publisher: IEEE.

[93] M. Gregory, E. Russo, and H. Singh. Electronic Health Record Alert-Related Workload as a Predictor of Burnout in Primary Care Providers. *Applied Clinical Informatics*, 08(03):686–697, 2017.

[94] J. J. Gross and O. P. John. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of personality and social psychology*, 85(2):348, 2003.

[95] D. E. Guest. Perspectives on the Study of Work-life Balance. *Social Science Information*, 41(2):255–279, June 2002.

[96] X. Gui, Y. Chen, C. Caldeira, D. Xiao, and Y. Chen. When Fitness Meets Social Networks: Investigating Fitness Tracking and Social Practices on WeRun. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1647–1659, Denver Colorado USA, May 2017. ACM.

[97] P. A. Hancock. A dynamic model of stress and sustained attention. *Human factors*, 31(5):519–537, 1989. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[98] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[99] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6(2):156–166, 2005.

[100] J. Hellhammer and M. Schubert. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology*, 37(1):119–124, Jan. 2012.

[101] T. Heponiemi, S. Kujala, S. Vainiomäki, T. Vehko, T. Lääveri, J. Vänskä, E. Ketola, S. Puttonen, and H. Hyppönen. Usability Factors Associated With Physicians' Distress and Information System–Related Stress: Cross-Sectional Survey. *JMIR Medical Informatics*, 7(4):e13466, Nov. 2019.

[102] J. Herbert. Fortnighly review. Stress, the brain, and mental illness. *BMJ: British Medical Journal*, 315(7107):530, 1997.

[103] J. Hernandez, D. McDuff, R. Fletcher, and R. W. Picard. Inside-out: Reflecting on your inner state. In *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 324–327, San Diego, CA, Mar. 2013. IEEE.

[104] J. Hernandez, D. McDuff, C. Infante, P. Maes, K. Quigley, and R. Picard. Wearable ESM: differences in the experience sampling method across wearable devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 195–205. ACM, 2016.

[105] J. Hernandez, D. McDuff, and R. W. Picard. Biowatch: estimation of heart and breathing rates from wrist motions. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 169–176. IEEE, 2015.

[106] J. Hernandez, D. J. McDuff, and R. W. Picard. Biophone: Physiology monitoring from peripheral smartphone motions. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7180–7183. IEEE, 2015.

[107] J. Hernandez, R. R. Morris, and R. W. Picard. Call center stress recognition with person-specific models. pages 125–134. Springer, 2011.

[108] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–60. ACM, 2014.

[109] J. Hernandez Rivera. *Towards wearable stress measurement*. PhD Thesis, Massachusetts Institute of Technology, 2015.

[110] R. W. Hilliard, J. Haskell, and R. L. Gardner. Are specific elements of electronic health record use associated with clinician burnout more than others? *Journal of the American Medical Informatics Association*, page ocaa092, July 2020.

[111] N. Hjortskov, A. H. Garde, P. Ørbæk, and M. Hansen. Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. *Stress and Health*, 20(2):91–98, Apr. 2004.

[112] V. Hollis, A. Pekurovsky, E. Wu, and S. Whittaker. On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–31, Sept. 2018.

[113] T. H. Holmes and R. H. Rahe. The social readjustment rating scale. *Journal of psychosomatic research*, 11(2):213–218, 1967.

[114] I. Houkes, Y. Winants, M. Twellaar, and P. Verdonk. Development of burnout over time and the causal order of the three dimensions of burnout among male and female GPs. A three-wave panel study. *BMC Public health*, 11(1):240, 2011.

[115] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar. cStress: towards a gold standard for continuous stress assessment in the mobile environment. pages 493–504. ACM, 2015.

[116] J. M. Hudson, J. Christensen, W. A. Kellogg, and T. Erickson. I'd be overwhelmed, but it's just one more thing to do: Availability and interruption in research management. In *Proceedings of the SIGCHI Conference on Human factors in computing systems*, pages 97–104. ACM, 2002.

[117] K. Hänsel, A. Alomainy, and H. Haddadi. Large scale mood and stress self-assessments on a smartwatch. pages 1180–1184. ACM, 2016.

[118] S. T. Innstrand, E. M. Langballe, E. Falkum, and O. G. Aasland. Exploring within-and between-gender differences in burnout: 8 different occupational groups. *International Archives of Occupational and Environmental Health*, 84(7):813–824, 2011.

[119] S. Intille, C. Haynes, D. Maniar, A. Ponnada, and J. Manjourides. EMA: Microinteraction-based ecological momentary assessment (EMA) using a smartwatch. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1124–1128, Heidelberg Germany, Sept. 2016. ACM.

[120] C. E. Izard. The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review*, 2(4):363–370, Oct. 2010.

[121] T. Jackson, R. Dawson, and D. Wilson. The cost of email interruption. *Journal of systems and information technology*, 5(1):81–92, 2001.

[122] G. Jessup, S. Bian, Y.-W. Chen, and A. Bundy. PIEL Survey application manual. Technical report, 2012.

[123] T. Kageyama, N. Nishikido, T. Kobayashi, Y. Kurokawa, T. Kaneko, and M. Kabuto. Self-Reported Sleep Quality, Job Stress, and Daytime Autonomic Activities Assessed in Terms of Short-Term Heart Rate Variability among Male White-Collar Workers. *INDUSTRIAL HEALTH*, 36(3):263–272, 1998.

[124] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702):1776–1780, 2004.

[125] A. D. Kanner, J. C. Coyne, C. Schaefer, and R. S. Lazarus. Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of behavioral medicine*, 4(1):1–39, 1981.

[126] R. Karasek, C. Brisson, N. Kawakami, I. Houtman, P. Bongers, and B. Amick. The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. *Journal of Occupational Health Psychology*, 3(4):322–355, 1998.

[127] Kate Crawford. Artificial Intelligence Is Misreading Human Emotion. *The Atlantic*, Apr. 2021.

[128] O. Kelly, K. Matheson, A. Martinez, Z. Merali, and H. Anisman. Psychosocial stress evoked by a virtual audience: relation to neuroendocrine activity. *CyberPsychology & Behavior*, 10(5):655–662, 2007. Publisher: Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA.

