Digital Biomarker Models for Prediction of Infectious Disease Susceptibility

by

Yaya Zhai

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in the University of Michigan
2021

Doctoral Committee:

Professor Alfred O. Hero, III, Chair
Professor Brian D. Athey
Professor Jun Li
Professor Kayvan Najarian
Associate Professor Maureen A. Sartor

Yaya Zhai

yayazhai@umich.edu

ORCID iD: 0000-0002-0907-866X

DEDICATION

This dissertation is dedicated to my parents, Yong'an Zhai and Qin Wang, and my husband, Gang Yan.

For their everlasting and unconditional love, understanding and support.

ACKNOWLEDGMENTS

This dissertation would never have been possible without the instruction, guidance and support from Professor Alfred Hero. I will be eternally grateful for everything he has done to help me with my doctoral researches. My research background before coming to UMich was mostly in biology, chemistry, and pharmaceutical sciences, all of which are wet-lab experiences. Professor Hero brought me into the realm of statistical learning and data analysis, which is completely new to me and really eye-opening. Professor Hero demonstrated how powerful, if not miraculous, customized statistical models and algorithms can be in detecting hidden patterns in medical data and health records, which is quite challenging due to the small sample size, high dimensionality and low signal-to-noise ratio. He also demonstrated how the distinct patterns will improve medical practice and wellness. His depth and width of knowledge and intuition in scientific research are unforgettable. It impresses me that he is dedicated to and rigorous in research while still being considerate and pleasant to his students who need assistance. Scientific research is never easy since it involves exploring the unknown, but he made it less difficult for me with his knowledge and continuous guidance and encouragement. He supported me in every way he can and he is always there whenever I need help. I owe my greatest appreciation to him and I am very fortunate to have the opportunity to have him as my mentor.

I want to express my gratitude to my collaborators, Dr. Geoffrey Ginsberg, Dr. Murali Doraiswamy, Dr. Ricardo Henao and Dr. Christopher Woods from Duke University, and Dr. Peter Song, Dr. Xichen She from UMich Biostatitics Department for the invaluable advice and help offered to my research. Collaborators from Duke University have provided me excellent data as well as biological insights. I had many inspiring discussions with Dr. Song and Dr. She.

I would like to thank my thesis committee members, Dr. Brian Athey, Dr. Jun Li, Dr. Kayvan Najarian, and Dr. Maureen Sartor. Every meeting has been a learning experience for me, and I am appreciative for all of your suggestions. Your patience and support mean a lot to me.

Thank you to directors and staffs from the Department of Bioinformatics and EECS, my course instructors and GSIs, my friends at Ann Arbor and members of the Hero Lab.

Finally, I would thank my family. My parents have always supported me wholeheartedly and to the best of their ability at every stage of my life, without which I would not have been able to achieve what I have today. My husband has always been at my side, and we've experienced so many memorable moments.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

ABSTRACT

Acute respiratory viral infection (ARVI) represents one of the most prevalent infectious diseases affecting mankind. With the threat of COVID-19 still looming over us, we have witnessed the substantial threat ARVI poses to world health and economy, extinguishing millions of lives and costing trillions of dollars. This sets the context for the research of this thesis: using digital biomarkers to distinguish between individuals who are susceptible to becoming severely infected and/or infectious before an infection is clinically detectable. The development of such biomarkers can have both clinical and epidemiological impact in terms of identifying individuals who are either vulnerable to severe infection or those who may become highly infectious. The digital biomarkers and associated analysis methods are developed and validated on longitudinal data collected by our clinical collaborators from two different ARVI challenge studies. The first study provides data on healthy human volunteers who were inoculated with the common cold and the second study provides data on volunteers inoculated with the flu. Digital biomarkers include molecular, physiological and cognitive data continuously collected from blood, wearable devices and cognitive testing of the study participants. The findings of our research on digitally measurable susceptibility factors are wide-ranging. We find that circadian rhythm at the molecular scale (biochronicity) plays an important role in mediating both the susceptibility and the response to severe infection, revealing groups of gene expression markers that differentiate the responses of low infected and high infected individuals. Using a high dimensional representation of physiological signals from a wearable device, we find that an infection response and its onset time can be reliably predicted at least 24 hours before peak infection time. We find that a certain measure of variability in pre-exposure cognitive function is highly associated with the post-exposure severity of infection.

CHAPTER 1

Introduction

1.1 Motivation and Research Background

First we provide motivation for the work described in this thesis. We then provide some background on the mechanisms of infection and susceptibility, digital biomarkers, circadian rhythm. We conclude the introduction with an outline of the contributions of the thesis.

1.1.1 Motivation

Acute respiratory infections have similar clinical symptoms including fever, runny nose, coughing, stuffed nose, sore throat, malaise etc. Colds, influenza, pneumonia, and COVID-19 are all examples of acute respiratory infections. Although numerous pathogens cause respiratory illness, over 80% cases are viral [MPR$^+$98]. The high diversity and fast mutation rates of viruses makes it difficult to develop effective detection, prevention and treatment strategies for acute respiratory viral infection.

Acute respiratory viral infection (ARVI) is a serious health and economic concern worldwide. While a mild common cold is often self-limited and can be resolved without treatment in 1-2 weeks for healthy individuals, other forms of ARVI can be much more dangerous. Acute respiratory viral infection is a leading cause of morbidity, hospitalization, and mortality with a global illness burden of millions of deaths [ML97]. COVID-19 alone has infected approximately 200 million individuals and claimed the lives of 4 million within less than 2 years [JHU]. ARVI treatment and control, including but not limited to quarantine, screening, medical treatment and vaccine, have global costs in the billions or trillions of dollars [JWL$^+$21]. A tantalizing hypothesis is that susceptibility to severe ARVI can be predicted using digital biomarkers before a person becomes symptomatically ill. If this were possible on a large scale, then one could forecast symptom severity and degree of infectiousness (viral shedding) of individuals in a population. This would clearly have profound implications on prospective disease prevention and treatment, and on our ability to manage pandemics by identifying and protecting vulnerable individuals. This thesis makes a step in this direction by investigating the value of digital biomarkers that are collected over time before

exposure and during infection. Using multi-modal digital biomarker data collected from challenge studies designed and executed by our clinical collaborators, we develop methods of analysis and make findings that may one day help early detection of susceptibility to ARVI.

1.1.2 Background: Human Viral Challenge (HVC) studies

Human Viral Challenge (HVC) studies have long been conducted to understand various aspects during infection by deliberately infecting volunteers with a pathogen of interest[LWNM+18]. They can be traced back to the famous smallpox challenge of James Phipps by Edward Jenner in 1796[DCM03]. HVC studies can provide valuable information on susceptibility factors, disease pathogenesis, diagnostic biomarkers and / or vaccine efficacy, depending on the design and purpose of the experiment [BHPB17].

An illustration of the experimental process and the data processing pipeline for the execution of a Human Viral Challenge (HVC) experiment and the analysis of its data is shown in Fig. 1.1. After a group of volunteers has been selected and pre-screened, they are monitored before and after being exposed to a viral pathogen. Associated molecular, physiological, and cognitive samples are gathered from each individual throughout time. Data extracted from these samples are called digital biomarkers. For molecular data these biomarkers are derived from gene sequencing data or metabolomic, proteomic assays. For physiological data, these biomarkers are derived from signals extracted from wearable devices, signals such as electrodermic activity (EDA), skin temperature, heart rate (HR), or acceleromenter (ACC) measuring physical movement. After inoculation some study participants may become symptomatically ill with viral shedding, some participants may have few or no symptoms but have viral shedding, others may have symptoms but no shedding, while others may have neither symptom nor shedding. The symptoms and shedding levels are also measured during the study, becoming part of the digital biomarker databases.

To build a predictive analysis framework for analyzing these digital biomarkers the data must first be pre-processed, often involving various forms of data cleaning, formatting and outlier detection. Then feature extraction is performed, often involving a combination of mathematical procedures, e.g., PCA or sparse logistic regression, and biological domain knowledge. Methods of statistical learning are increasingly employed to develop robust and generalizable predictors based on a minimum of domain knowledge.

Data from two HVC studies are used in this dissertation. The first study performed in 2015 at the University of Virginia, called the Biochronicity study, recruited 24 volunteers and their molecular, physiological, hormonal and cognitive samples were collected both before and after the inoculation of Human Rhinovirus (HRV). This experiment was designed to explore the role of biochronicity in infection susceptibility and infection response. The results of Chapter 2 and Chapter 4 are based on analysis of data from this experiment. The other study, called Prometheus,

challenged 39 individuals with influenza virus. The Prometheus challenge study was performed at University College London in 2017. The experiment was designed to establish the feasibility of predicting infectiousness, measured by amount of shedding, of infected participants before peak symptom occurs. The wearable data from this experiment was used in Chapter 3. Detailed protocol and information of these experiments will be described in details in corresponding chapters.

1.1.3 Background: Mechanisms of infection and susceptibility

Systemic acute respiratory viral infection occurs when a viral pathogen enters the body of the host, often through the naso-pharyngal mucosa, and replicate sufficiently that it is able to travel and invade multiple cells throughout the respiratory tract. The success of the virus in infecting the host is determined by many factors, the major factor being the effectivenss of host immune response. Molecular mechanism of host immune response has been investigated both *in vitro* and *in vivo*. Viruses are detected by pattern recognition receptors (PRRs) which recognize pathogen-associated molecular patterns (PAMPs). Important PRRs include NOD-like receptors (NLRs), Toll-like receptors (TLRs), retinoic acid-inducible gene I (RIG-I), and other cytosolic virus sensors. Interferons (IFNs) and other proinflammatory mediators (e.g. cytokines, chemokines) are released when these PRRs are activated, triggering the host's innate and adaptive immune responses[NCB16]. A set of peripheral blood gene expression signatures containing different PPRs has been identified to precisely distinguish individuals with symptomatic ARVIs from resistant individuals [ZCV+09b]. Different temporal host molecular responses were examined in symptomatic hosts and asymptomatic subjects. Multiple PRRs-mediated antiviral and inflammatory responses were invoked in the former while the latter has increased expression of genes with annotations related to antioxidant responses and cell-mediated responses [HZR+11]

There are many factors that influence population vulnerability. Sometimes genotype is one reason why some people are more vulnerable than others. For example, ACE2 gene is known to encode the entry receptor for SARS-CoV-2, whose genetic polymorphisms may influence the binding affinity and the subsequent cell invasion efficiency [GPPK+20]. Aside from genotype, infection susceptibility is influenced by lifestyle. Sleep deprivation has been shown in studies to have a negative impact on immune system components that are important for host resistance to infectious disease [PL16]. Specifically, short sleep duration and sleep disruptions indicate greater susceptibility to upper respiratory infection [PJDHC15, CDA+09]. In addition, longer duty hours, stress, shift work and poor sleep quality appeared to be a risk factor to infectious diseases such as COVID-19 [RCW+20].

1.1.4 Background: Digital Health

Wearable digital devices, databases, and technologies have fundamentally changed healthcare and disease prevention [MDB+17] Device technologies are now commonly utilized in a variety of medical and healthcare settings. For example, cardiac pacemakers help arrhythmia patients and digital hearing aids provide relief to the hearing impaired. Sequencing technologies are used to perform genotyping and screen for genetic predisposition to certain diseases. Database technologies such as digitalized health information systems facilitate health care workers to better serve patients and health care researchers to make important scientific discoveries.

Among wearable digital devices, the most relevant to this dissertation are wearable biosensors and online mental assessment tools, which collect data directly from patient interaction with tablets or smart phones. These devices can monitor physiological health and psychological health, respectively [Lup13, BSB19]. Wearable physiological biosensors can capture a wide range of physiological data, including activities (measured with accelerometer), body temperature, blood pressure, blood volume pulse, electrocardiogram(ECG), electrodermal activity, etc. Likewise, mental states like stress, and cognitive functions, like response time, can be regularly and conveniently sensed by user-engagement with applications on portable devices. Mental and cognitive states have been found to be associates with physical health and vice-versa [Gaz00, CWH+05, VLM+11]. Thus a combination of wearable physiological biosensors and portable mental and cognitive state sensors allow for real-time monitoring of health conditions without the need for intrusive methods or hospitalization. This could lead to better patient compliance with treatments or prevention protocols. As previously indicated, poor lifestyle choices have long been recognized as risk factors for ARVI. It is a theoretically plausible conjecture that if we pick the right combinations of sensors with appropriate feature extraction, we may infer knowledge on lifestyle. This might permit us to go one step further to forecast disease vulnerability on an individual level.

1.1.5 Background: Circadian rhythm

Periodic phenomenon are observed to be ubiquitous at many different levels and scales of biological activity. A prominent example is the set of cyclic variations in animal and human behaviors, including sleep–wake cycles and feeding. Physiology parameters like body temperature blood pressure also fluctuate with a period of around 24 hours. With the advancement of molecular biology, concentrations of some molecules were also found to be cyclic, including hormone secretion and genes expression [SLL+02, ZLB+14]. A phenomenon that exhibits periodicity with period $\sim$24h is siad to have circadian rhythm (stemming from Latin *circa diem*, about a day). Circadian rhythms manifests across behavior, physiology, metabolism, and gene expression. However, the circadian cycling phases may be different depending on lifestyle, sleep habits, and body tissue type

that is assayed.

How is circadian rhythm organized in so many different aspects and tissues? The physiological structural basis is a hierarchical network of central and peripheral oscillators. The master pacemaker locates at ∼15,000-cell suprachiasmatic nucleus (SCN) [GR10, EMSC09], a region of the brain in the hypothalamus. Circadian rhythm in peripheral tissues is subject to orchestration by the central clock through endocrine and neural output [AS05, SGS+15]. It is worth mentioning that a central pacemaker of SCN is not necessary for peripheral tissues to maintain circadian rhythm. Many cultured *in vitro* peripheral tissue experiments have exhibited comparable self-sustained circadian oscillations in the absence of the central pacemaker SCN [YYL+04]. The relationship between central pacemaker and peripheral clock represents two characteristics of circadian rhythm: self-sustained and entrainable. The peripheral clock is self-sustained while it is under orchestration of the central pacemaker SCN. With entrainment, circadian clock is able to adjust phases to adapt to external environmental cues, including light/dark cycles, food intake, stress, and exercise.

The underlying molecular mechanism of the self-sustained circadian clock is based on a complicated network of interlocked transcriptional-translational feedback loops (TTFLs) [THKM08]. The core components of mammalian circadian clock machinery are the clock proteins. An illustraction of TTFLs is shown in Fig. 1.2. This figure was published in [THKM08]. Besides TTFL, other mechanisms, like redox states and post-translation modification of proteins, regulate circadian rhythm as well [EGZ+12, GV07].

Interrupted circadian rhythm has been proven to increase risk for many diseases. Disrupted rhythm can also lead to health hazards, as illustrated by the fact that shift workers are more vulnerable for diseases including cardiovascular disease, stroke, type 2 diabetes, and cancer [SLS+01, SHMS09, SBG+07, PSSH11, BFS+09] .

On the other hand, some diseases also affect the circadian clock. It is known that Alzheimer's disease, cancer, and severe sepsis cause abnormal circadian rhythm in mammals [AOCC15, MBCG05, CC08, SCVO13, MVDC+10, MDKK+02, PB08, TYH+01, OYD+08]. Traditionally circadian rhythm was expected to be totally lost during severe acute infection. Recently a study challenged the traditional idea with LPS-induce endotoxemic mice [HCS+14]. They found that despite most of baseline cyclers ceased to be periodic after infection, more genes astonishingly gained periodicity during endotoxemia. Functional annotation analysis was performed to discover the enrichment of immune-related processes in the novel cyclers in endotoxemia. A follow on *in vitro* research experiment provided additional supporting that a unique set of novel cyclers were gained in mice with tissue-specific disruption of *BMAL1* [ZHW+19], a clock gene. This set has almost the same size compared to the number of cyclers only found in $BMAL1^{flox/flox}$ mice.

1.2 Thesis Outline and Contributions

In Chapter 2, a bio-longitudinal viral challenge study was performed on a human cohort during which different types of biomarkers were collected for analysis. We investigated the temporal evolution of these biomarkers and evaluated their periodicity in basal state and after inoculation. This is the first systematic study on the interplay between circadian rhythm and immune response in a human *in vivo* experiment. While a large number of novel cyclers were identified after inoculation, this observation, together with two earlier studies [HCS+14, ZHW+19], suggests that circadian components could have quite different composition when the host is perturbed from a normal condition, e.g., by introduction of a viral pathogen. We make several striking discoveries in the context of our analysis of the Biochronicity human viral challenge study. First, baseline cycling genes selectively undergo cycle extinguishment depending on their phase angles; those that attain minimum cycle amplitude nearer the time of inoculation are more likely to lose periodicity. Second, baseline cyclers are much stronger in low shedders than in high shedders, which suggests that chronotype strength could potentially be used as a predictor of susceptibility. A transcription regulation database was incorporated to map the genes involved in immune response and circadian rhythm with their upstream transcription factors. Key parts of JAK-STAT pathway and NF-κB and its subunit RELA are at the center of the immune response module and the circadian module, respectively. We also find that these two sets of master regulators are buffered by other genes discovered in our analysis. These buffering genes might explain the resilience of the clock to immune response and the emergence of circadian rhythm in certain genes after inoculation.