[129] S. Khairat, G. Burke, H. Archambault, T. Schwartz, J. Larson, and R. Ratwani. Focus Section on Health IT Usability: Perceived Burden of EHRs on Physicians at Different Stages of Their Career. *Applied Clinical Informatics*, 09(02):336–347, Apr. 2018.

[130] S. Khairat, C. Coleman, P. Ottmar, T. Bice, R. Koppel, and S. S. Carson. Physicians' gender and their use of electronic health records: findings from a mixed-methods usability study. *Journal of the American Medical Informatics Association*, 26(12):1505–1514, Dec. 2019.

[131] S. Khairat, C. Coleman, P. Ottmar, D. I. Jayachander, T. Bice, and S. S. Carson. Association of Electronic Health Record Use With Physician Fatigue and Efficiency. *JAMA Network Open*, 3(6):e207385, June 2020.

[132] B. Kikhia, T. G. Stavropoulos, S. Andreadis, N. Karvonen, I. Kompatsiaris, S. Sävenstedt, M. Pijl, and C. Melander. Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors*, 16(12):1989, 2016.

[133] G. Kim and H. C. Kim. Designing of multimodal feedback for enhanced multitasking performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3113–3122. ACM, 2011.

[134] Y.-H. Kim, J. H. Jeon, E. K. Choe, B. Lee, K. Kim, and J. Seo. TimeAware: Leveraging Framing Effects to Enhance Personal Productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 272–283, San Jose California USA, May 2016. ACM.

[135] Y.-H. Kim, J. H. Jeon, B. Lee, E. K. Choe, and J. Seo. OmniTrack: A Flexible Self-Tracking Approach Leveraging Semi-Automated Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–28, Sept. 2017.

[136] G. Kinman and F. Jones. Lay representations of workplace stress: What do people really mean when they say they are stressed? *Work & Stress*, 19(2):101–120, Apr. 2005.

[137] C. Kirschbaum and D. H. Hellhammer. Salivary cortisol in psychobiological research: an overview. *Neuropsychobiology*, 22(3):150–169, 1989. Publisher: Karger Publishers.

[138] M. Kivimäki, M. Virtanen, M. Elovainio, A. Kouvonen, A. Väänänen, and J. Vahtera. *Work stress in the etiology of coronary heart disease–a meta-analysis.* 2006.

[139] R. Kocielnik, N. Sidorova, F. M. Maggi, M. Ouwerkerk, and J. H. Westerink. Smart technologies for long-term stress monitoring at work. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 53–58. IEEE, 2013.

[140] A. Kolakowska. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In *2013 6th International Conference on Human System Interactions (HSI)*, pages 548–555, Sopot, Poland, June 2013. IEEE.

[141] S. Koldijk, M. A. Neerincx, and W. Kraaij. Detecting Work Stress in Offices by Combining Unobtrusive Sensors. *IEEE Transactions on Affective Computing*, 9(2):227–239, Apr. 2018.

[142] S. Koldijk, M. Sappelli, M. Neerincx, and W. Kraaij. Unobtrusive monitoring of knowledge workers for stress self-regulation. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 335–337. Springer, 2013.

[143] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij. The SWELL knowledge work dataset for stress and user modeling research. In *Proceedings of the 16th international conference on multimodal interaction*, pages 291–298. ACM, 2014.

[144] T. L. Kraft and S. D. Pressman. Grin and bear it: The influence of manipulated facial expression on the stress response. *Psychological science*, 23(11):1372–1378, 2012.

[145] K. Kushlev and E. W. Dunn. Checking email less frequently reduces stress. *Computers in Human Behavior*, 43:220–228, 2015. Publisher: Elsevier.

[146] H. M. Lagraauw, J. Kuiper, and I. Bot. Acute and chronic psychological stress as risk factors for cardiovascular disease: Insights gained from epidemiological, clinical and experimental studies. *Brain, Behavior, and Immunity*, 50:18–30, Nov. 2015.

[147] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell. A survey of mobile phone sensing. *IEEE Communications Magazine*, 48(9):140–150, Sept. 2010.

[148] R. LaRose, M. S. Eastin, and J. Gregg. Reformulating the Internet paradox: Social cognitive explanations of Internet use and depression. *Journal of online behavior*, 1(2):1092–4790, 2001.

[149] R. Larson and M. Csikszentmihalyi. The experience sampling method. In *Flow and the foundations of positive psychology*, pages 21–34. Springer, 2014.

[150] A. Lazar, C. Koehler, J. Tanenbaum, and D. H. Nguyen. Why we use and abandon smart devices. pages 635–646. ACM, 2015.

[151] R. S. Lazarus. Puzzles in the study of daily hassles. *Journal of behavioral medicine*, 7(4):375–389, 1984.

[152] R. S. Lazarus. From psychological stress to the emotions: A history of changing outlooks. *Annual review of psychology*, 44(1):1–22, 1993. Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

[153] R. S. Lazarus and S. Folkman. *Stress, appraisal, and coping.* Springer publishing company, 1984.

[154] S. Leroy. Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes*, 109(2):168–181, 2009.

[155] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union, 1966. Issue: 8.

[156] J. A. Levine, I. T. Pavlidis, L. MacBride, Z. Zhu, and P. Tsiamyrtzis. Description and clinical studies of a device for the instantaneous detection of office-place stress. *Work*, 34(3):359–364, 2009. Publisher: IOS Press.

[157] I. Li, A. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 557, Atlanta, Georgia, USA, 2010. ACM Press.

[158] I. Li, A. K. Dey, and J. Forlizzi. Using context to reveal factors that affect physical activity. *ACM Transactions on Computer-Human Interaction*, 19(1):1–21, Mar. 2012.

[159] T. A. Lieu and G. L. Freed. Unbounded–Parent-Physician Communication in the Era of Portal Messaging. *JAMA pediatrics*, 173(9):811–812, 2019.

[160] J. G. Logan and D. J. Barksdale. Allostasis and allostatic load: expanding the discourse on stress and cardiovascular disease. *Journal of Clinical Nursing*, 17(7b):201–208, July 2008.

[161] F. T. Ltd. Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability. White Paper.

[162] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 351–360. ACM, 2012.

[163] U. Lundberg and K. Palm. Workload and catecholamine excretion in parents of preschool children. *Work & Stress*, 3(3):255–260, July 1989.

[164] D. Lupton. Self-tracking, health and medicine. *Health Sociology Review*, 26(1):1–5, Jan. 2017.

[165] Y. Lutchyn, P. Johns, M. Czerwinski, S. Iqbal, G. Mark, and A. Sano. Stress is in the eye of the beholder. pages 119–124. IEEE, 2015.

[166] Y. Lyu, X. Luo, J. Zhou, C. Yu, C. Miao, T. Wang, Y. Shi, and K.-i. Kameyama. Measuring photoplethysmogram-based stress-induced vascular response index to assess cognitive load and stress. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 857–866. ACM, 2015.

[167] Léger, S. Drouin, D. L. Collins, T. Popa, and M. Kersten-Oertel. Quantifying attention shifts in augmented reality image-guided neurosurgery. *Healthcare Technology Letters*, 4(5):188–192, Oct. 2017.

[168] W. Maehr. *eMotion: Estimation of user's emotional state by mouse motions*. VDM Verlag, 2008.

[169] L. Mamykina, E. D. Mynatt, and D. R. Kaufman. Investigating health management practices of individuals with diabetes. In *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, page 927, Montréal, Québec, Canada, 2006. ACM Press.

[170] G. Mark. Multitasking in the digital age. *Synthesis Lectures On Human-Centered Informatics*, 8(3):1–113, 2015. Publisher: Morgan & Claypool Publishers.

[171] G. Mark, V. M. Gonzalez, and J. Harris. No task left behind?: examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*, page 321, Portland, Oregon, USA, 2005. ACM Press.

[172] G. Mark, D. Gudith, and U. Klocke. The cost of interrupted work: more speed and stress. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 107, Florence, Italy, 2008. ACM Press.

[173] G. Mark, S. Iqbal, M. Czerwinski, and P. Johns. Focused, Aroused, but so Distractible: Temporal Perspectives on Multitasking and Communications. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 903–916, Vancouver BC Canada, Feb. 2015. ACM.

[174] G. Mark, S. T. Iqbal, M. Czerwinski, and P. Johns. Bored mondays and focused afternoons: the rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3025–3034, Toronto Ontario Canada, Apr. 2014. ACM.

[175] G. Mark, S. T. Iqbal, M. Czerwinski, P. Johns, and A. Sano. Neurotics Can't Focus: An *in situ* Study of Online Multitasking in the Workplace. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1739–1744, San Jose California USA, May 2016. ACM.

[176] G. Mark, S. T. Iqbal, M. Czerwinski, P. Johns, A. Sano, and Y. Lutchyn. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1717–1728. ACM, 2016.

[177] G. Mark, S. Voida, and A. Cardello. A pace not dictated by electrons: an empirical study of work without email. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 555–564. ACM, 2012.

[178] G. Mark, Y. Wang, and M. Niiya. Stress and multitasking in everyday college life: an empirical study of online activity. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 41–50, Toronto, Ontario, Canada, 2014. ACM Press.

[179] A. Mathur, M. Van den Broeck, G. Vanderhulst, A. Mashhadi, and F. Kawsar. Tiny habits in the giant enterprise: understanding the dynamics of a quantified workplace. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pages 577–588, Osaka, Japan, 2015. ACM Press.

[180] G. Matthews, T. J. Sparkes, and H. M. Bygrave. Attentional overload, stress, and simulate driving performance. *Human Performance*, 9(1):77–101, 1996.

[181] M. Matthews, J. Snyder, L. Reynolds, J. T. Chien, A. Shih, J. W. Lee, and G. Gay. Real-time representation versus response elicitation in biosensor data. pages 605–608. ACM, 2015.

[182] T. Matthews, M. Czerwinski, G. Robertson, and D. Tan. Clipping lists and change borders: improving multitasking efficiency with peripheral information design. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 989–998. ACM, 2006.

[183] S. M. Mattingly, J. M. Gregg, P. Audia, A. E. Bayraktaroglu, A. T. Campbell, N. V. Chawla, V. Das Swain, M. De Choudhury, S. K. D'Mello, A. K. Dey, G. Gao, K. Jagannath, K. Jiang, S. Lin, Q. Liu, G. Mark, G. J. Martinez, K. Masaba, S. Mirjafari, E. Moskal, R. Mulukutla, K. Nies, M. D. Reddy, P. Robles-Granda, K. Saha, A. Sirigiri, and A. Striegel. 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*, pages 1–8, Glasgow Scotland Uk, May 2019. ACM.

[184] M. Mazmanian and I. Erickson. The product of availability: understanding the economic underpinnings of constant connectivity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 763–772, Toronto Ontario Canada, Apr. 2014. ACM.

[185] M. Mazmanian, J. Yates, and W. Orlikowski. Ubiquitous email: individual experiences and organizational consequences of blackberry use. In *Academy of Management Proceedings*, volume 2006, pages D1–D6. Academy of Management Briarcliff Manor, NY 10510, 2006.

[186] R. R. McCrae and P. T. Costa Jr. A five-factor theory of personality. *Handbook of personality: Theory and research*, 2(1999):139–153, 1999.

[187] N. McDonald, S. Schoenebeck, and A. Forte. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, Nov. 2019.

[188] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. AffectAura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 849–858. ACM, 2012.

[189] D. J. McDuff, J. Hernandez, S. Gontarek, and R. Picard. Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4000–4004. ACM, 2016.

[190] B. S. McEwen and T. Seeman. Stress and affect: Applicability of the concepts of allostasis and allostatic load. 2003.

[191] B. S. McEwen and E. Stellar. Stress and the individual: mechanisms leading to disease. *Archives of internal medicine*, 153(18):2093–2101, 1993.

[192] D. C. McFarlane and K. A. Latorella. The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17(1):1–61, 2002.

[193] N. Meiran. Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1423–1442, 1996.

[194] S. Melamed, A. Shirom, S. Toker, S. Berliner, and I. Shapira. Burnout and risk of cardiovascular disease: Evidence, possible causal paths, and promising research directions. *Psychological bulletin*, 132(3):327, 2006. Publisher: American Psychological Association.

[195] A. N. Meyer, G. C. Murphy, T. Zimmermann, and T. Fritz. Design Recommendations for Self-Monitoring in the Workplace: Studies in Software Development. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–24, Dec. 2017.