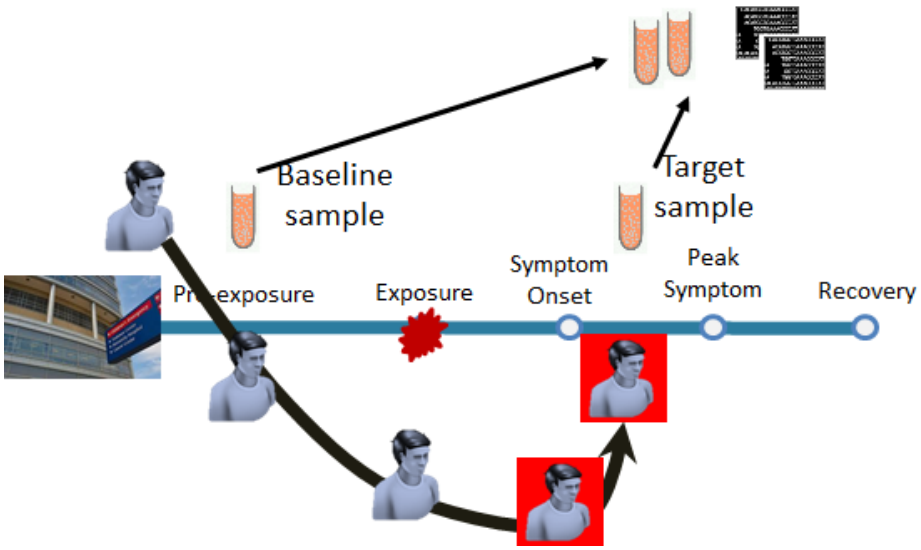
In Chapter 3, an adaptive transfer learning algorithm was introduced to address the data covariate shift problem in wearable data. Using a multivariate hidden Markov model (HMM) and Fisher's linear discriminant analysis (FLDA) the algorithm adaptively adjusts to shifts in distribution over time. The proposed algorithm is unsupervised and learns to label events without requiring *a priori* information about true event states. Simulations validate the effectiveness of the proposed adaptive algorithm, which significantly outperforms other event classification methods. When applied to early time points in the HVC data, the algorithm extracts sleep/wake features that are predictive of both infection and infection onset time.

In Chapter 4, a score, named cognitive performance variability (CPV), is developed to represent the variability of an individual's performance on a repeated sequence of cognitive tests during pre-exposure baseline, which is found to be highly associated with future viral shedding. CPV is robust to sampling and sessions. Our results suggest that pre-exposure cognitive function variability is closely related to viral shedding, opening the intriguing possibility that periodic cognitive testing could be useful in predicting susceptibility to infection.

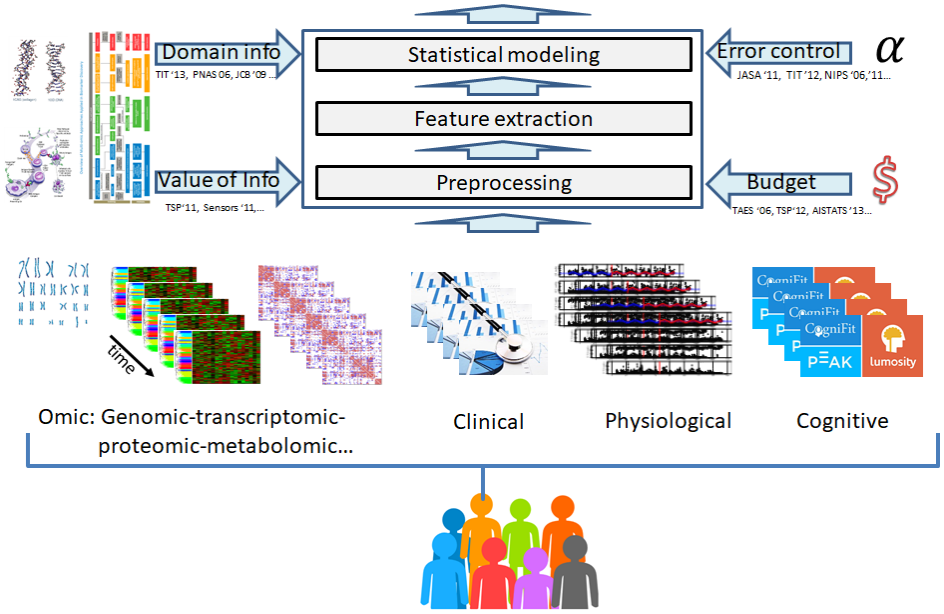
1.3 List of relevant publications/preprints

This dissertation research leads to several publications that are published, or in preparation:

- She, Xichen, Yaya Zhai, Ricardo Henao, Christopher Woods, Christopher Chiu, Geoffrey S. Ginsburg, Peter XK Song, and Alfred O. Hero. Adaptive multi-channel event segmentation and feature extraction for monitoring health outcomes. *IEEE Transactions on Biomedical Engineering*, 2020. (co-first author)
- Yaya Zhai, Ricardo Henao, Christopher Woods, Geoffrey S. Ginsburg, and Alfred O. Hero. Rhinovirus infection and the rhythms of health: a multispectral challenge study in a human cohort. (in preparation)
- Yaya Zhai, P. Murali Doraiswamy, Christopher W. Woods, Ronald B. Turner, Geoffrey S. Ginsburg, Alfred O. Hero. Cognitive performance variability is associated with severity of respiratory infection. (in preparation)



(a) Illustration of the experimental process of Human Viral Challenge (HVC) experiment. After a group of volunteers has been selected and pre-screened, they are monitored before and after being exposed to the virus. Related biological samples or other signal, information responses will be gathered throughout time. Participants will have various levels of symptoms after being inoculated, and they will eventually recover.



(b) Illustration of data processing of Human Viral Challenge (HVC) experiment. Pre-processing of diverse forms of data acquired in the experiment with various origins will be done first, and then features will be extracted, usually using domain expertise. Statistical learning approaches would be employed to answer the issue that the HVC research is meant to address.

Figure 1.1: Illustration of experimental process and data processing of Human Viral Challenge (HVC) experiment.

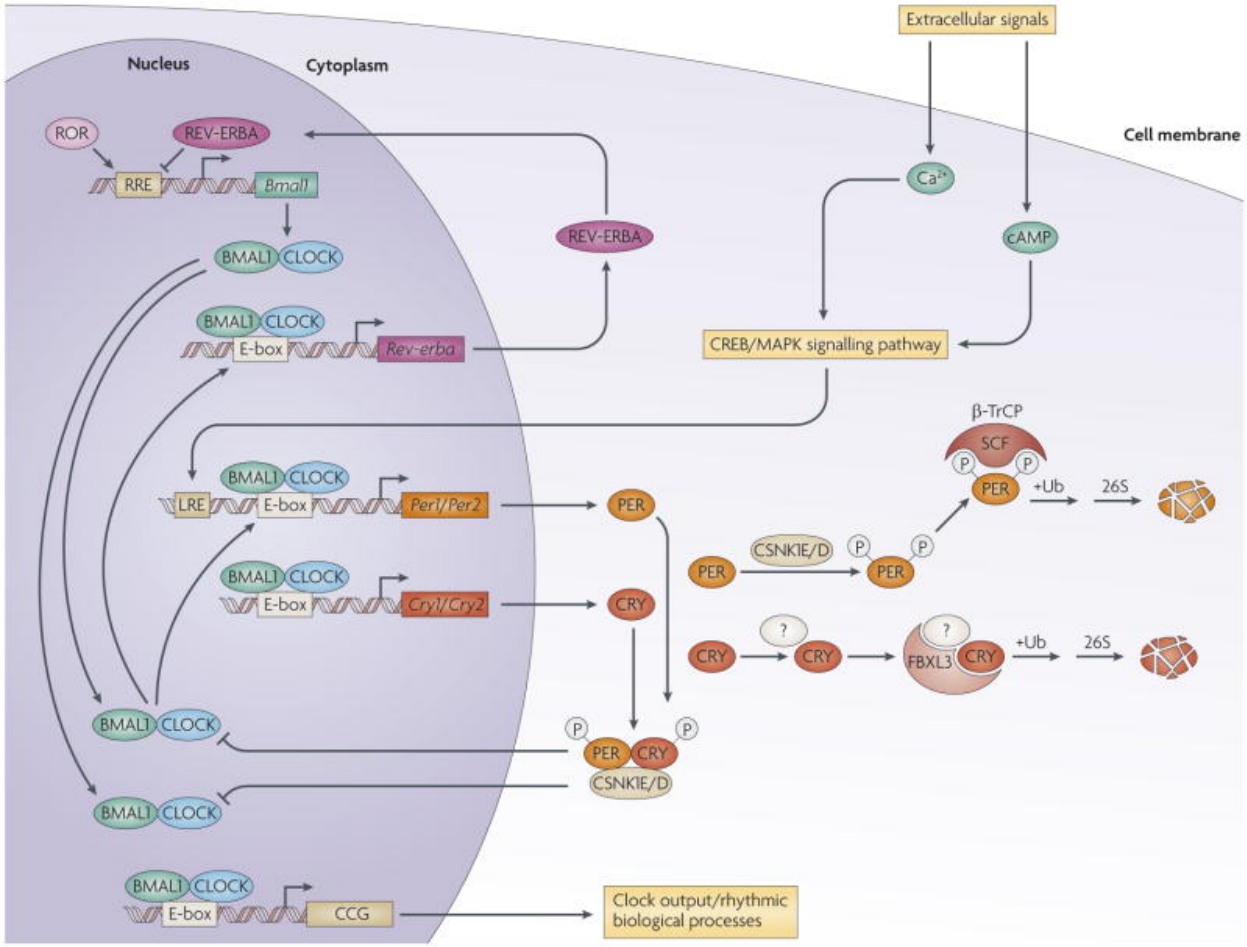


Figure 1.2: Illustration of transcriptional-translational feedback loops (TTFL) as the mammalian circadian clock machinery. This figure is cited from [THKM08].

CHAPTER 2

Rhinovirus Infection and the Rhythms of Health: A Multispectral Challenge Study in a Human Cohort

2.1 Introduction

The rhythms of health are manifested across behavioral, physiological, metabolic, and genomic domains of the human condition. For a healthy individual, sleep, mental state, blood pressure, heart rate, and molecular (gene, protein, metabolite) expression express in a steady (circadian) pattern within a 24 hour cycle and from day to day, but are subject to modulation by external environmental stimuli. Mechanisms underlying circadian rhythms and their contribution to the health of individuals have been elucidated in [THKM08]. Ways that these rhythms become perturbed due to exposure to stressors in the environment have been studied in [ALML+14]. Links between circadian rhythms and optimal timing of drug treatments to achieve maximal effectiveness, i.e., circadian chemotherapy and other chronotherapies have also been of interest [Hru85].

This chapter studies the sensitivity of circadian rhythms to an acute health perturbation: a one-time exposure to an infection-causing pathogen. Specifically, a challenge study was designed and a cohort of human volunteers was recruited to participate in an ambitious longitudinal study involving monitoring over 19,000 biomarkers from 5 different biomarker types, including clinical, cognitive, physiological, metabolic, and genomic modalities measured in blood, nasal fluids, human-machine-interactions, and activity logged by wearable devices. Importantly, this challenge study was bio-longitudinal – multiday time samples of biomarkers were collected both and after inoculation. These samples were collected from the cohort over an 8 day period, collected at least 3 times per day, and an inoculation by live virus (human rhinovirus, HRV) was administered on the fourth day, perturbing the healthy baseline of each subject in the study. This challenge study offers a unique glimpse into the exposome over time with a rich multimodality view across multiple scales. This is the first systematic study on the interplay between circadian rhythm and immune response in an *in vivo* human experiment.

Our analysis reveals novel exposome chronotypes associated with the effect of the viral inoculation, manifested across all biomarker types collected in the challenge study. Many circadian

biomarkers largely remain resilient to the viral insult. Biomarkers that are strongly circadian throughout the baseline and post-inoculation time periods include: cortisol (a steroid associated with stress [HWK09]), DDIT4 (a gene identified as a monitor for human circadian rhythm in blood transcriptome [LMLP$^+$17, Hug17, LCJ$^+$21]), and heart rate (modulated by the wake/sleep cycle). We find that among all cycling biomarkers, it is the circadian rhythm of gene expression that is by far the most perturbed by exposure. We observe that the perturbation by inoculation results in quenching (extinguishment) of the circadian cycling of certain baseline cycling genes while it results in recruitment circadian cycling of some baseline non-cycling genes. Among biomarkers that are recruited is the poly(A) RNA polymerase D7 (PAPD7), a gene that adds poly(A) to various classes of nuclear RNAs and has been found to be essential to Hepatitis B Virus RNA stabilization [OCH13, ZYL$^+$19]. As a large number of novel cyclers were identified after inoculation, this observation, together with two earlier studies [HCS$^+$14, ZHW$^+$19], suggests that circadian components could have quite different compositions when the host is perturbed from a normal condition, e.g., by introduction of a viral pathogen.

Our analysis also reveals that the quenching of gene cycling is phase-selective. Specifically, among genes of all cycling phases, cycling genes with phase minima around inoculation time (8am) are the most likely to lose their periodicity after inoculation. Conversely, cycling genes that have phases achieving their maximum expression level at inoculation time, are most likely to retain their periodicity. Circadian cycling of metabolites is also most resilient at inoculation time where the metabolites are at their maximum concentration, but the phase-selective quenching effect is not as strong. Such phase-selective disruption of circadian cycling may have significant preventative health implications.

Last but not least, our analysis reveals that those individuals who are not severely infected, as measured by viral shedding and symptom, exhibit considerably stronger circadian cyclers both in the basal condition and after inoculation then the rest of the cohort. This suggests that a lack of strong biochronicity might be a predictor of susceptibility. In particular, if these findings hold up in broader cohorts, the strength of a certain circadian cycling genes could be a possible predictor of infectiousness, safeguarding susceptible people and aiding in infectious disease prevention and treatment.

2.2 Results

2.2.1 Human viral challenge study

A longitudinal human rhinovirus (HRV) challenge study was designed at Duke University and the University of Virginia and took place in 2015 over a nine day period. A total of 24 volunteers were recruited into the study and 19 participants completed the study. One of these participants had a

failed inoculation and was omitted from our analysis. The age range of the remaining 18 participants was between 18 and 23, two thirds of these participants were male, and 4 were non-caucasian. For a more detailed demographic summary see participant demographics are shown in Fig. 2.7 in the Supplementary Materials. The study was reviewed and approved by the Institutional Review Boards at Duke University and at University of Virginia. Written informed consent was obtained from all participants. Exclusion criteria included pregnancy, chronic respiratory illness, high blood pressure, tobacco/drug/alcohol history, and serum antibody levels above 1:4 titers. The participants were not isolated during the study.

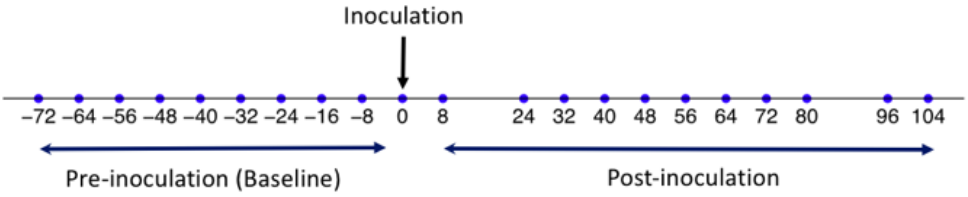
Fig. 2.1A summarizes the data collection protocol. On day four at 8am (denoted as time 0 on Fig. 2.1A) each participant was inoculated via intranasal drops of diluted Human Rhinovirus strain type 16 with a dose of 100TCID50 in 1mL Lactated Ringer's Solution. Between 7am and 9am on the morning of the fourth day of the study participants were inoculated with live HRV virus. A histogram of the inoculation times is shown in Figure 2.8 in the Supplementary Materials. Molecular, physiological, cognitive, clinical, and shedding data were collected at multiple time points, annotated in time units (hours) on Fig. 2.1A.

The biomarker type, subtype and sampling rates of the data collection are indicated in Fig. 2.1B. Blood samples were collected three times per day from which molecular data was extracted using RNAseq (genes), p180 (metabolites), and steroid assay (hormones). Physiological signals, including physical activity (ACC), blood volume pulse (BVP), electrodermal activity (EDA) and skin temperature (TEMP), were recorded continuously over time from wearable wristbands (Empatica E4 devices). Cognitive ability, measured by 18 Lumos scores, and 8 self-reported symptom subtypes were measured three times daily based on subject inputs to a tablet device. Finally viral shedding, quantified by PCR viral titrations, was collected once per day from nasal washes.

Study participants reacted differently to the viral inoculation, with some becoming severely symptomatically ill while others remaining largely symptom free. All participants had some level of viral shedding; the total accumulated shedding over the post-inoculation periods is shown in Fig. 2.1C. The entire population was divided into two sub-groups (low shedders and high shedders) depending on whether their shedding failed to reach or exceeded the population median of 5.7 titers (vertical red line in the figure). In the sequel we will show that these two sub-groups have distinct chronotypes in terms of their circadian cycling and non-cycling during baseline and post-inoculation (Fig. 2.5). Shedding and symptom evolution over time for each subject is included in the Supplementary Material (Fig. 2.9). There is a good linear correlation between total shedding and total symptom (Fig. 2.10). Accumulated symptoms scores of all participants are shown in Fig. 2.11. The population are ordered by increasing total shedding levels from left to right. All high shedders have more severe symptom than low shedders except for participant 19, indicating the good coherence between shedding and symptom. We use total shedding in the following analysis to

represent infectiousness for it is more objective compared to self-reported symptom score.