[196] S. Michie. Causes and management of stress at work. *Occupational and environmental medicine*, 59(1):67–72, 2002. Publisher: BMJ Publishing Group Ltd.

[197] Min Wu, Hong Cao, Hai-Long Nguyen, K. Surmacz, and C. Hargrove. Modeling perceived stress via HRV and accelerometer sensor streams. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1625–1628, Milan, Aug. 2015. IEEE.

[198] C. Morgan and S. R. Cotten. The Relationship between Internet Activities and Depressive Symptoms in a Sample of College Freshmen. *CyberPsychology & Behavior*, 6(2):133–142, Apr. 2003.

[199] C. A. Moyer, D. T. Stern, S. J. Katz, and A. M. Fendrick. We got mail": electronic communication between physicians and patients. *Am J Manag Care*, 5(12):1513–1522, 1999.

[200] A. Muaremi, B. Arnrich, and G. Tröster. Towards measuring stress with smartphones and wearable devices during workday and sleep. *BioNanoScience*, 3(2):172–183, 2013. Publisher: Springer.

[201] J. E. Munoz, F. Pereira, and E. Karapanos. Workload management through glanceable feedback: The role of heart rate variability. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6, Munich, Germany, Sept. 2016. IEEE.

[202] D. R. Murphy, A. N. D. Meyer, E. Russo, D. F. Sittig, L. Wei, and H. Singh. The Burden of Inbox Notifications in Commercial Electronic Health Records. *JAMA Internal Medicine*, 176(4):559, Apr. 2016.

[203] D. R. Murphy, B. Reis, D. F. Sittig, and H. Singh. Notifications Received by Primary Care Practitioners in Electronic Health Records: A Taxonomy and Time Analysis. *The American Journal of Medicine*, 125(2):209.e1–209.e7, Feb. 2012.

[204] D. R. Murphy, T. Satterly, T. D. Giardina, D. F. Sittig, and H. Singh. Practicing Clinicians' Recommendations to Reduce Burden from the Electronic Health Record Inbox: a Mixed-Methods Study. *Journal of General Internal Medicine*, 34(9):1825–1832, Sept. 2019.

[205] H. Mönnikes, J. Tebbe, M. Hildebrandt, P. Arck, E. Osmanoglou, M. Rose, B. Klapp, B. Wiedenmann, and I. Heymann-Mönnikes. Role of stress in functional gastrointestinal disorders. *Digestive Diseases*, 19(3):201–211, 2001.

[206] Y. Nakashima, J. Kim, S. Flutura, A. Seiderer, and E. André. Stress recognition in daily work. In *International Symposium on Pervasive Computing Paradigms for Mental Health*, pages 23–33. Springer, 2015.

[207] R. W. Neufeld. Evidence of stress as a function of experimentally altered appraisal of stimulus aversiveness and coping adequacy. *Journal of Personality and Social Psychology*, 33(5):632–646, 1976.

[208] D. Nunan, G. R. H. Sandercock, and D. A. Brodie. A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults: REVIEW OF SHORT-TERM HRV VALUES. *Pacing and Clinical Electrophysiology*, 33(11):1407–1417, Nov. 2010.

[209] B. O'Conaill and D. Frohlich. Timespace in the workplace: Dealing with interruptions. In *Conference companion on Human factors in computing systems*, pages 262–263. ACM, 1995.

[210] P. O'Connor, J. Nguyen, and J. Anglim. Effectively coping with task stress: A study of the validity of the Trait Emotional Intelligence Questionnaire–Short Form (TEIQue–SF). *Journal of personality assessment*, 99(3):304–314, 2017. Publisher: Taylor & Francis.

[211] S. Ohly, A. S. Göritz, and A. Schmitt. The power of routinized task behavior for energy at work. *Journal of Vocational Behavior*, 103:132–142, Dec. 2017.

[212] J. Okkonen, T. Heimonen, R. Savolainen, and M. Turunen. Assessing Information Ergonomics in Work by Logging and Heart Rate Variability. In *International Conference on Applied Human Factors and Ergonomics*, pages 425–436. Springer, 2017.

[213] T. Okoshi, H. Nozaki, J. Nakazawa, H. Tokuda, J. Ramos, and A. K. Dey. Towards attention-aware adaptive notification on smart phones. *Pervasive and Mobile Computing*, 26:17–34, Feb. 2016.

[214] T. Okoshi, J. Ramos, H. Nozaki, J. Nakazawa, A. K. Dey, and H. Tokuda. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile

environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pages 475–486, Osaka, Japan, 2015. ACM Press.

[215] A. J. Oldehinkel, J. Ormel, N. M. Bosch, E. M. C. Bouma, A. M. Van Roon, J. G. M. Rosmalen, and H. Riese. Stressed out? Associations between perceived and physiological stress responses in adolescents: The TRAILS study: Perceived and physiological stress responses. *Psychophysiology*, 48(4):441–452, Apr. 2011.

[216] A. Oulasvirta and P. Saariluoma. Long-term working memory and interrupting messages in human–computer interaction. *Behaviour & Information Technology*, 23(1):53–64, 2004.

[217] D. Ouyang, J. H. Chen, J. Hom, and J. Chi. Internal medicine resident computer usage: an electronic audit of an inpatient service. *JAMA internal medicine*, 176(2):252–254, 2016.

[218] S. Palojoki, T. Pajunen, K. Saranto, and L. Lehtonen. Electronic Health Record-Related Safety Concerns: A Cross-Sectional Survey of Electronic Health Record Users. *JMIR Medical Informatics*, 4(2):e13, May 2016.

[219] P. Paredes and M. Chan. CalmMeNow: exploratory research and design of stress mitigating mobile interventions. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, page 1699, Vancouver, BC, Canada, 2011. ACM Press.

[220] J. H. Park and M. Liu. Multitasking in e-learning environments: users' multitasking strategies and design implications. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1991–1996. ACM, 2012.

[221] L. Parkitny and J. McAuley. The Depression Anxiety Stress Scale (DASS). *Journal of Physiotherapy*, 56(3):204, 2010.

[222] I. Pavlidis, M. Dcosta, S. Taamneh, M. Manser, T. Ferris, R. Wunderlich, E. Akleman, and P. Tsiamyrtzis. Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors. *Scientific reports*, 6:25651, 2016. Publisher: Nature Publishing Group.

[223] J. G. Phillips and L. Reddie. Decisional style and self-reported Email use in the workplace. *Computers in Human Behavior*, 23(5):2414–2428, Sept. 2007.

[224] R. W. Picard. Affective computing. 1995.

[225] R. W. Picard, S. Fedor, and Y. Ayzenberg. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, 8(1):62–75, 2016. Publisher: SAGE Publications Sage UK: London, England.

[226] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1175–1191, 2001. Publisher: Ieee.

[227] T. G. Pickering. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current hypertension reports*, 3(3):249–254, 2001.

[228] S. D. Pressman and S. Cohen. Does positive affect influence health? *Psychological bulletin*, 131(6):925, 2005.

[229] S. P. Proctor, R. White, T. Robins, D. Echeverria, and A. Rocskay. Effect of overtime work on cognitive function in automotive workers. *Scandinavian Journal of Work, Environment & Health*, 22(2):124–132, Apr. 1996.

[230] J. C. Quick and C. D. Spielberger. Walter Bradford cannon: pioneer of stress research. *International Journal of Stress Management*, 1(2):141–143, 1994. Publisher: Springer.

[231] M. Rabbi, S. Ali, T. Choudhury, and E. Berke. Passive and In-Situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11*, page 385, Beijing, China, 2011. ACM Press.

[232] A. Raij, A. Ghosh, S. Kumar, and M. Srivastava. Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment. pages 11–20. ACM, 2011.

[233] M. Reed, I. Graetz, N. Gordon, and V. Fung. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. *Am J Manag Care*, 21(12):e632–e639, 2015.

[234] M. E. Reed, J. Huang, R. Brand, D. Ballard, C. Yamin, J. Hsu, and R. Grant. Communicating through a patient portal to engage family care partners. *JAMA internal medicine*, 178(1):142–144, 2018.

[235] M. E. Reed, J. Huang, A. Millman, I. Graetz, J. Hsu, R. Brand, D. W. Ballard, and R. Grant. Portal Use Among Patients With Chronic Conditions: Patient-reported Care Experiences. *Medical care*, 57(10):809–814, 2019.

[236] L. Reinecke, S. Aufenanger, M. E. Beutel, M. Dreier, O. Quiring, B. Stark, K. Wölfling, and K. W. Müller. Digital Stress over the Life Span: The Effects of Communication Load and Internet Multitasking on Perceived Stress and Psychological Health Impairments in a German Probability Sample. *Media Psychology*, 20(1):90–115, Jan. 2017.

[237] R. Reiner. Integrating a portable biofeedback device into clinical practice for patients with anxiety disorders: Results of a pilot study. *Applied Psychophysiology and Biofeedback*, 33(1):55–61, 2008.

[238] X. Ren, B. Yu, Y. Lu, and A. Brombacher. Exploring Cooperative Fitness Tracking to Encourage Physical Activity among Office Workers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–20, Nov. 2018.

[239] K. Renaud, J. Ramsay, and M. Hair. " You've got e-mail!"... shall I deal with it now? Electronic mail from the recipient's perspective. *International Journal of Human-Computer Interaction*, 21(3):313–332, 2006.

[240] D. Reynolds. Gaussian Mixture Models. In S. Z. Li and A. Jain, editors, *Encyclopedia of Biometrics*, pages 659–663. Springer US, Boston, MA, 2009.

[241] K. Richardson and R. Benbunan-Fich. Examining the antecedents of work connectivity behavior during non-work time. *Information and Organization*, 21(3):142–160, 2011.

[242] V. Rivera-Pelayo, A. Fessl, L. Müller, and V. Pammer. Introducing Mood Self-Tracking at Work: Empirical Insights from Call Centers. *ACM Transactions on Computer-Human Interaction*, 24(1):1–28, Mar. 2017.

[243] N. H. Rod, M. Grønbaek, P. Schnohr, E. Prescott, and T. S. Kristensen. Perceived stress as a risk factor for changes in health behaviour and cardiac risk profile: a longitudinal study. *Journal of Internal Medicine*, 266(5):467–475, Nov. 2009.

[244] S. Rosenthal, A. K. Dey, and M. Veloso. Using decision-theoretic experience sampling to build personalized mobile phone interruption models. pages 170–187. Springer, 2011.

[245] H. Rusko, T. Rönkä, A. Uusitalo, U. Kinnunen, S. Mauno, T. Feldt, M. Kinnunen, K. Martinmäki, A. Hirvonen, S. Hyttinen, and others. Stress and relaxation during sleep and awake time, and their associations with free salivary cortisol after awakening. In *Nordic Ergonomics Society congress*, 2006.

[246] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[247] H. S. Saag, K. Shah, S. A. Jones, P. A. Testa, and L. I. Horwitz. Pajama Time: Working After Work in the Electronic Health Record. *Journal of general internal medicine*, pages 1–2, 2019.

[248] P. Sanches, K. Höök, E. Vaara, C. Weymann, M. Bylund, P. Ferreira, N. Peira, and M. Sjölinder. Mind the body!: designing a mobile stress management application encouraging personal reflection. pages 47–56. ACM, 2010.

[249] A. Sano, A. J. Phillips, Z. Y. Amy, A. W. McHill, S. Taylor, N. Jaques, C. A. Czeisler, E. B. Klerman, and R. W. Picard. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. pages 1–6. IEEE, 2015.

[250] A. Sano and R. W. Picard. Stress recognition using wearable sensors and mobile phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 671–676. IEEE, 2013.

[251] H. Sarker, M. Tyburski, M. M. Rahman, K. Hovsepian, M. Sharmin, D. H. Epstein, K. L. Preston, C. D. Furr-Holden, A. Milam, and I. Nahum-Shani. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. pages 4489–4501. ACM, 2016.

[252] H. Sato and J.-i. Kawahara. Selective bias in retrospective self-reports of negative mood states. *Anxiety, Stress & Coping*, 24(4):359–367, July 2011.

[253] S. Schieman and M. C. Young. Are communications about work outside regular working hours associated with work-to-family conflict, psychological distress and sleep problems? *Work & Stress*, 27(3):244–261, July 2013.