A



B

Biomarker type	Subtype	Sampling rate
Molecular (Blood draw)	RNAseq, p180, hormones	3 times/day
Physiology (E4 wearable)	ACC, BVP, EDA, TEMP signals	32Hz, 64Hz, 4Hz, 4Hz
Cognitive testing (Lumos)	18 Lumos scores	3 times/day
Clinical symptoms (Diaries)	8 symptom subtypes	3 times/day
Shedding (Nasal wash)	PCR viral titrations	1 time/day (post-inoc only)

C

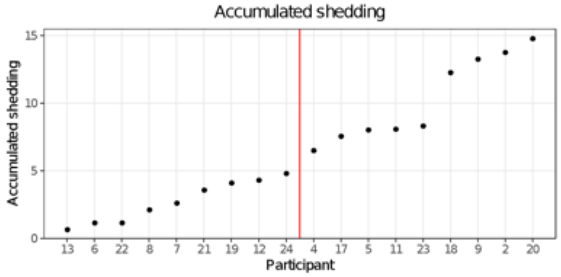


Figure 2.1: Challenge study protocol, data collection and clinical phenotype. A. Timeline of study indicating the 3-day pre-inoculation period and the 5-day post-inoculation period along with times of blood samples shown as bullets on timeline. HRV viral inoculation was administered on day 4 at 8AM (0 hours on timeline) immediately after sample collection. B. 5 diverse biomarker types were collected over the course of 8-day study from 18 human volunteers. C. By computing the median accumulated shedding level (indicated by the red vertical line), the entire population can be divided into two sub-groups (low shedders and high shedders) based on their accumulated shedding levels in response to the viral challenge. These two sub-groups were found to have very distinct types of circadian patterns in both baseline and post-inoculation (see Fig. 2.5).

2.2.2 Distinct biomarker chronotypes over baseline and after viral exposure

We evaluated the circadian cycling of each biomarker by applying a statistical periodicity test (JTK_CYCLE [HHK10]) at various levels of significance for the basal state (-72 to 0 hours) and after viral exposure (0 to 104 hours), respectively. Consistent with previous studies [LMLP+17], we found that many protein coding genes, metabolomic molecules, and physiological signals display strong circadian behaviors in baseline. The chronotypes of some baseline cycling biomarkers are resilient to the perturbation caused by pathogen exposure while others have disrupted chronotypes as evidenced by cycling phase shift, cycling attenuation, or cycling extinguishment. Interestingly,

some additional biomarkers were found whose circadian cycling appears to be recruited only after inoculation.

2.2.2.1 The circadian landscape of different biomarkers in baseline and after exposure

The cycling of biomarkers was distinguished on the basis of four different strength levels (Strong, Medium, Weak and Non-cycling), determined by the level of significance of the JTK_CYCLE cycling test (see Supplementary Material 2.6 for a precise definition and illustrations, e.g. Fig. 2.12). A biomarker is cycling if it is a strong, medium or weak cycler and otherwise it is said to be non-cycling.

Figures 2.2A & 2.2B show the proportion and number of circadian cycling biomarkers at different cycling strength levels in baseline (BL) and post-inoculation (PI), respectively. Among all biomarker types, Lumos cognitive biomarkers had the least amount of circadian cycling, exhibiting no cycling in either BL or PI according to the JTK_CYCLE significance test at level 0.05. Participants' cognitive performance improved over time as they became more familiar with the games. That is the reason why Lumos variables has no cycling property. The cognitive biomarkers will not be discussed further.

Among the other biomarker types, gene expression (RNAseq) had the most cyclers at all cycling strength levels (Top panel of Fig. 2.2A) in both baseline and post-inoculation time intervals, but it had the lowest proportion of cyclers relative to the total number of RNAseq biomarkers (Top of Fig. 2.2B). 45 (0.23%) genes are strongly periodic in baseline, while 364 (1.86%) genes have medium periodicity strength and another 1,475 (7.55%) genes are weakly periodic in baseline. In post-inoculation, there are only 11 (0.06%) strong gene cyclers, 109 (0.56%) medium gene cyclers but 2,374 (12.14%) weak gene cyclers. The most enriched KEGG pathway of significant baseline cycles with any strength is Chemokine signaling pathway, while the post-inoculation novel cyclers has enrichment in Protein processing in endoplasmic reticulum. Metabolites measured by p180 assay (Second from top panel of Figs. 2.2A and 2.2B) have 7 (3.85%) strong cyclers, 24 (13.19%) medium cyclers and 64 (35.16%) weak cyclers in baseline, and the numbers go down to 2 (1.1%) strong cyclers, 20 (10.99%) medium cyclers and 62 (34.07%) weak cyclers after viral exposure. Metabolites including mino acid, urea and different kinds of lipids were found to be rhythmic in a previous research. [EMSC13].

Physiological features (second from bottom panel of Figs. 2.2A and 2.2B) were extracted from E4 wristband measurements using the methods described in [SZH+20] (Supplementary Materials). These physiological biomarkers exhibit the highest proportion of strong cyclers as compared to genes and metabolites. In baseline, all features related to heart rate (HR), activity (ACC) and body temperature (TEMP) are strongly cycling while in EDA, there are no strong cyclers (EDA_sd is medium cycling). It is not surprising that heart rate and body temperature fluctuate around a 24-hour

period in tandem with the sleep-wake cycle [MMW$^+$00, BWV$^+$01]. However, the rhythmicity of ACC_median degrades from strong to medium cycling and TEMP_median and TEMP_mean are transformed to weak cyclers after viral exposure. Interestingly, while non-cyling in BL, in post-inoculation EDA_median and EDA_mean emerge as medium strength cyclers in PI. Despite circadian variations are not neglectable [KKB$^+$18], electrodermal activity (EDA) is more regulated by the activation of the autonomic nervous system, when sweat gland stimulation increases the electrical conductivity of the skin[ZFS91]. It is also related with stressful events [SAS$^+$09].

Among all biomarkers, the cycling behavior of hormones (bottom panel of Figs. 2.2A and 2.2B) is most resilient to the viral perturbation; there is no change in their cycling strength after viral exposure. Two out of the three hormones measured in this study, melatonin and cortisol, have previously been reported as markers of circadian rhythm [JCB07]. These hormones retain their strong periodicity from BL to PI. Another hormone, dehydroepiandrosterone (DHEAS), is weak cycling both in basal state and after exposure. (See Supplementary Material Fig. 2.18, Fig. 2.19 and Fig. 2.20 for their concentration over time.)

2.2.2.2 Impact of perturbation on circadian cycling

In addition to the general circadian landscape of the sets of biomarkers that cycle in baseline and post-inoculation, shown in Figs. 2.2A and 2.2B, the fate of individual BL cyclers is also of interest. Figures 2.2C and 2.2D show that a cycler's ability to maintain its cycling after inoculation depends on its strength of cycling at baseline. If a baseline cycler become non-cycling after inoculation, it is called a BL-only chronotype. Conversely, a biomarker is a PI-only chronotype if it is non-cycling in baseline and acquires significant periodicity after exposure. Another type of cycler are the resilient chronotype who maintain their rhythm in both BL and PI. The number of these three chronotypes are shown in Fig. 2.2C for different cycling strength levels (JTK_CYCLE false discovery rate, FDR). Proportionately to all baseline cyclers (including both the BL-only and resilient chronotypes) we show the percent of each chronotype in Fig. 2.2D. The rightmost bar in Figs. 2.2C and 2.2D indicates cyclers at any strength level (FDR < 0.05) and demonstrates that approximately 30% (611 of the 1,992) baseline-cycling biomarkers are resilient chronotypes, i.e., they continue cycling after inoculation. Furthermore, there are more biomarkers that only cycle in post-inoculation (1,982, 99.5%) than those that only cycle in baseline (1,381, 69.3%). However, in both cases most of post-inoculation chronotypes are weak cyclers, i.e., they have JTK_Cycle FDR between 0.001 and 0.05.

The leftmost bar of the barplot in Figs. 2.2C and 2.2D shows only the strongest cyclers (JTK_Cycle FDR < 0.000001), which only includes just 67 biomarkers. These strong baseline and post-inoculation cycling biomarkers are listed along with their FDR and cycling phases in the Supplementary materials and include biomarkers of every type except for Lumos, which, as pointed

out above, does not exhibit cycling at any strength level (JTK_CYCLE FDR < 0.05). Approximately 27% (17) of the 63 strong BL-cycling biomarkers continue to cycle after inoculation. Remarkably, there are many more genes with the BL-only chronotype than with the PI-only chronotype.

From weaker to stronger cyclers (left to right of Figs. 2.2C and 2.2D) we observe that the proportion of genes with resilient chronotype, starts to decrease from 27% to $\sim$10% as the FDR cutoff loosens from 0.000001 to 0.001, then increases to $\sim$30% as the FDR cutoff increases to 0.05. The proportion of genes with PI-only chronotype has a nearly monotonic increasing trend as the FDR cutoff increases, and as the number of all baseline cyclers also increases. The perturbations of cycling strength due to inoculation will be further discussed in Sec. 2.2.5.1.

2.2.2.3 Impact of perturbation on biomarker expression levels.

In addition to having an effect on circadian cycling, the viral inoculation has an effect on the average measured levels of the biomarkers. For the RNAseq, 467 (2.4%) genes were up-regulated and 2,106 (10.8%) genes were down-regulated after inoculation when compared to baseline. A pathway enrichment analysis (See 2.4 in Methods for details) shows that the most enriched pathways for the up-regulated genes are NOD-like receptor signaling (FDR 7.86e-12) and Influenza A (FDR 7.86e-12), while the most enriched pathways for the down-regulated genes are Spliceosome (FDR 5.68e-23), Ribosome(FDR 2.70e-13), RNA transport (FDR 2.70e-13) and protein processing in endoplasmic reticulum (FDR 4.11e-11).

As contrasted to the genes, none of the metabolites' concentrations rose significantly after inoculation. However, 15 (8.2%) metabolite concentrations dropped in value with FDR < 0.05. The mean levels of hormones and physiological features do not vary significantly between baseline and post-exposure.

In summary, we found that genes have the largest number of cyclers and their rhythmicity is the most sensitive to circadian disruption by viral exposure compared to other biomarker types. Thus in the following subsections (Secs. 2.2.3-2.2.4) we focus on highlighting the cycling behaviors of genes in baseline and how they change after inoculation. Similar analysis on other biomarker types are included in the Supplementary Material.

2.2.3 Circadian genes in basal state

As described in 2.2.2.1, there are 45 STRONG cyclers, 364 MEDIUM cyclers and 1,475 WEAK cyclers in baseline (basal state) before perturbation. These sum up to a total of 9.6% of the entire transcriptome. A KEGG pathway enrichment analysis (See 2.4 in Methods for details) was performed and 108 out of 329 pathways tested are significantly enriched in the set of baseline cycling genes, among which the top ranking pathways are Chemokine signaling pathway(FDR

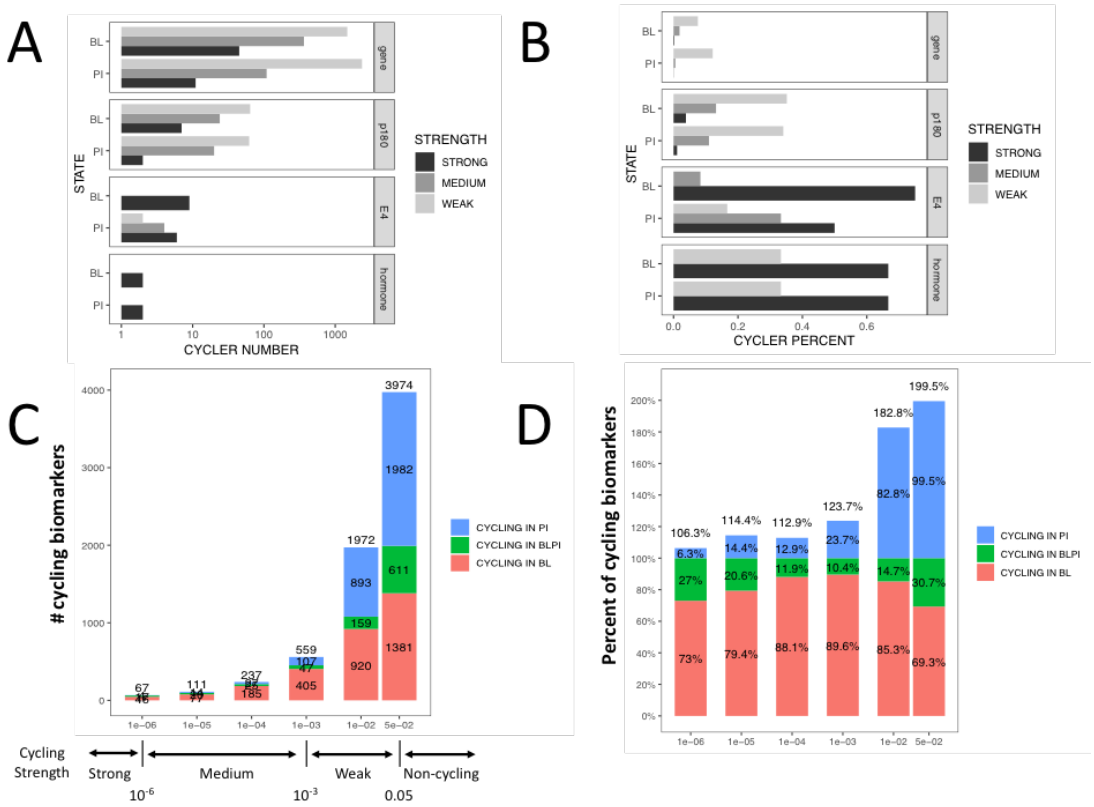


Figure 2.2: Cycling profiles of various biomarkers in baseline and post-inoculation. Panels A&B: Count and proportion of cycling biomarkers with different strength in baseline (BL) and post-inoculation (PI). (Strong cycler: FDR $<$ 1e-6; Medium cycler: 1e-6 $<$ FDR $<$ 1e-3; Weak cycler: 1e-3 $<$ FDR $<$ 0.05). The genes have the lowest proportion but the largest number of cyclers, and they are most susceptible to circadian disturbance due to viral exposure. Most physiological features by E4 are strongly periodic. Hormones (Cortisol, Melatonin, DHEAS) are the most resilient to the perturbation. Panel C: Total number of cycling biomarkers at different strengths having unique cycling features (only cycling in baseline; cycling in both baseline and post-inoculation; only cycling in post-inoculation). Panel D: Proportional biomarker cycling corresponding to Panel C. Two types of BL cyclers add up to 100%. As the cycling strength threshold weakens from left to right, the proportion of BL cyclers that maintain their periodicity (green bar segment) first starts decreasing from FDR 0.000001 to 0.001, then increases as FDR cutoff goes to 0.05. The PI-only cyclers (blue bar segment) increases both in number and in proportion as FDR cycling threshold is loosened from strong to weak.

7.30E-13), Pathways in cancer(FDR 1.64E-12), Cytokine-cytokine receptor interaction (FDR 1.18E-08), Natural killer cell mediated cytotoxicity (FDR 7.60E-08). The functions and pathways enriched by baseline cyclers with different phases are depicted in 2.3A.

The phase and strength distribution of cyclers in the basal state is shown in 2.3B. Phase 20h corresponds more than half (51%) of all baseline cyclers, and this phase accounts for even higher composition in STRONG and MEDIUM cyclers (69% and 67%, respectively), compared to WEAK

cyclers (46%). Other than Phase 20, Phase 12h has more strong cyclers than the rest, including DDIT4, the gene with most significant periodicity in both baseline and post-inoculation. (See Supplementary Material 2.15 for a plot of DDIT4 over time.)