[254] W. Schlotz, R. Kumsta, I. Layes, S. Entringer, A. Jones, and S. Wüst. Covariance Between Psychological and Endocrine Responses to Pharmacological Challenge and Psychosocial Stress: A Question of Timing:. *Psychosomatic Medicine*, 70(7):787–796, Sept. 2008.

[255] D. Schneider and K. Harknett. Consequences of Routine Work-Schedule Instability for Worker Health and Well-Being. *American Sociological Review*, 84(1):82–114, Feb. 2019.

[256] S. M. Schueller. Personality fit and positive interventions: Extraverted and introverted individuals benefit from different happiness increasing strategies. *Psychology*, 3(12):1166, 2012.

[257] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2009. Publisher: IEEE.

[258] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2010.

[259] T. D. Shanafelt, L. N. Dyrbye, C. Sinsky, O. Hasan, D. Satele, J. Sloan, and C. P. West. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. In *Mayo Clinic Proceedings*, volume 91, pages 836–848. Elsevier, 2016.

[260] J. A. Shenson, R. M. Cronin, S. E. Davis, Q. Chen, and G. P. Jackson. Rapid growth in surgeons' use of secure messaging in a patient portal. *Surgical endoscopy*, 30(4):1432–1440, 2016.

[261] H. Shibata and K. Omura. Docking window framework: supporting multitasking by docking windows. In *Proceedings of the 10th asia pacific conference on Computer human interaction*, pages 227–236. ACM, 2012.

[262] S. Shiffman, A. A. Stone, and M. R. Hufford. Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, 4(1):1–32, Apr. 2008.

[263] J. Siegrist. Effort-reward imbalance at work and health. In *Research in Occupational Stress and Well-being*, volume 2, pages 261–291. Emerald (MCB UP ), Bingley, 2002.

[264] N. L. Sin, R. P. Sloan, P. S. McKinley, and D. M. Almeida. Linking Daily Stress Processes and Laboratory-Based Heart Rate Variability in a National Sample of Midlife and Older Adults. *Psychosomatic Medicine*, 78(5):573–582, June 2016.

[265] H. Singh, C. Spitzmueller, N. J. Petersen, M. K. Sawhney, and D. F. Sittig. Information Overload and Missed Test Results in Electronic Health Record–Based Settings. *JAMA Internal Medicine*, 173(8):702, Apr. 2013.

[266] H. Singh, C. Spitzmueller, N. J. Petersen, M. K. Sawhney, M. W. Smith, D. R. Murphy, D. Espadas, A. Laxmisan, and D. F. Sittig. Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey. *Journal of the American Medical Informatics Association*, 20(4):727–735, July 2013.

[267] E. Smets, P. Casale, U. Großekathöfer, B. Lamichhane, W. De Raedt, K. Bogaerts, I. Van Diest, and C. Van Hoof. Comparison of machine learning techniques for psychophysiological stress detection. In *International Symposium on Pervasive Computing Paradigms for Mental Health*, pages 13–22. Springer, 2015.

[268] R. E. Smith, C. Fagan, N. L. Wilson, J. Chen, M. Corona, H. Nguyen, S. Racz, and Y. Shoda. Internet-based approaches to collaborative therapeutic assessment: New opportunities for professional psychologists. *Professional Psychology: Research and Practice*, 42(6):494, 2011.

[269] J. Snyder, M. Matthews, J. Chien, P. F. Chang, E. Sun, S. Abdullah, and G. Gay. Moodlight: Exploring personal and social implications of ambient display of biosensor data. pages 143–153. ACM, 2015.

[270] J. Y. Soh, S.-H. Jung, W. C. Cha, M. Kang, D. K. Chang, J. Jung, J. Lee, J. S. Choi, and K. Kim. Variability in Doctors' Usage Paths of Mobile Electronic Health Records Across Specialties: Comprehensive Analysis of Log Data. *JMIR mHealth and uHealth*, 7(1):e12041, Jan. 2019.

[271] S. Sonnentag, L. Reinecke, J. Mata, and P. Vorderer. Feeling interrupted—Being responsive: How online messages relate to affect at work. *Journal of Organizational Behavior*, 39(3):369–383, 2018.

[272] C. D. Spielberger, R. L. Gorsuch, and R. E. Lushene. Manual for the state-trait anxiety inventory. 1970.

[273] A. Springer, V. Hollis, and S. Whittaker. Dice in the Black Box: User Experiences with an Inscrutable Algorithm. *arXiv:1812.03219 [cs]*, Dec. 2018. arXiv: 1812.03219.

[274] A. Spurgeon, J. M. Harrington, and C. L. Cooper. Health and safety problems associated with long working hours: a review of the current position. *Occupational and Environmental Medicine*, 54(6):367–375, June 1997.

[275] S. Sriramprakash, V. D. Prasanna, and O. R. Murthy. Stress detection in working people. *Procedia computer science*, 115:359–366, 2017. Publisher: Elsevier.

[276] A. Starczewski and A. Krzyżak. Performance Evaluation of the Silhouette Index. In L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, editors, *Artificial Intelligence and Soft Computing*, volume 9120, pages 49–58. Springer International Publishing, Cham, 2015. Series Title: Lecture Notes in Computer Science.

[277] Steven Aldana. Wearables and Wellness Programs: The Complete Guide, July 2020.

[278] M. A. Stults-Kolehmainen and R. Sinha. The effects of stress on physical activity and exercise. *Sports medicine*, 44(1):81–121, 2014.

[279] A. Stys and T. Stys. Current clinical applications of heart rate variability. *Clinical Cardiology*, 21(10):719–724, 1998.

[280] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss. Activity-aware mental stress detection using physiological sensors. In *International Conference on Mobile Computing, Applications, and Services*, pages 282–301. Springer, 2010.

[281] M. Swan. Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International journal of environmental research and public health*, 6(2):492–525, 2009.

[282] M. Swan. Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *Journal of Personalized Medicine*, 2(3):93–118, Sept. 2012.

[283] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel. Influence of Mental Stress on Heart Rate and Heart Rate Variability. In R. Magjarevic, J. H. Nagel, J. Vander Sloten, P. Verdonck, M. Nyssen, and J. Haueisen, editors, *4th European Conference of the International Federation for Medical and Biological Engineering*, volume 22, pages 1366–1369. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. Series Title: IFMBE Proceedings.

[284] M. Tai-Seale, E. C. Dillon, Y. Yang, R. Nordgren, R. L. Steinberg, T. Nauenberg, T. C. Lee, A. Meehan, J. Li, A. S. Chan, and others. Physicians' Well-Being Linked To In-Basket Messages Generated By Algorithms In Electronic Health Records. *Health Affairs*, 38(7):1073–1078, 2019.

[285] S. Tak, A. Cockburn, K. Humm, D. Ahlström, C. Gutwin, and J. Scarr. Improving window switching interfaces. In *IFIP Conference on Human-Computer Interaction*, pages 187–200. Springer, 2009.

[286] M. Tarafdar, Q. Tu, B. S. Ragu-Nathan, and T. S. Ragu-Nathan. The Impact of Technostress on Role Stress and Productivity. *Journal of Management Information Systems*, 24(1):301–328, July 2007.

[287] S. E. Taylor, M. E. Kemeny, G. M. Reed, J. E. Bower, and T. L. Gruenewald. Psychological resources, positive illusions, and health. *American psychologist*, 55(1):99, 2000.

[288] O. o. t. N. C. f. H. I. Technology. Office-based Physician Electronic Health Record Adoption, 2019.

[289] T. Teisala, S. Mutikainen, A. Tolvanen, M. Rottensteiner, T. Leskinen, J. Kaprio, M. Kolehmainen, H. Rusko, and U. M. Kujala. Associations of physical activity, fitness, and body composition with heart rate variability–based indicators of stress and recovery on workdays: a cross-sectional study. *Journal of Occupational Medicine and Toxicology*, 9(1):16, 2014.

[290] S. Thomée, M. Eklöf, E. Gustafsson, R. Nilsson, and M. Hagberg. Prevalence of perceived stress, symptoms of depression and sleep disturbances in relation to information and communication technology (ICT) use among young adults–an explorative prospective study. *Computers in Human Behavior*, 23(3):1300–1321, 2007.

[291] J. Tomaka, J. Blascovich, J. Kibler, and J. M. Ernst. Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 73(1):63–72, 1997.

[292] J. G. Trafton and C. A. Monk. Task interruptions. *Reviews of human factors and ergonomics*, 3(1):111–126, 2007.

[293] P. Tsiamyrtzis, M. Dcosta, D. Shastri, E. Prasad, and I. Pavlidis. Delineating the operational envelope of mobile and conventional EDA sensing on key body locations. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5665–5674. ACM, 2016.

[294] A. Uusitalo, T. Mets, K. Martinmäki, S. Mauno, U. Kinnunen, and H. Rusko. Heart rate variability related to effort at work. *Applied Ergonomics*, 42(6):830–838, Nov. 2011.

[295] S. Vainiomäki, A.-M. Aalto, T. Lääveri, T. Sinervo, M. Elovainio, P. Mäntyselkä, and H. Hyppönen. Better Usability and Technical Stability Could Lead to Better Work-Related Well-Being among Physicians. *Applied Clinical Informatics*, 08(04):1057–1067, 2017.

[296] M. Van Der Hulst and S. Geurts. Associations between overtime and psychological health in high and low reward jobs. *Work & Stress*, 15(3):227–240, July 2001.

[297] H. Van Steenis and J. Tulen. The effects of physical activities on cardiovascular variability in ambulatory situations [ECG/accelerometry analysis]. volume 1, pages 105–108. IEEE, 1997.

[298] T. B. VanItallie. Stress: A risk factor for serious illness. *Metabolism*, 51(6):40–45, June 2002.

[299] L. M. Vizer. Different strokes for different folks: individual stress response as manifested in typed text. pages 2773–2778. ACM, 2013.

[300] L. M. Vizer, L. Zhou, and A. Sears. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886, 2009. Publisher: Elsevier.

[301] N. Vogelzangs, A. T. F. Beekman, Y. Milaneschi, S. Bandinelli, L. Ferrucci, and B. W. J. H. Penninx. Urinary Cortisol and Six-Year Risk of All-Cause and Cardiovascular Mortality. *The Journal of Clinical Endocrinology & Metabolism*, 95(11):4959–4964, Nov. 2010.

[302] S. Voida, D. J. Patterson, and S. N. Patel. Sensor Data Streams. In *Ways of Knowing in HCI*, pages 291–321. Springer, 2014.

[303] M. von Entreß-Fürsteneck, A. Buchwald, and N. Urbach. Will I or Will I Not? Explaining the Willingness to Disclose Personal Self-Tracking Data to a Health Insurance Company. 2019.

[304] J. Wajcman and E. Rose. Constant Connectivity: Rethinking Interruptions at Work. *Organization Studies*, 32(7):941–961, July 2011.

[305] M. Waldner, M. Steinberger, R. Grasset, and D. Schmalstieg. Importance-driven compositing window management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 959–968. ACM, 2011.

[306] M. Wallerg\aard, P. Jönsson, G. Johansson, and B. Karlson. A virtual reality version of the Trier Social Stress Test: A pilot study. *Presence*, 20(4):325–336, 2011. Publisher: MITP.

[307] E. J. Wang, J. Zhu, M. Jain, T.-J. Lee, E. Saba, L. Nachman, and S. N. Patel. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 425. ACM, 2018.

[308] J. K. Wang, D. Ouyang, J. Hom, J. Chi, and J. H. Chen. Characterizing electronic health record usage patterns of inpatient medicine residents using event log data. *PloS one*, 14(2):e0205379, 2019.

[309] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. pages 3–14. ACM, 2014.

[310] W. Wang, G. M. Harari, R. Wang, S. R. Müller, S. Mirjafari, K. Masaba, and A. T. Campbell. Sensing Behavioral Change over Time: Using Within-Person Variability Features from Mobile Sensing to Predict Personality Traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–21, Sept. 2018.

[311] A. Warr, E. H. Chi, H. Harris, A. Kuscher, J. Chen, R. Flack, and N. Jitkoff. Window Shopping: A Study of Desktop Window Switching. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3335–3338. ACM, 2016.