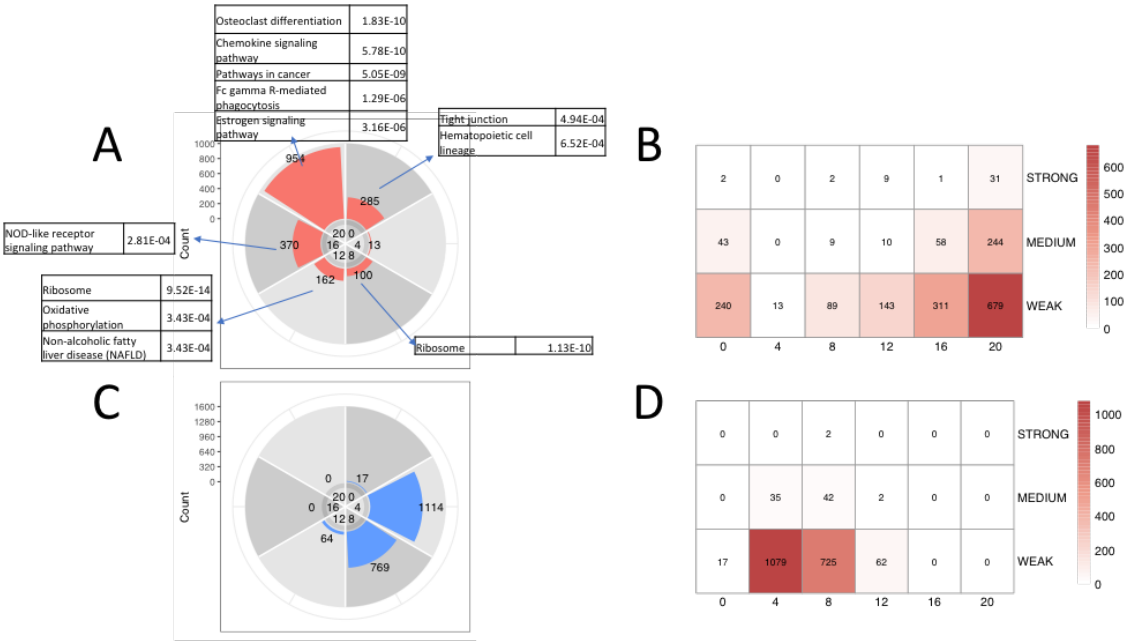


Figure 2.3: Phase and strength distribution of baseline (BL) and post-inoculation (PI) cyclers, and corresponding biological function annotations. Panel A: the phase distribution of baseline cyclers and correspondingly enriched KEGG pathways of baseline cyclers. Panel B: the co-distribution of phase (in hours as estimated by JTK_Cycle) and strength of baseline cyclers. Panel C: the phase distribution of novel recruited cyclers after exposure. Panel D: the co-distribution of phase and strength of novel recruited cyclers after viral exposure.

2.2.4 Novel gene cyclers are recruited after inoculation

The number of novel recruited cyclers after viral exposure is similar to the number of cyclers in baseline ($1,964$ compared to $1,884$). In Fig. 2.3C and Fig. 2.3D we show the phase and strength distribution of the novel cyclers. Most of these novel cyclers have Phase 4 or Phase 8, which is just the opposite of baseline phases. There are only 2 STRONG cyclers, CAT and TLR2. (See Supplementary Material Fig. 2.16 and Fig. 2.17 for their expression over time.)

Only one pathway, Protein processing in endoplasmic reticulum, is enriched by novel cyclers. 54 out of 129 members in this pathway are novel cyclers only found in post-exposure. This pathway is not enriched by any baseline cyclers.

Compared to baseline cyclers, these novel cyclers emerged only after inoculation have a smaller ratio in STRONG and MEDIUM cyclers (4.1% in total cyclers in contrast to 21.7% in all cyclers in baseline).

2.2.5 Influence of inoculation on the circadian transcriptome

2.2.5.1 Changes in cycling strength and phase of resilient baseline cyclers

After the disturbance induced by viral exposure, the baseline circadian cycling of some genes persists (resilient chronotype) while the cycling is quenched for others and fails to persist. Taking into account the different cycling strength of cyclers, 530 baseline cyclers (28%) successfully survived the disruption. However, most of them undergo a change in phase relative to baseline. The effect of viral exposure on cycling strength and phase is shown in Fig. 2.4. We observe, by comparing Figs. 2.4A and 2.4B, that the distribution of cycling strengths of resilient genes does not change much; 83% of cyclers still maintain the same cycling strength category (strong, medium, weak) after inoculation. In contrast the distribution of cycling phases of resilient genes does shift significantly; from Phase 0 and Phase 20 in baseline, to Phase 4 and Phase 8 after inoculation (Fig. 2.4B).

2.2.5.2 Disruption of baseline-cycling genes is phase-selective

The proportion of baseline cycling genes that retain periodicity for different cycling strength thresholds depends strongly on the baseline phase. This result is shown in Fig. 2.4C with 6 different strength thresholds ranging from p-value$<$0.05 (weakest) to p-value$<$0.000001 (strongest) computed by JTK_CYCLE. Up to a p-value cutoff at 0.001, there is a consistent phase-selective quenching pattern of baseline cycling genes. Figure 2.4D redraws this proportion of resilient cycling genes at p-value cutoff 0.001 in a way that more explicitly reveals the phase-selective quenching phenomenon. The figure shows that the proportion of resilient cycling genes is minimized at a phase of 20hrs, corresponding to cycling genes with minimal circadian expression levels occurring at 8am, near the time of inoculation, and that this resiliency steadily increases as the phase of a cycling gene approaches 8hrs where the maximum gene expression occurs at 8am.

This finding indicates that a cycling gene presenting its maximal expression level near inoculation time (8am) has higher likelihood of retaining its periodicity. Conversely, genes presenting minimal expression levels near inoculation time have higher likelihood of undergoing circadian extinction. In other words, the phase of a cycling gene at the time of insult determines its circadian resilience.

2.2.6 A biochronicity genotype that differentiates between low and high shedders.

The cohort was divided into two balanced groups of low and high shedders, as explained in Sec. 2.2.1 (see Fig. 2.1C). Interestingly, there are no significant differentially expressed genes in baseline between the two sub-cohort groups according to t-test at 0.05 FDR. In contrast, there is an astonishing difference in their baseline cycling strength. The top 20 genes having largest

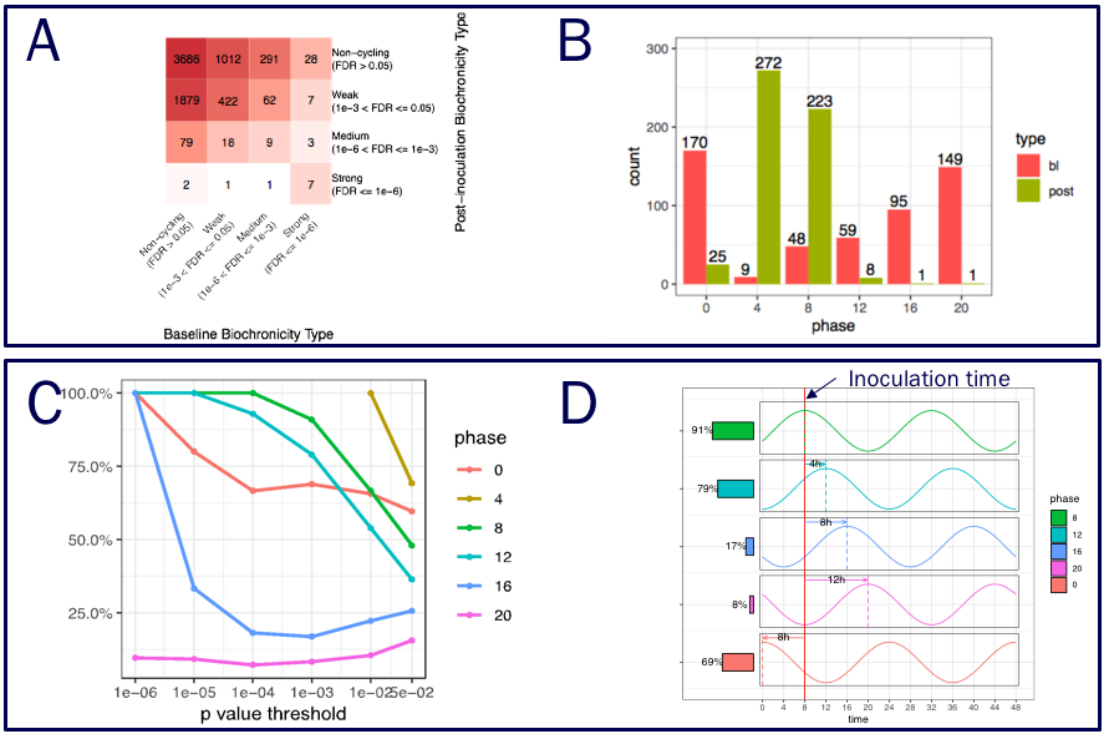


Figure 2.4: Effect of inoculation on baseline circadian rhythm. Panels A and B show the changes in strength and phase of resilient cyclers remaining rhythmic after inoculation. Panels C and D show the phase-selective periodicity quenching phenomenon.) Panel A: Changes in cycling strength from baseline to post-inoculation. Stronger cycling genes are more resilient to viral exposure. Panel B: Phase distribution of periodic genes in baseline and post-inoculation. While approximately 60% of cyclers have phase of 20h or 0h in baseline, only 93% of these cyclers shift their phase (to 4h and 8h) after inoculation. Panel C: Proportions of BL cyclers retaining periodicity at different phases (colored curves) and at different p-value thresholds. Panel D. For a p-value cutoff of 0.001 the proportions (colored bars at left) of baseline cyclers that retain their periodicity after inoculation are strongly dependent their phases. Cyclers whose phase maximum occurs near the time of inoculation (8h) are much more likely to be resilient to inoculation and this resilience decreases as the phase maximum shifts to other times of day.

baseline cycling strength difference were identified by the ratio in their JTK_CYCLE p-values in sub-cohorts. The cycling p-value in high shedders of the 20th gene is more than $7,000$ times that in low shedders. With the exception of phase 4h, we notes that they cover the full range of phases $(0, 4, 8, 12, 16, 20)$h reported by JTK_CYCLE. Phase 4 is missing from these 20 genes due to the fact that there are very few baseline cyclers at phase 4, as reported in Sec. 2.2.3. Gene profiles over time are shown for low shedders in Fig. 2.5A and for high shedders in Fig. 2.5B. In each subfigure, the strip at left indicates the cycling status of the gene in BL and PI (red color denotes cycling), and differential expression of the gene BL to PI (red color denotes differential expression). The heatmap show the expression levels of the entire time course of these genes. For each gene and each

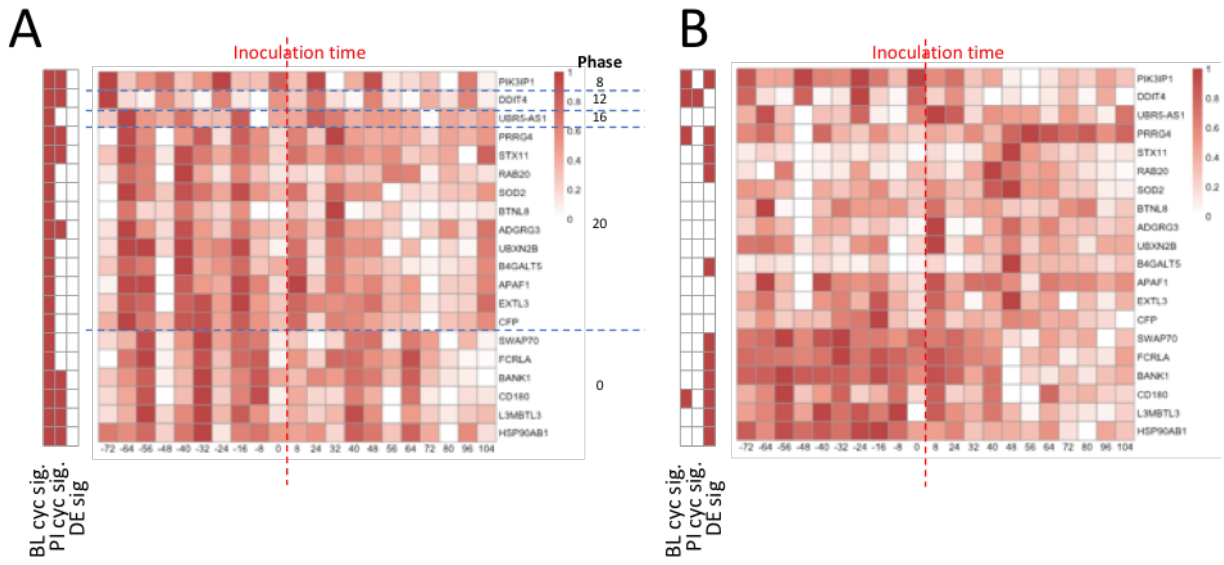


Figure 2.5: Heatmaps of a group of 20 BL-cycling genes, ordered by increasing phase, that best discriminate the low shedders (A) from the high shedders (B) in the challenge study . The 20×3 strips on the left of the heatmaps denote whether a gene is BL cycling, PI cycling, or differentially expressed (DR) at FDR level of significance of 0.05 (JTK_cycle test). The heatmaps exhibit the time courses of gene expression over time from the start of the challenge study to its end. Each gene is normalized relative to its maximum value. Panel A: for low shedders the 20 genes are all cycling in BL and 9 are cycling in PI, while none have differential expression in their respective average levels over BL and PI. Panel B. In contrast to Panel A, for high shedders, only 4 of these 20 genes are cycling in BL and only 1 is cycling in PI. On the other hand, more than half (11) of these 20 genes are differential expressed from BL to PI. Thus these 20 genes specify a biochronicity genotype differentiating expression patterns between low and high shedders.

sub-cohort respectively, their expressions were normalized to the range between 0 and 1, so that cycling and change in expression after inoculation can be visually compared. In low shedders, all the genes shown in the figure are cycling in baseline, and about half (9 out of 20) continues to be cycling after inoculation. In contrast, baseline cycling strength is much weaker in high shedders, with only 4 cycling genes. Only 1 of them, DDIT4, retains cycling after inoculation. DDIT4 is one of the most significant cycling genes both in BL and PI.

2.3 Discussion

2.3.1 Enriched pathways of BL cyclers and novel post-inoculation cyclers

In order to understand the biological functions of cyclers, a pathway enrichment analysis was performed, as in Method 2.4. All KEGG pathways were tested, and about 1/3 have been significantly enriched in baseline cyclers. In the enriched pathways, about 80% of the pathways are related to

immune system has been targeted. It is known that various immune processes fall under the control of circadian rhythm, including the regulation of proinflammatory cytokine secretion, trafficking of myeloid and lymphocyte subsets, differentiation and maturation of clinically important leukocyte subsets [HAB+20]. Our results confirm the outsized role of immune related pathways in regulating the healthy basal state which we measured over the 3 day baseline.

Two of the KEGG-annotated Development pathways (Osteoclast differentiation and Axon guidance) were found to be enriched by the baseline cyclers. It would be intriguing to see whether there are known connections between development and circadian rhythm. Circadian rhythm has been previously found to have a role in bone resorption mice[XOF+16]. Furthermore, one gene that has a role in development, the axon-guidance roundabout gene, was reported to be able to alter the pace of the Drosophila circadian clock [BBFC08]. Together with our findings that both of the two development pathways are significantly enriched by baseline cyclers, this may suggest that there might be a tight bi-directional connection between circadian rhythm and development.

It is of interest that, except for ARNTL that is only a WEAK cycler in our baseline data, members in the core CLOCK gene modules are not detected as cyclic at baseline in our dataset. One reason for this may be misalignment of phase: the CLOCK module was not well phase synchronized across study participants. It is also possible that the whole blood assay used in our study contains cells masks the CLOCK chronotype due to the phase diversity of circulating blood cells. Another possibility is that only peripheral tissues are under control of the master clock in the suprachiasmatic nucleus(SCN). One research group hypothesized that core clocks might not be necessary in order for *in vivo* tissues to be cyclic [HCS+14]. Another group identified many biomarkers for human circadian phase in whole blood that are not classical clock genes or clock-controlled genes. [LMLP+17].

Only one KEGG pathway, Protein processing in endoplasmic reticulum, has been enriched by novel cyclers after viral exposure. This is perhaps because most of the novel cyclers are WEAK cyclers and they distribute evenly from different biological pathways.

2.3.2 Novel recruited cyclers may be critical to launching effective immune response.

Despite the fact that more than 70% of the baseline cyclers ceased to be cyclic after inoculation, there are almost same number of novel cyclers after viral exposure as the number of all baseline cycler, just with fewer STRONG and MEDIUM cyclers, as shown in Result 2.2.4.Historically, circadian rhythm has been expected to be largely lost during severe acute infection [HCS+14] . However, recently two studies have challenged this idea, claiming that unique sets of novel cyclers are recruited after infection in LPS-induced endotoxemic mice [HCS+14] and in mice with tissue-specific disruption of *BMAL1* [ZHW+19]. Although it is difficult to compare our novel cyclers to those discovered in these previous studies, which are specific to species and tissue types,

our study adds evidence to these findings for the human host. Specifically, our findings suggest that circadian components recruited after viral exposure could have quite different composition than in the basal state.

2.3.3 Phase-selective cycling resilience to disturbance

To the best of our knowledge this paper is the first to report such a phase-selective response in the context of transcriptional disruption of circadian rhythms due to exposure to infection. Other phase-selective gene responses to intervention treatments have been observed, e.g., chemotherapy drugs can be more effective if administered at particular circadian phases [SLBG$^+$15]. But such intervention phase-selectively targets the circadian rhythm of specific tissues, while the mechanisms governing the influence on circadian rhythm by physiological or pathological environments may be qualitatively different. The phase-selective cycling resilience and quenching after viral inoculation merits additional confirmatory study, going beyond the relatively small cohort size and limited temporal resolution (thrice daily).