[312] J. Warshaw, T. Matthews, S. Whittaker, C. Kau, M. Bengualid, and B. A. Smith. Can an Algorithm Know the "Real You"?: Understanding People's Reactions to Hyper-personal Analytics Systems. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 797–806, Seoul, Republic of Korea, 2015. ACM Press.

[313] P. C. Wason. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12(3):129–140, July 1960.

[314] M. Weigl, J. Beck, M. Wehler, and A. Schneider. Workflow interruptions and stress atwork: a mixed-methods study among physicians and nurses of a multidisciplinary emergency department. *BMJ Open*, 7(12):e019074, Dec. 2017.

[315] A. T. WELFORD. Stress and performance. *Ergonomics*, 16(5):567–580, 1973.

[316] J. I. Westbrook, E. Coiera, W. T. M. Dunsmuir, B. M. Brown, N. Kelk, R. Paoloni, and C. Tran. The impact of interruptions on clinical task completion. *Quality and Safety in Health Care*, 19(4):284–289, Aug. 2010.

[317] J. I. Westbrook, M. Z. Raban, S. R. Walter, and H. Douglas. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: a prospective, direct observation study. *BMJ Quality & Safety*, 27(8):655–663, Aug. 2018.

[318] C. D. Wickens. Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):449–455, June 2008.

[319] J. Wijsman, B. Grundlehner, H. Liu, J. Penders, and H. Hermens. Wearable physiological sensors reflect mental stress state in office-like situations. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 600–605. IEEE, 2013.

[320] D. C. Williams, R. W. Warren, M. Ebeling, A. L. Andrews, and R. J. Teufel II. Physician Use of Electronic Health Records: Survey Study Assessing Factors Associated With Provider Reported Satisfaction and Perceived Patient Impact. *JMIR Medical Informatics*, 7(2):e10949, Apr. 2019.

[321] X. Xu, P. Chikersal, A. Doryab, D. K. Villalba, J. M. Dutcher, M. J. Tumminia, T. Althoff, S. Cohen, K. G. Creswell, J. D. Creswell, J. Mankoff, and A. K. Dey. Leveraging Routine Behavior and Contextually-Filtered Features for Depression Detection among College Students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–33, Sept. 2019.

[322] R. Yang, E. Shin, M. W. Newman, and M. S. Ackerman. When fitness trackers don't 'fit': end-user difficulties in the assessment of personal tracking device accuracy. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, pages 623–634, Osaka, Japan, 2015. ACM Press.

[323] R. M. Yerkes and J. D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology*, 18(5):459–482, 1908.

[324] S. Yoon and K.-p. Lee. A Study on Notification System Design of Smartphone Messenger Considering the User's Stress. *Archives of Design Research*, 28(2):75, May 2015.

[325] B. Yu, J. Hu, M. Funk, and L. Feijs. DeLight: biofeedback through ambient light for stress intervention and relaxation assistance. *Personal and Ubiquitous Computing*, 22(4):787–805, Aug. 2018.

[326] J. Zhai and A. Barreto. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In *2006 international conference of the IEEE engineering in medicine and biology society*, pages 1355–1358. IEEE, 2006.

[327] J. Zhai and A. Barreto. Stress Recognition Using Non-invasive Technology. In *FLAIRS Conference*, pages 395–401, 2006.

[328] P. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez. Affective Computing—A Rationale for Measuring Mood With Mouse and Keyboard. *International Journal of Occupational Safety and Ergonomics*, 9(4):539–551, Jan. 2003.

[329] T. Âkerstedt. Psychosocial stress and impaired sleep. *Scandinavian journal of work, environment & health*, pages 493–501, 2006.

# Appendix A

# Recruitment flyer

A STUDY OF WORKPLACE STRESS USING WEARABLE DEVICES

## STRESSED AT WORK?

## Track your stress with a smartwatch

We are recruiting participants for a research study about stress in the workplace. This study may help us to better understand events, interactions, and contexts that surround high-stress episodes.

Data collection is unobtrusive, using a smartwatch and a computer application.

You will be asked to wear a smartwatch, install a mobile application, and install a computer application. These applications will run for the period of **3 weeks** and will not interfere with your work.

*You will receive a $50 gift card at the end of the study*

**Fill out this form to determine your eligibility:**

For more information, contact Fatema Akbar at fatemaa@uci.edu

**RESEARCH TEAM**

| **Lead Researcher** | **Faculty Sponsor** |
| --- | --- |
| Fatema Akbar<br>PhD Student<br>Department of Informatics<br>Donald Bren School of Information and Computer Sciences<br>University of California, Irvine<br><br>Email: fatemaa@uci.edu | Gloria Mark<br>Professor<br>Department of Informatics<br>Donald Bren School of Information and Computer Sciences<br>University of California, Irvine |

UCI IRB Approved: 12-11-2019 | MOD# 27056 | HS# 2019-4895

Figure A.1: Recruitment flyer sent by email to all UCI employees.

# Appendix B

# Screener survey

Thank you for your interest in participating in the ”In-Situ Study of Workplace Stress Using Wearable Devices”.

This survey will determine your eligibility to participate.

Please answer the following questions:

What is your age?

What is your sex? Male Female

What is the highest grade or level of school that you have completed?

High School Graduate

Some College

Graduated 2-year College

Graduated 4-year College

Post Graduate (e.g. MS, PhD)

Prefer not to answer

Do you have access to a smartphone with internet access? Yes No

Do you have access to a work computer with internet access? Yes No

How much of your workday do you spend working on the computer?

All of my workday

Most of my workday

About half of my workday

Less than half of my workday

None at all

What are your typical work hours?

9am to 5pm

8am to 4pm

Other (please specify)

Are you a UCI employee? Yes No

Are you enrolled in another research study about managing stress? Yes No

Are you taking cardiac medications (e.g. beta-blockers, diltiazem, verapamil, or digoxin)?

Yes No

Do you use pacemakers or implantable cardiac defibrillators? Yes No

Have you been previously diagnosed with atrial or ventricular arrhythmias? Yes No

What is your current height and weight?

height:

weight (rounded to nearest pound):

Please enter your first name and last name in the form below.

First Name:

Last Name

What is your email address? Email Address

What is the best daytime telephone number to reach you at?

Daytime Telephone Number:

Please leave any additional information that you would like us to know in the space provided below.

Powered by Qualtrics