2.3.4 Master regulator analysis

It is of great interest to understand the molecular mechanisms that relate the observed immune response to viral exposure to changes in circadian rhythm. To build understanding, we used a human transcription factor database TRRUST v2 [HCL$^+$18]. First, a set of 14 genes are identified as differentially expression from baseline to post-inoculation in full cohort, as well as their transcription factor relationship is confirmed by correlation analysis This constitutes the immune response module on the left of the network shown in Fig. 2.6, containing the JAK-STAT pathway at the center. This pathway is well understood as a primary mediator of the inflammatory response of the immune system [SL03, VKO17].

Subsequently, transcription factors of cyclers were identified in the same manner, forming the circadian rhythm module on the right of the network in Fig. 2.6 with NF-κB1 and its subunit RELA at central positions. NF-κB1 was reported to interact with the core clock protein BMAL1 and to modify the mammalian circadian clock [SWE$^+$20]. Interim connections connecting the two modules were discovered from the TRRUST database. The two modules are connected by three paths, each having two genes(PPARA - NFKBIA; TP53 - BCL3; PML - TNFAIP3). This topological structure might explain the reason why not all the cyclers from baseline were disabled by the inoculation, and that NF-κB1 related pathways might be a provide the mechanism for recruitment of novel cyclers after exposure. This buffering structure reminded us of the regulatory network of *E. coli* metabolism, which could demonstrate both homeostasis and flexibility of response. [SJ08]Among theses buffering genes, NFKBIA, BCL3 and TNFAIP3 belong to the TNF signaling pathway.

Hepatitis C pathway contains TP53 and PPARA. PML is related to Acute myeloid leukemia and Influenza A pathways.

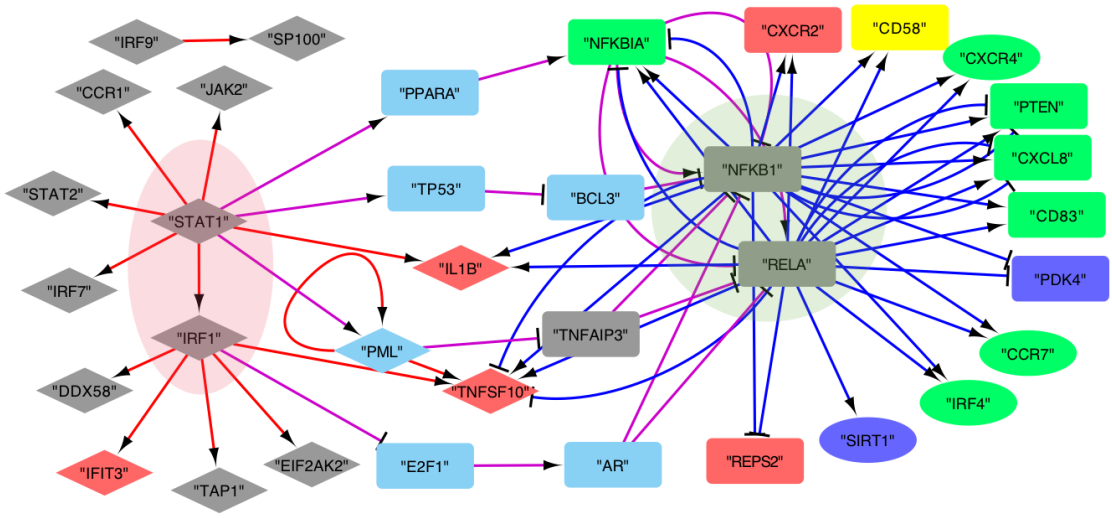


Figure 2.6: Master regulator identified with the transcription factor (TF) database [HCL$^+$18] . The immune response module on the left is comprised of TF-target gene pairs with significantly different expression from baseline to post-inoculation. JAK-STAT pathway has a central position in this module. The circadian rhythm module is on the right of the network with NF-κB1 and its subunit RELA in the center. The two master regulator sets are connected by paths of length two, by finding all possible paths from STAT1 and IRF1 to NF-κB1 and RELA without rings. This structure in the transcriptional network might help explain the resilience of certain clock-related genes and the emergence of circadian rhythm in other genes after inoculation.

2.3.5 Difference in baseline between sub-cohorts

Our discovery that cycling strength differ in low shedders and high shedders suggests potential markers of infection susceptibility. The set of genes have similar mean expression levels over BL and PI but differ in the strength of their chronotypes. For the molecular states we have shown in Sec. 2.2.6 that the weakness and instability of the circadian cycling genes over baseline that differentiates the severely infected subjects from the others . Markers of association are not necessarily indicative of risk factors, so more work is needed to establish if these associations reflect causal relationships or not. Future experiments with larger cohort size, longer baseline and higher temporal sampling resolution would allow us to confirm these findings.

2.3.6 Limitations

Our findings provide interesting perspective on circadian rhythm of blood transcriptome and the interaction between circadian rhythm homeostasis and viral infection. It would be interesting to

perform follow up experiments to further validate our observations. For example, we found that many baseline cyclers are involved in immune related pathways. This is possibly due to the fact that this study took human whole blood as the biological sample, where leukocytes are a major component. It could provide more information if we can combine flow cytometry and RNAseq, which could enable us to examine the circadian rhythm of each cell type or specific cells of interest in human whole blood.

In addition, we observed the phenomenon of phase-selective baseline cycler extinguishment. But with the entire cohort inoculated at the same time of day, it remains unknown whether this phase-selective quenching is based on time of day when the cyclers peaks occur and invariant to inoculation time, or based on the relative distance between the acrophases of cyclers and inoculation time. Additional experiment with inoculations at different time of day may help elucidate this question. If the latter is true, circadian rhythmicity in disease susceptibility may be due to different groups of baseline cyclers are quenched following inoculation.

Limited by the number and frequency of samples, we could only identify significantly cyclic genes on a population / sub-population level, which prevented us from building a predictive model to estimate infection susceptibility with individual baseline cycling strength.

2.4 Materials and Methods

Challenge study A human viral challenge (HVC) study was performed with human rhinovirus (HRV) in 2015 at the University of Virginia. A total of 24 volunteers were recruited and 19 participated completed the study. One of these participants had a failed inoculation and was omitted from our analysis.

The age range of the remaining 18 participants was between 18 and 33, two thirds of these participants were male, and 4 were non-caucasian. The demographics of the 18 individuals included in our analysis is shown in 2.7.

The study was reviewed and approved by the Institutional Review Boards at Duke University and at University of Virginia . Written informed consent was obtained from all participants. Exclusion criteria included pregnancy, chronic respiratory illness, high blood pressure, tobacco/drug/alcohol history, and serum antibody levels above 1:4 titers. The participants were not isolated during the study.

The challenge study lasted 9 days over which various types of biomarkers were continuously collected from participants using wearable wristbands (Empatica E4), whole blood assays (mRNA, p180 metabolites, steroids), nasal-pharyngal washes (viral shedding), cognitive testing (Lumos), and self-reported clinical data (symptoms). On day four at 8am (denoted as 0 hours in this paper) each participant was inoculated via intranasal drops of diluted Human Rhinovirus strain type 16

with a dose of 100TCID50 in 1mL Lactated Ringer's Solution.

Their molecular, physiological, hormonal and cognitive samples were collected both before and after the exposure of HRV. After inoculation, the participants underwent daily nasal lavage each morning to determine the amount of viral shedding and a cumulative shedding score was computed by summing up each person's shedding titrations over the post-inoculation period. The modified Jackson Score requires subjects to rank symptoms of upper respiratory infection (coughing, sore throat, sneezing, headache, etc) on a scale of 0 – 3 of "no symptoms", "just noticeable", "bothersome but can still do activities" and "bothersome and cannot do daily activities". Self-reported symptom scores were also recorded and a cumulative modified Jackson score was computed using the formula described in the section entitled "Case Definitions" in [ZCV+09a].

RNA-seq pre-processing The mapped count data was transformed into TPM (Transcripts per million)[WKL12] for further analysis.

Periodicity estimation The periodicity of all biomarkers were tested following the standard protocol of JTK_CYCLE method [HHK10]. The p-values are adjusted to false discovery rate for multiple testing.

Pathway enrichment analysis The Fisher Exact Test (FET) was used to assign the level of significance (p-values) to the overlap of gene set of interest and 329 pathways in the KEGG database [KGK+04] respectively. The p-values are adjusted to false discovery rate to account for multiple testing.

Master regulator analysis A transcription factor database [HCL+18] is employed to discover the regulatory relationships in gene sets. For a given gene list, all upstream transcription regulators are identified. With all these pairs, a gene network can be visualized in Cytoscape.

2.5 Conclusion

We presented an analysis of biochronicity of a spectrum of biomarkers collected from a human bio-longitudinal viral challenge study. The temporal evolution of these biomarkers fell into several chronotypes with different patterns of circadian cycling in the basal state (baseline) as contrasted to post-inoculation. Many baseline gene expression cyclers are associated with immune response pathways. Although some baseline cyclers maintained their periodicity after inoculation, many of these underwent a shift in cycling phase and strength. In addition, novel cyclers emerged after inoculation. Some baseline cycling genes selectively undergo phase-selective quenching and those that attain minimum cycle amplitude nearer the time of inoculation are more likely to lose periodicity.

The baseline cyclers are more numerous and much stronger in low shedders than in high shedders. Using a transcription regulation analysis on the circadian and differentially expressed genes, we proposed a buffered network mechanism that mediates between the immune response, governed by the JAK-STAT pathway, and the circadian module, governed by the NF-κB signaling pathway and RELA. The two master regulators are buffered by a small number of genes, which may play a role in ensuring resiliency and emergence of novel circadian cycling after inoculation.

The data and analysis result for this paper has been made publicly available at Google Drive (`https://drive.google.com/drive/folders/1DIHrd79dSbexhehwNXvtnJ6OgHjp2R74?usp=sharing`).

2.6 Supplementary Material

Demographics of Volunteers in Human Viral Challenge A table of demographics of volunteers in this challenge study is shown in Fig. 2.7.

		Low shedder	High shedder	Total
Median Age (range)		21 (20-24)	20 (19-33)	20(19-33)
Gender	Male	8	4	12
	Female	1	5	6
Ethnicity	White	7	7	14
	Black	2	0	2
	Asian	0	2	2
Dominant hand	Right	9	9	18
	Left	0	0	0
Median Systolic BP (range)		120 (100-130)	110(100-120)	112(100-130)
Median Diastolic BP (range)		70 (60-80)	70 (58-79)	70 (58-80)
Median Pulse (range)		64 (60-72)	72 (60-80)	67 (60-80)
Median Respiratory Rate (range)		20 (16-20)	18 (16-20)	19 (16-20)
Median Weight (range)		71 (61-89)	65 (55-73)	67 (55-89)
Getup time	before 6:30	0	1	1
	6:30-7:45	1	0	1
	7:45-9:45	5	6	11
	9:45-11	3	1	4
	11-12	0	1	1
Bed time	10:15-12:30	5	5	10
	12:30-1:45	4	2	6
	1:45-3	0	2	2
Productive time	Morning	4	3	7
	Afternoon	3	1	4
	Evening	2	3	5
	Night	0	2	2

Figure 2.7: Demographic distribution of the 18 subjects included in the analysis of the viral challenge study.

Time courses of shedding and symptom Time courses of shedding and symptom are shown in Fig. 2.9.

Relationship of accumulated shedding and accumulated symptom Accumulated shedding and accumulated symptom are linearly correlated, with a Pearson correlation of 0.82 and p-value

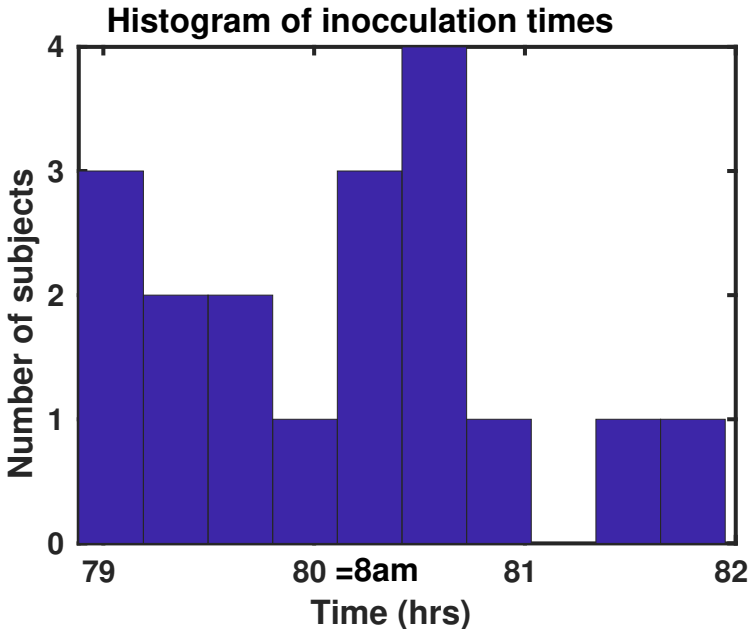


Figure 2.8: Histogram of the inoculation times for the 18 participants. Eighty hours corresponds to 8am on the morning of the 4th day of the challenge study. All but 2 participants are inoculated between of 7 and 9am.

3.28^{-5}, 95% Confidence Interval at $[0.57, 0.93]$. The scatter plot is shown in Fig. 2.10. The accumulated symptom scores is also shown in Fig. 2.11. They are ordered by increasing total shedding levels from left to right. All high shedders have more severe symptom than low shedders except for participant 19.

Definition and illustration of cyclers with different strength Definition of cycling strength categories: We measure strength of cycling of a biomarker using the level of significance of the JTK_CYCLE periodicity test. For each biomarker time course tested JTK_CYCLE outputs a Benjamini-Hochberg FDR p-value or, in the case that there are too few unique values in the time course, a "Fail". We consider any biomarkers that "Fail" as non-cycling. We define 4 different cycling strength levels as follows:

- Strong cycling have smallest BH-Q: FDR $\leq 10^{-6}$
- Medium cycling have smaller BH-Q: $10^{-6} \leq$ FDR $\leq 10^{-3}$
- Weak cycling have small BH-Q: $10^{-3} \leq$ FDR ≤ 0.05
- Non cycling have large BH-Q: $0.05 <$ FDR

These three types of cyclers are illustrated in Fig. 2.12.

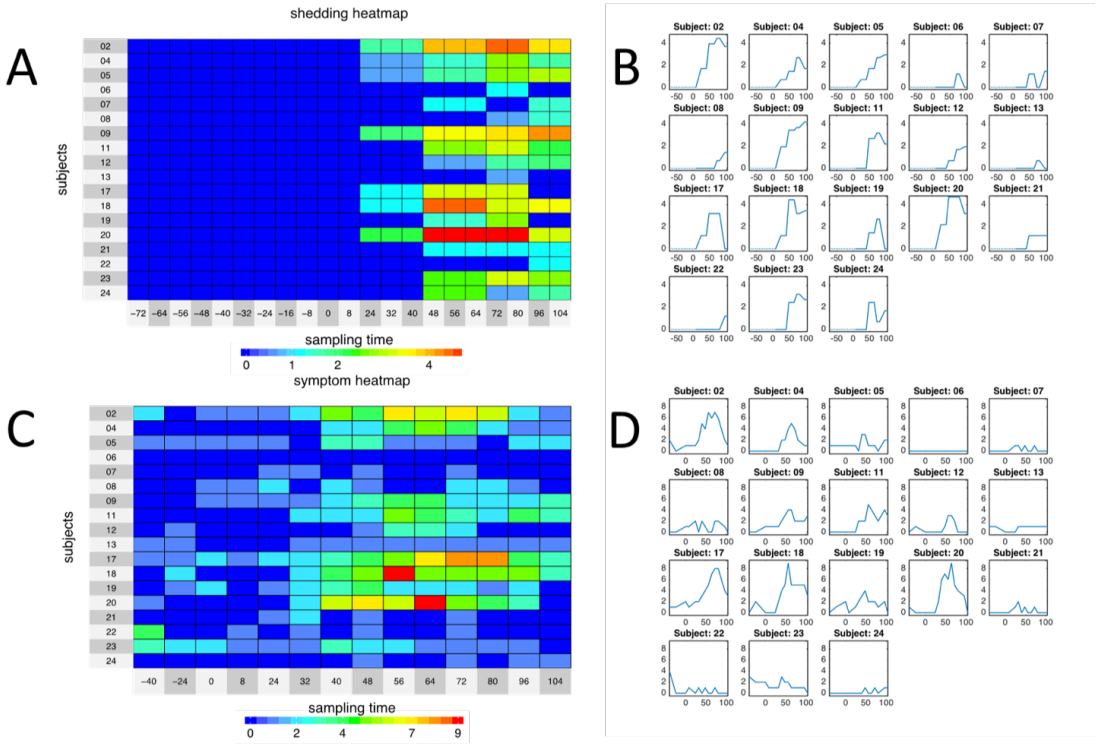


Figure 2.9: Shedding and symptom data for the 18 subjects in the cohort. A-B: Shedding heatmap (A) and shedding curves (B) over the full timecourse of the challenge study. C-D: Symptom heatmap (C) and symptom curves (D), where symptom is a aggregate of 8 different symptoms computed as a modified Jackson score. All subjects exhibit a detectable amount of post-inoculation viral shedding. All subjects except for Subject 6 complained of symptoms.

STRONG cyclers A list of all STRONG cyclers are shown in Fig. 2.13 and Fig. 2.14

Gallery of biomarkers Time course and periodogram of some biomarkers mentioned in the paper: DDIT4 (Fig. 2.15), CAT (Fig. 2.16), TLR2 (Fig. 2.17), etc.

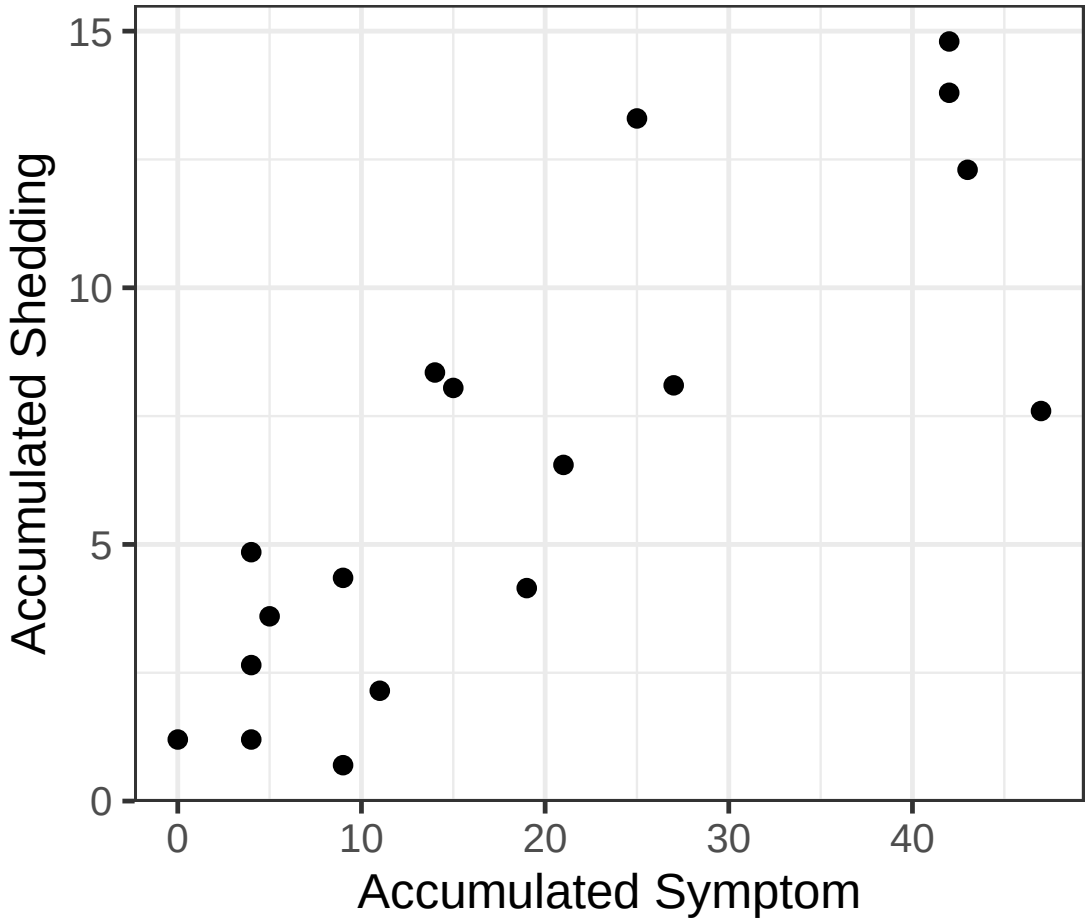


Figure 2.10: The linear correlation between accumulated shedding and accumulated symptom. (Pearson correlation 0.82 with p-value 3.28^{-5}.)

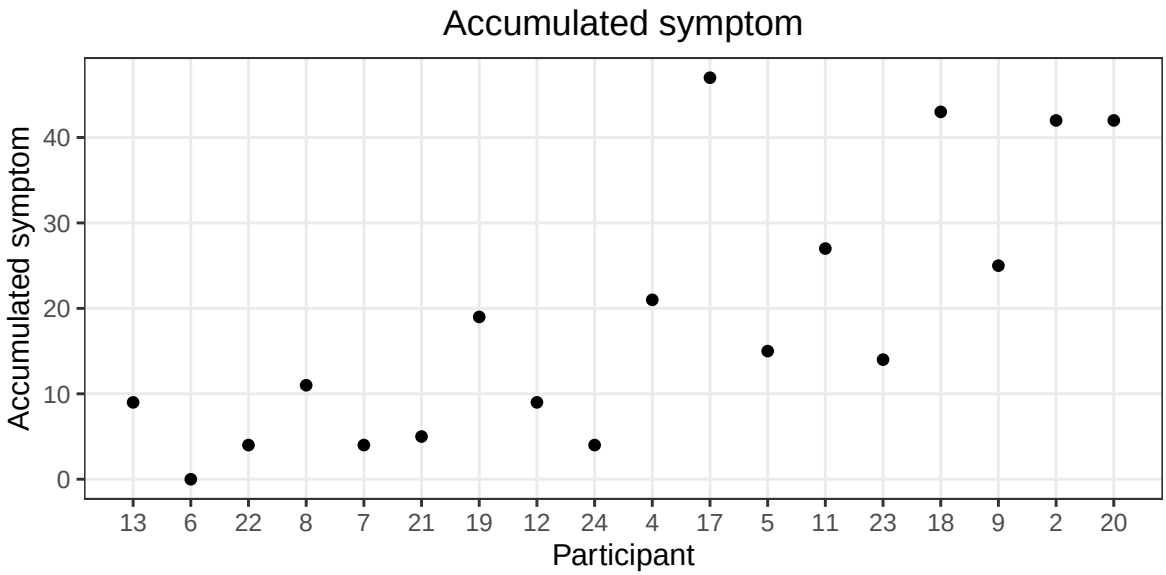


Figure 2.11: Accumulated symptoms of participants ordered by increasing total shedding levels from left to right. All high shedders have more severe symptom than low shedders except for participant 19.

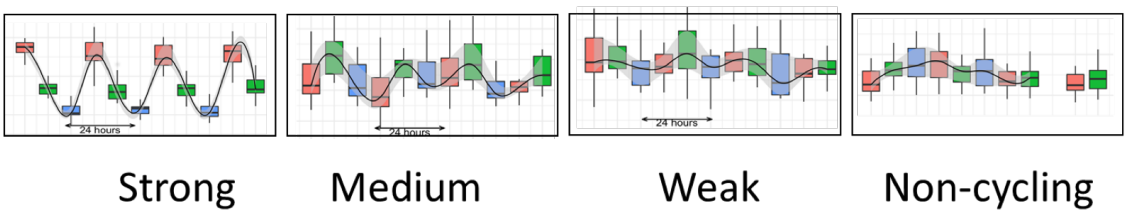


Figure 2.12: Illustration of biomarkers that cycle with different circadian cycling strength levels.

Type	Name	BH.Q	LAG	Type	Name	BH.Q	LAG	Type	Name	BH.Q	LAG
Genes	DDIT4	1.40E-21	12	Genes	FAM126B	1.00E-07	20	Genes	RAB31	9.00E-07	20
	TSC22D3	8.69E-12	12		STX11	1.14E-07	20		CLEC4E	9.71E-07	16
	DUSP1	7.48E-11	12		EGLN1	2.12E-07	20		SLC2A3	9.81E-07	20
	FKBP5	9.02E-11	12		TMCC3	2.12E-07	20	p180	lysoPC.a.C18.2	1.51E-11	0
	FOSL2	2.09E-10	20		CD69	2.26E-07	8		alpha.AAA	1.06E-07	0
	CPPED1	7.01E-10	20		MGAM	2.41E-07	20		lysoPC.a.C20.3	1.06E-07	0
	CD180	1.35E-09	0		CNTNAP3	2.87E-07	20		lysoPC.a.C16.0	2.03E-07	0
	PIK3IP1	3.21E-09	8		LILRA5	2.87E-07	20		Met.SO	4.71E-07	0
	PRRG4	3.86E-09	20		MAN1A1	2.87E-07	20		Orn	6.58E-07	0
	RHOB	3.87E-09	20		MSL1	2.87E-07	20		lysoPC.a.C20.4	7.57E-07	0
	MRVI1	4.36E-09	20		RAB11FIP1	2.87E-07	20	Hormones	cortisol	2.84E-45	12
	MYADM	1.04E-08	0		FRAT2	3.47E-07	20		melatonin	1.00E-09	4
	KIAA0825	1.94E-08	20		ARHGAP9	3.54E-07	20	E4	HR_mean	1.14E-78	17
	LITAF	1.94E-08	20		CXCL8	3.71E-07	12		HR_median	4.75E-77	17
	KLF9	2.01E-08	12		TUBA1A	4.31E-07	20		ACC_sd	1.48E-67	16.5
	CXCR4	4.27E-08	12		ARHGEF40	4.37E-07	20		HR_sd	5.51E-47	16
	BCL6	4.40E-08	20		GPAT3	4.37E-07	20		ACC_mean	1.42E-38	16
	TNFRSF10C	7.83E-08	20		BASP1	4.40E-07	20		TEMP_sd	2.26E-31	16
	ABCA1	7.97E-08	20		CHSY1	4.99E-07	20		ACC_median	3.54E-18	16
	ADGRG3	9.05E-08	20		TNFSF8	6.15E-07	12		TEMP_mean	7.37E-18	0.5
	RBM3	9.09E-08	12		NRBF2	7.69E-07	20		TEMP_median	2.15E-17	0.5

Figure 2.13: Circadian biomarkers cycling over baseline found to be statistically significant using JTK_CYCLE analysis. The cyclic phases of these circadian biomarkers (in hours in 3rd column of each category) are indicated. Only biomarkers with FDR significance level less than or equal to 0.000001 are shown. A total of 63 biomarkers are discovered with p-values under the FDR threshold (denoted BH.Q in the Table) including: 45 genes; 7 p180 metabolites; 2 hormones; and 9 E4 features.

Type	Name	BH.Q	LAG	Type	Name	BH.Q	LAG
gene	DDIT4	2.09E-13	12	p180	Met.SO	4.13E-08	0
	TLR2	4.24E-11	8		lysoPC.a.C18.2	4.13E-08	0
	CCND3	8.91E-10	8	hormone	cortisol	4.48E-22	12
	FKBP5	9.90E-10	8		melatonin	2.69E-08	4
	CXCR4	5.17E-09	8	E4	HR_median	9.90E-60	17.5
	KLF9	5.84E-09	8		HR_mean	1.63E-58	17.5
	NFKBIA	1.04E-08	8		ACC_sd	8.59E-58	17
	DUSP1	2.32E-08	12		HR_sd	6.37E-39	17
	PIK3IP1	5.37E-08	8		ACC_mean	1.66E-21	16.5
	TSC22D3	6.86E-08	8		TEMP_sd	2.72E-13	16
	CAT	7.04E-07	8				

Figure 2.14: Circadian biomarkers cycling over post-exposure period found to be statistically significant using JTK_CYCLE analysis. The cyclic phases of these circadian biomarkers (in hours in 3rd column of each category) are indicated. Only biomarkers with FDR significance level less than or equal to 0.000001 are shown. Note that there are substantially fewer strongly cycling biomarkers than at baseline 2.13.

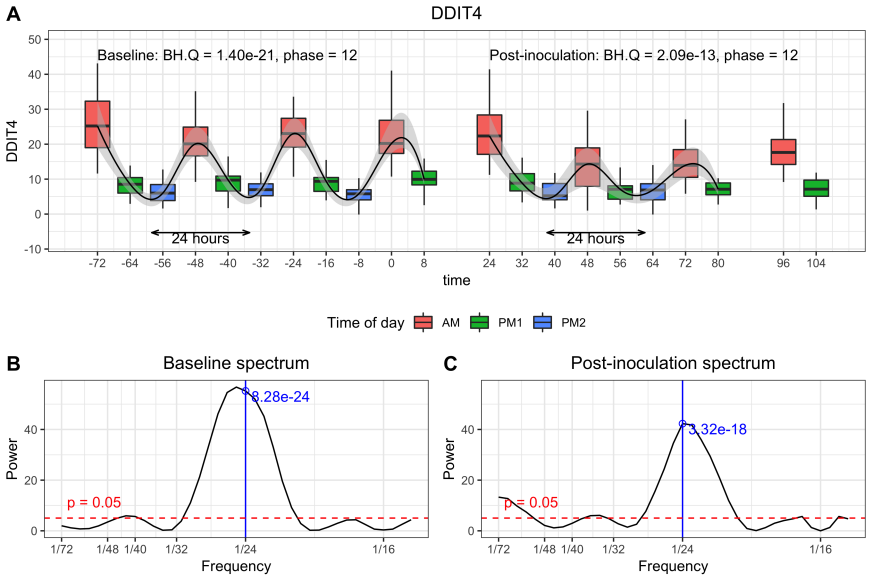


Figure 2.15: Time course of DDIT4, and periodogram in baseline and post viral exposure.

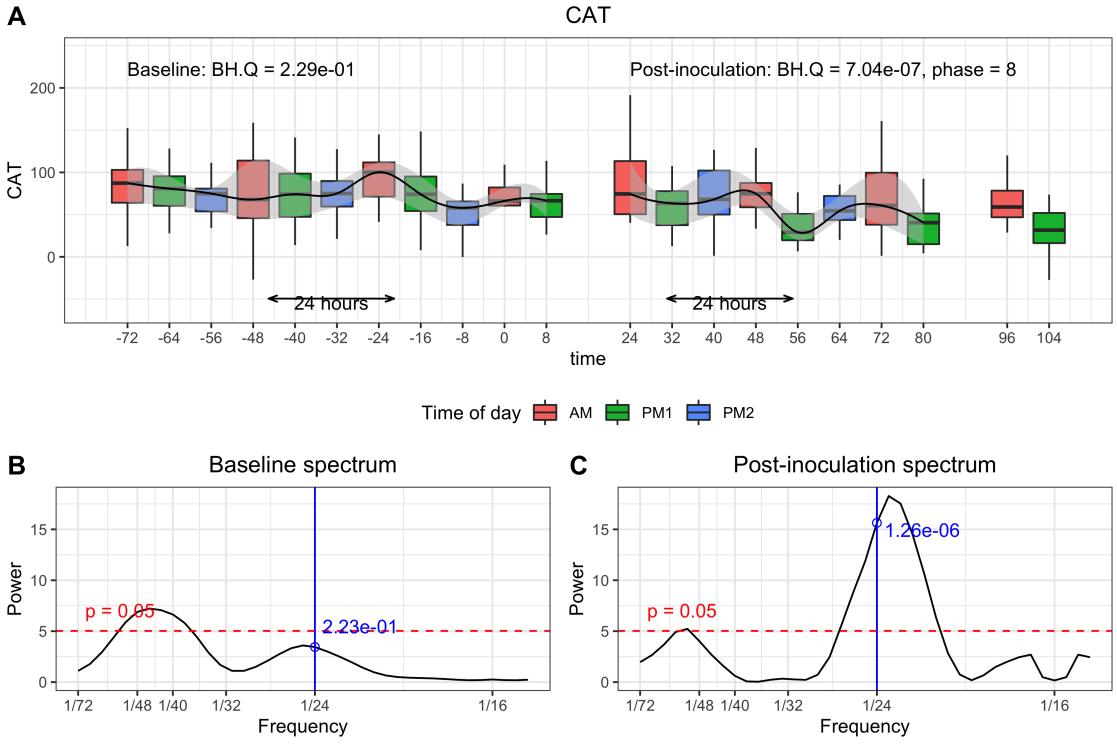


Figure 2.16: Time course of CAT, and periodogram in baseline and post viral exposure.

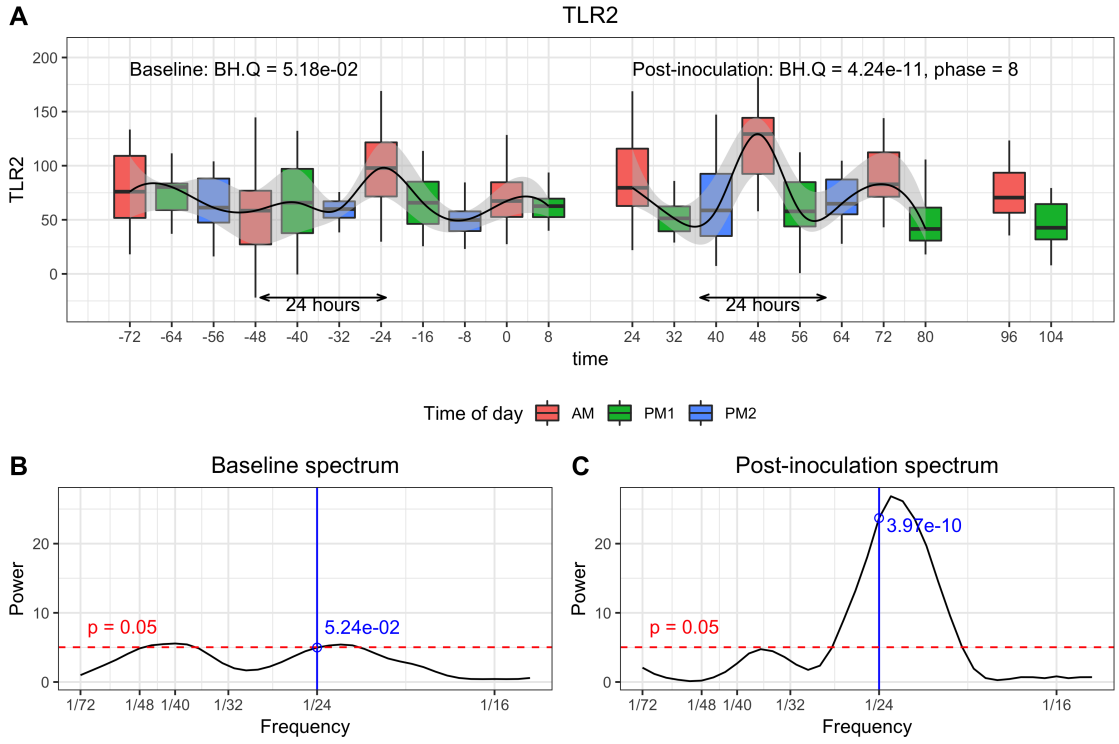


Figure 2.17: Time course of TLR2, and periodogram in baseline and post viral exposure.

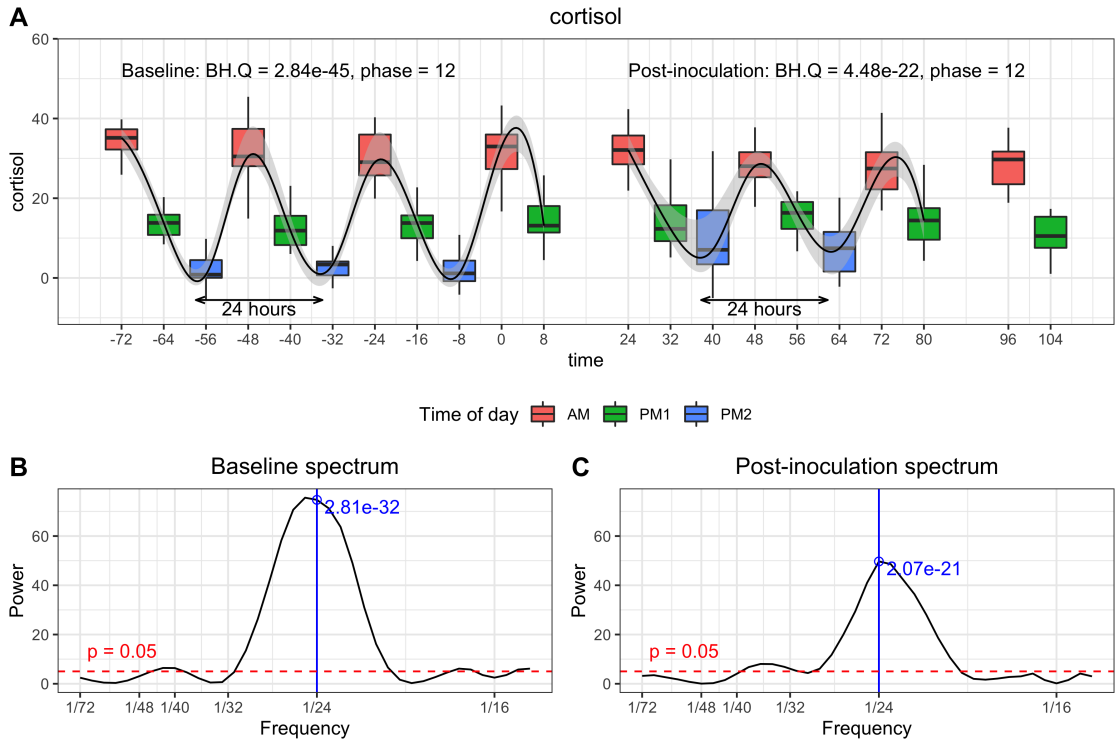


Figure 2.18: Time course of cortisol, and periodogram in baseline and post viral exposure.

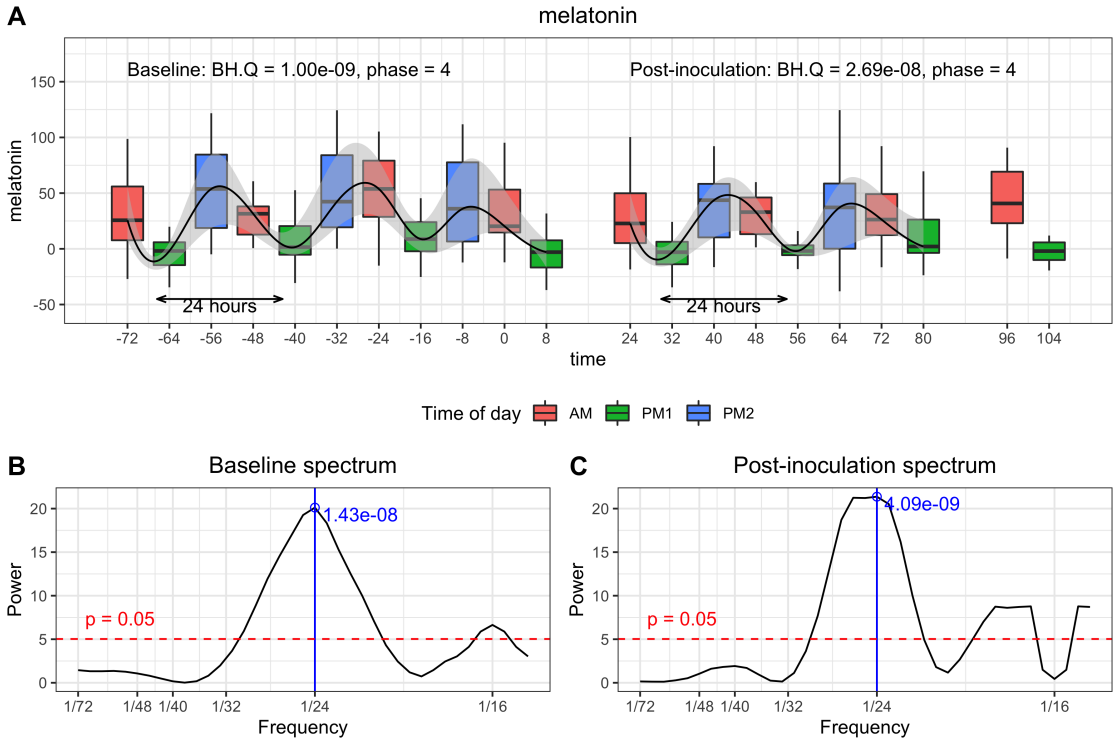


Figure 2.19: Time course of melatonin, and periodogram in baseline and post viral exposure.

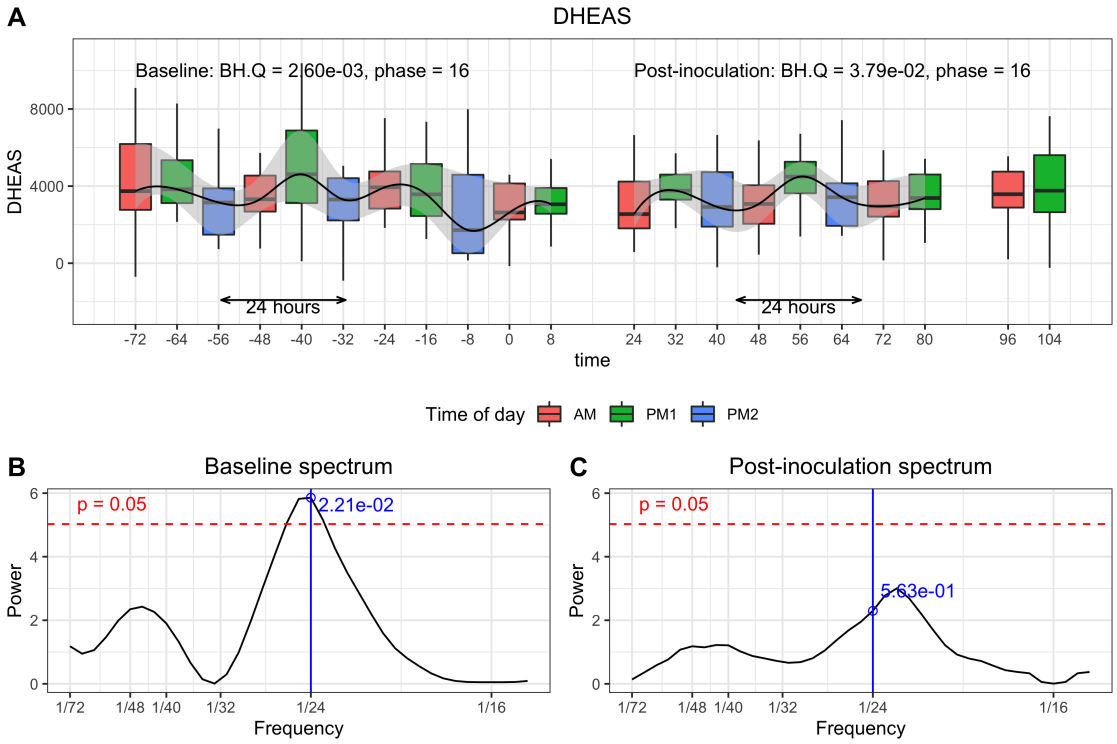


Figure 2.20: Time course of DHEAS, and periodogram in baseline and post viral exposure.

CHAPTER 3

Adaptive Sleep Detection and Physiological Features for Infection State and Onset Prediction

3.1 Introduction

Many physiological time series involve dynamic transition among event states. For example, the transitions in human circadian cycling alternate between sleep and wake states over a 24-hour period. The sleep state can itself be a dynamic process, switching over different stages of rapid eye movement (REM) and non-REM sleep states [CD+05]. As another example, the transitions in menstrual cycles go through follicular and luteal phase [SOG+10], transitioning to an ovulation state. With the emergence of cheap wearable multi-channel physiological monitoring devices, there has been much interest in automating the detection of physiological event states. The potential impact of automated event detection is far reaching, potentially improving a person's health awareness and aiding the management of disease.

A principal impediment to automating event detection for wearable devices is the intrinsically high variability of the measured physiological signals over time. An especially challenging situation is when the wearer of the device undergoes a strong perturbation, such as exposure to a pathogen that results in infection. Most event detection algorithms that are trained over a healthy baseline time period will have difficulty adapting when the device's signals become strongly perturbed away from baseline. Modern machine learning approaches to training that incorporate transfer learning can mitigate these difficulties. This paper introduces an adaptive transfer learning method for automating event detection for wearable devices and demonstrates its ability to adapt to strong perturbations from a healthy baseline in the context of a human viral challenge study.

Another challenge to automating event detection in time series data is the lack of available ground truth event labels. The training of the automated algorithm must thus be unsupervised, based only on the observable physiological signals measured by the device. Effective unsupervised learning is only possible when features extracted from these signals have statistical distributions that strongly depend on the event states. Selection of strong discriminating features is therefore a crucial step in designing an unsupervised learning algorithm. The adaptive transfer learning algorithm

introduced here generates a set of strongly discriminating event-labeled features that are based on statistical summaries of the device signals at multiple time scales.

A state-of-the-art unsupervised algorithm for detecting unobserved event state sequences from time series data is the hidden Markov model (HMM) [ZML17]. The HMM is a generative model in which the hidden states are latent random variables that condition the joint probability distribution of the data. HMM methods have been widely applied to many health applications [BKN17, KMB$^+$10, Yu12, TMC$^+$13, BDTH12]. However, the HMM is not well adapted to data whose statistical distribution may shift over time. When there are external factors that cause perturbations to the distribution, the training data may be mismatched to the post-training data and the HMM will fail to perform as well as expected. In the machine learning literature this phenomenon is known as *data covariate shift* [MTRAR$^+$12]. Many methods have been proposed to address the covariate shift problem, principally for supervised learning [WK96, KJ00, KŽ09, PY$^+$10, GŽB$^+$14].

The adaptive unsupervised transfer learning method for event state classification introduced in this paper is designated by the acronym HMM-FLDA. It involves learning a multivariate HMM latent state prediction model that initializes a sequential version of Fisher's linear discriminant analysis (FLDA) to adapt the initialized HMM to perturbed post-baseline data. Our adaptive transfer learning method is introduced as the key component of a complete data processing pipeline that includes pre-processing for local feature extraction and abnormal signal detection, adaptive event classification using HMM-FLDA, and post-processing for constructing strong high dimensional event-labeled features for inferring perturbation-dependent events, e.g., health outcome after exposure to a viral pathogen.

We illustrate the proposed analysis pipeline for a human viral challenge (HVC) experiment in which physiological data from multiple subjects is collected over multiple days from wearable wristband devices. On the second day of the experiment the subjects were exposed to a H1N1 flu virus that caused some to become infected, as clinically determined by viral shedding which does not occur before the 4th day. For each subject, the HMM-FLDA is trained on the first two days and nights of wearable data to generate sleep/wake features. Using these features we are able to predict if the subject will get infected and the onset time, defined as the first day that shedding is detected.

The remainder of this paper is organized as follows: we introduce in detail the proposed transfer learning algorithm in Section 3.2. Section 3.3 presents a numerical simulation study emulating the HVC experimental data, but with ground truth event labels. Section 3.4 applies the proposed analysis pipeline to wearable data collected in the experimental HVC study. In Section 3.5 we end the paper with discussion and conclusions. Supporting information on the algorithm, the pipeline, and the HVC data is presented in the supplementary Appendices.

3.2 Adaptive unsupervised event monitoring

Figure 3.1a summarizes the proposed adaptive event labeling procedure and Algorithm 1 provides pseudo-code for its principal steps. Figure 3.1b shows the entire data processing and analysis pipeline, from sensor data capture to event-labeled feature extraction to predictive models using these features.

Algorithm 1 HMM-FLDA gradual self-training event state classifier

Initialize with: $\{\hat{y}_t, \boldsymbol{x}_t\}_{t=1}^{t_{N_0}}$ ▷ Use HMM to classify baseline data
while $t_{N_0} < t < t_N$ **do** ▷ Continue over all batches of samples
  for l in 1 to L **do** ▷ Iterate over candidate training sizes d_l
    procedure FISHER'S LDA(train = $(t - d_l, t]$, test = $(t, t + \Delta_t]$)
      $\boldsymbol{w} = \boldsymbol{S}_W^{-1}(\bar{\boldsymbol{x}}_{1\cdot} - \bar{\boldsymbol{x}}_{0\cdot})$ ▷ Update FLDA weights on train data
      $z_t = \boldsymbol{w}^T \boldsymbol{x}_t$
      $\hat{y}_t(d_l) = \mathbb{1}\{(z_t - \bar{z}_{0\cdot})^2/\hat{\sigma}_0^2 - (z_t - \bar{z}_{1\cdot})^2/\hat{\sigma}_1^2 > \log(\gamma\, \hat{\sigma}_1^2/\hat{\sigma}_0^2)\}$ ▷ Classify test data with updated weights
      $\text{SI}(d_l) = \text{SI}(\text{train}, \text{test}; d_l)$ ▷ Compute separability index (3.5)
  $d_{(1)} =$ **which.max**$(\text{SI}(d_l))$ ▷ Select optimal training window length
  $\hat{y}_t = \hat{y}_t(d_{(1)})$ ▷ Classified labels of current batch
  $t = t + \Delta_t$ ▷ Move on to the next batch

The pre-processed data comes in the form of a $p \times N$ matrix $\boldsymbol{X} = [\boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_N}]$ where $\boldsymbol{x}_t \in \mathbb{R}^p$ is a p-dimensional real valued feature vector, available at a set of time instants $\{t_i\}_{i=1}^N$. We assume a statistical time series model for $\boldsymbol{x}_t$ with hidden states that determine the joint probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$. Specifically, associated with each time sample $\boldsymbol{x}_t$ is a latent (hidden) variable $y_t \in \{0, 1, \ldots, K\}$, which labels the event associated with $\boldsymbol{x}_t$. Later in the paper we will specialize to the binary case where $K = 2$ and the events correspond to a person's sleep or wake state at a particular time. We denote $\boldsymbol{y} = [y_{t_1}, \ldots, y_{t_N}]$ the time sequence of a subject's hidden event states. In the time series setting, $\boldsymbol{y}$ determines the conditional probability distribution $p_{\boldsymbol{X}|\boldsymbol{y}}$ and the marginal probability distribution $p_{\boldsymbol{X}}$. We construct an event classifier function based on an HMM with a parametric (Gaussian) conditional distribution of $\boldsymbol{x_t}$ given y_t that may gradually shift over time, e.g., due to a perturbation after exposure to a pathogen. The parameters of the classifier function are estimated from the data in a two phase process involving: (1) an initial training phase where the HMM model is fitted to a batch of (healthy) baseline data, followed by (2) an adaptation phase where a naive Bayes Gaussian model is adapted to future batches of possibly perturbed data using Fisher Linear Discriminant Analysis (FLDA) in a sequential transfer learning framework.

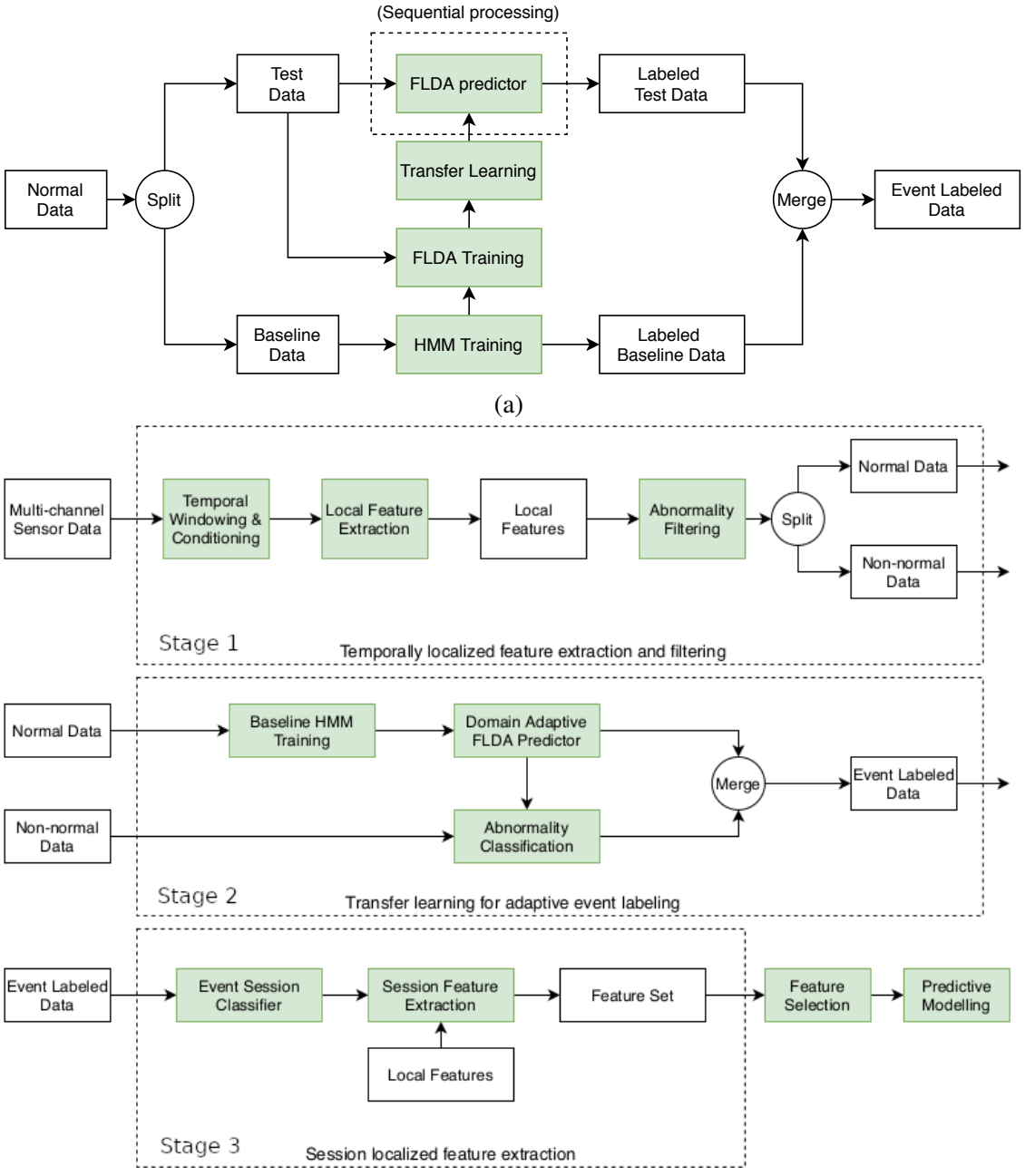


Figure 3.1: **(a)** The proposed HMM-FLDA unsupervised transfer learning algorithm based on a multivariate hidden Markov model (HMM) and Fisher's linear discriminant analysis (FLDA) to segment event states; **(b)** The three stages of proposed data processing pipeline, discussed in Section 3.4.1. The HMM-FLDA procedure depicted in **(a)** appears in the second stage of the pipeline.

3.2.1 Baseline training: HMM initial segmentation

For $N_0 < N$ define the *baseline* (BL) segment of the data $\boldsymbol{X}$ as the first N_0 time samples $\boldsymbol{X}_{\rm BL} = [\boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_{N_0}}]$ and the *post-baseline* (PBL) segments of the data as the remaining part $\boldsymbol{X}_{\rm PBL} = [\boldsymbol{x}_{t_{N_0+1}}, \ldots, \boldsymbol{x}_{t_N}]$. The BL segment is used as a training set for learning the parameters of a multivariate HMM [VS+10]. The HMM is a Markov model: modeling the observations as an N-length segment of a Markov random process $\{(\boldsymbol{x}_t, y_t)\}_{t=-\infty}^{\infty}$, where transitions $y_{t-1} \to y_t$ between event states have probabilities specified by a $K \times K$ state transition matrix $\boldsymbol{A}$. The Markov property implies that the joint distribution $p_{\boldsymbol{X},\boldsymbol{y}}$ factorizes as: $p_{\boldsymbol{X},\boldsymbol{y}} = p_{\boldsymbol{x}_{t_1},y_{t_1}} \prod_{i=2}^{N_0} p_{\boldsymbol{x}_{t_i},y_{t_i}|\boldsymbol{x}_{t_{i-1}},y_{t_{i-1}}}$, implying an analogous factorization of the marginal $p_{\boldsymbol{X}}$ and the conditionals $p_{\boldsymbol{X}|\boldsymbol{y}}$. In the special case of a Gaussian HMM model each factor $p_{\boldsymbol{x}_t|y_t}$ is a multivariate Gaussian conditional distribution, denoted as

$$\boldsymbol{x}_t \mid (y_t = k) \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad k = 1, \ldots, K, \tag{3.1}$$

where $\boldsymbol{\mu}_k$ is a p-dimensional mean parameter and $\boldsymbol{\Sigma}_k$ is a $p \times p$ covariance matrix parameter that must be learned in addition to the state transition matrix $\boldsymbol{A}$.

There are several methods available for learning multivariate HMM's [GS11, ZML17] which could be adapted to our setting. Most methods are iterative and many use a variant of the expectation-maximization (EM) algorithm to find maximum likelihood estimates of the unknown parameters. Important practical considerations are whether the HMM algorithm provides a final estimate of labeling accuracy, important for self diagnostics, if it converges sufficiently rapidly to a global maximum, and if it has fast computation time per iteration, which tends to increase as K and p increase. In some applications, it may be desirable to estimate the number of event states K, for which automated model selection versions of HMM are available. For a comprehensive review of HMM implementations the reader can refer to [ZML17, Ch. 8]. In our pipeline we use a reasonably fast multivariate HMM algorithm implemented by EM iterations that approximates the maximum likelihood estimator of the HMM parameters. In the Expectation (E) step, the marginal likelihood is computed by a variant of the forward-backward algorithm [LH02], which calculates the gradient of the log-likelihood (score) function in a single pass.

3.2.2 Post-baseline adaptation: sequential transfer learning

If the feature distribution varies over time, the static baseline-trained HMM will have difficulty classifying and segmenting events in the post-baseline data $\boldsymbol{X}_{\rm PBL}$. To address this difficulty we introduce a transfer learning strategy that is initialized with the HMM on the initial-training data and sequentially updates the event classifier over successive batches of test data, continually adapting to changes in distribution. In transfer learning the batches are called target domains and the objective is to design a classifier that continually adapts to them, a property called domain adaptation [KF11],

[MTRAR+12]. To achieve this objective, we use an unsupervised *gradual self-training* procedure. In supervised transfer learning, "self-training" means that domain adaptation is done with an unlabeled test set and "gradual" means that the adaptation is done sequentially over time, updating over successive batch pairs. As compared to direct self-training procedures, which try to adapt to all test batches at once, gradual self-training procedures are better suited to online applications and are more robust to smoothly varying shifts in distribution. Recent theory establishes that gradual procedures are provably better when the class distributions on successive pairs of batches are close in Wasserstein distance [KML20]. This closeness condition is satisfied when the distributions are quasi-stationary, an assumption common in time series analysis [Ada98] and online supervised learning [SS+12].

Our unsupervised gradual self-training procedure consists of three components: 1) an unsupervised version of Fisher's LDA that uses class labels predicted by the HMM to project the data $\boldsymbol{X}_{\mathrm{BL}}$ to a one dimensional space that maximally separates these classes. 2) application of this projection to the next batch of samples in $\boldsymbol{X}_{\mathrm{PBL}}$ followed by event classification using a naive Bayes classifier; 3) using the class labels to update the projection by reapplying Fisher's LDA to the batch filtered by a tapered sliding-window. This process is sequentially repeated for each successive batch, resulting in a continuous adaptation of the event classifier.

Fisher's LDA is most commonly applied in the context of supervised learning for dimension reduction and classification [Fis36, FHT01] when a labeled sample is available for training. Here we adapt LDA to the unsupervised context of latent event classification. After applying HMM to the baseline sample $\boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_{N_0}}$ we obtain estimated class labels $\hat{y}_{t_1}, \ldots, \hat{y}_{t_{N_0}}$. We then train Fisher's LDA classifier on the putatively labeled sample $\{(\boldsymbol{x}_{t_1}, \hat{y}_{t_1}), \ldots, (\boldsymbol{x}_{t_{N_0}}, \hat{y}_{t_{N_0}})\}$. Fisher's LDA uses dimensionality reduction to learn a classifier. By applying a weight $\boldsymbol{w} \in \mathbb{R}^p$ to $\boldsymbol{x}_t$ FLDA generates a projection score $z_t = \boldsymbol{w}^T \boldsymbol{x}_t$. The weight vector is optimized so that these scores attain the largest possible separation of the classes. Fisher's LDA accomplishes classification of a novel sample $\boldsymbol{x}_{t'}$ by using the trained weights to compute the projected score $\boldsymbol{w}^T \boldsymbol{x}_{t'}$ that is used in an optimal naive Bayes LDA classifier to predict its unknown label $y_{t'}$.

Specifically, for the case of $K = 2$ classes $y_t \in \{0, 1\}$, in the dimensionality reduction stage of FLDA the class separability of the projection scores $z_t = \boldsymbol{w}^T \boldsymbol{x}_t$ is measured by the following ratio of *between-class* variation and *within-class* variation:

$$J(\boldsymbol{w}) = \frac{(\bar{z}_{1\cdot} - \bar{z})^2 + (\bar{z}_{0\cdot} - \bar{z})^2}{\sum_{t \in G_1} (z_t - \bar{z}_{1\cdot})^2 + \sum_{t \in G_0} (z_t - \bar{z}_{0\cdot})^2} \,, \tag{3.2}$$

where $G_k = \{t \leq N_0 : y_t = k\}$ denotes the data indices of $\boldsymbol{x}_t$ coming from event class k, $\bar{z}_{k\cdot} = |G_k|^{-1} \sum_{t \in G_k} z_t$ for $k \in \{0, 1\}$ are the class-specific sample mean projected score and $\bar{z}$ is

the overall sample mean projected score. The optimal $\boldsymbol{w}$ that maximizes (3.2) is [FHT01]

$$\boldsymbol{w} = \boldsymbol{S}_W^{-1}(\bar{\boldsymbol{x}}_{1\cdot} - \bar{\boldsymbol{x}}_{0\cdot}), \tag{3.3}$$

where $\bar{\boldsymbol{x}}_{k\cdot} = |G_k|^{-1} \sum_{t \in G_k} \boldsymbol{x}_t$ is the class-specific sample mean and $\boldsymbol{S}_W = \sum_{k=0,1} \sum_{t \in G_k} (\boldsymbol{x}_t - \bar{\boldsymbol{x}}_{k\cdot})(\boldsymbol{x}_t - \bar{\boldsymbol{x}}_{k\cdot})^T$ is the *within-class* scatter matrix of the $\boldsymbol{x}_t$'s.

Under the multivariate normal model (3.1) for $\boldsymbol{x}_t$, the z_t's are conditionally independent Gaussian distributed random variables with conditional mean μ_k and conditional variance σ_k^2 given $y_t = k$. Assuming independent conjugate priors on μ_k and σ_k^2, conditioned on the training data $\boldsymbol{X}_{\mathrm{BL}}$, the maximum a posteriori (MAP) estimator of the class label $y_{t'}$ of a novel test sample $\boldsymbol{x}_{t'}$ based on its projected score $z_{t'}$ is given by the naive Bayesian classifier [FHT01]:

$$y_{t'} = \mathbb{1}\left\{ \frac{(z_{t'} - \bar{z}_{0\cdot})^2}{\hat{\sigma}_0^2} - \frac{(z_{t'} - \bar{z}_{1\cdot})^2}{\hat{\sigma}_1^2} > \log\left(\gamma \, \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right) \right\}, \tag{3.4}$$

where $\mathbb{1}(\mathcal{A})$ is the binary indicator function equaling 1 when the logical preposition $\mathcal{A}$ is true. Here, as above, $\bar{z}_k$ is the class-specific sample mean and $\hat{\sigma}_k^2$ the within class sample variance of $\{z_t : t > N_0\}$ $\hat{\sigma}_k^2 = (|G_k| - 1)^{-1} \sum_{t \in G_k} (z_t - \bar{z}_{k\cdot})^2$. The threshold parameter γ in (3.4) is the prior odds ratio $P(y_{t'} = 0)/(1 - P(y_{t'} = 0))$ and can be adjusted to account for class imbalance and to trade-off the two types of classification errors. The naive Bayes classifier classifies the label of the test sample based on its relative distance to each of the class centroids $\bar{z}_k$ weighted by the class dispersions $\hat{\sigma}_k^2$.

We next describe the proposed unsupervised transfer learning procedure using gradual self-training to sequentially update the HMM-FLDA algorithm as, outlined in Algorithm 1 and depicted in Fig. 3.1a. The post-baseline data is successively divided into disjoint test batches of length Δ_t secs., called the test window length, a tuning parameter whose selection is application dependent. The current batch is used as a test sample to update the self-trained HMM-FLDA classifier determined from the previous batch. Let $t_n^{tst} = t_{n-1}^{tst} + \Delta t$ be the start time of the n-th test window and denote the n-th test batch as $\mathcal{T}_{tst} = \{\boldsymbol{x}_{t_i}\}_{t_i \in [t_n^{tst}, t_n^{tst} + \Delta_t)}$. The update proceeds as follows. For a specified size d, define the n-th training set $\mathcal{T}_{trn} = \{\boldsymbol{x}_{t_i}\}_{t_i \in [t_n^{tst} - d, t_n^{tst})}$. Note that the feature instances in $\mathcal{T}_{trn}$ have already been assigned class labels $\{\hat{y}_{t_i}\}_{t_i \in [t_n^{tst} - d, t_n^{tst})}$ in the previous update of HMM-FLDA. Hence, the first step in updating the classifier is to compute the updated FLDA weight vector $\boldsymbol{w}$, as defined in (3.3), using $\mathcal{T}_{trn}$ and its previously assigned class labels. The second step is to use this updated weight vector to assign predicted labels $\{\hat{y}_{t_i}\}_{t_i \in [t_n^{tst}, t_n^{tst} + \Delta t)}$ to instances in $\mathcal{T}_{tst}$.

The size d of $\mathcal{T}_{trn}$ is adapted from batch to batch by optimizing a separability index (SI) defined over $\mathcal{T}_{trn} \cup \mathcal{T}_{tst}$. Among the many possible SI measures that could be used, we adopt Thornton's SI

[Tho98], a competitive measure for assessing class separation originally introduced for supervised classification problems [AR07, CMS12, Gre01]. Also called the geometric separability index (GSI), Thornton's measure has seen wide application in health, robotics, geology and other fields [MR10, CL14, SR14, TMN10, RWM15]. We use the following simple unsupervised modification of the GSI [Tho98], computed on the merged training and test samples for the batch:

$$\text{SI}(d) = \frac{\sum_{t \in \mathcal{T}_{trn}(d) \cup \mathcal{T}_{tst}} \{\hat{y}_t + \tilde{y}_t + 1\} \textbf{ MOD } 2}{|\mathcal{T}_{trn}(d)| + |\mathcal{T}_{tst}|}, \tag{3.5}$$

where $\tilde{y}_t$ is the predicted class of the nearest neighbor of $\boldsymbol{x}_t$ in the set of merged training and test samples. As compared to Thornton's original SI, defined for supervised classification problems where the true labels y_t are known, the unsupervised modification (3.5) uses the predicted labels $\hat{y}_t$ in place of y_t. Here the nearest neighbors are determined by the "projection distance" defined, for samples at time t and t', as $d(t, t') = |\hat{\boldsymbol{w}}^T(\boldsymbol{x}_t - \boldsymbol{x}_{t'})|$. Rank ordering the SI indices $\text{SI}(d_{(1)}) > \ldots > \text{SI}(d_{(L)})$ yields $d_{(1)}$ as the optimal training window length for the batch.

3.3 Numerical simulation study

We performed a simulation of HMM-FLDA that emulates the experimental study presented in Section 3.4, but with ground truth event states. Two dimensional data $\boldsymbol{x}_t = [x_t^{(2)}, x_t^{(2)}]$ is simulated from conditional distributions given the latent event state y_t, which randomly switches between wake ($y_t = 0$) and sleep ($y_t = 1$) with a mean cadence of approximately 24 hours. To emulate the perturbation effect of viral infection, each of these conditional distributions are fixed during the baseline training period but may undergo slowly varying time shifts in the post-baseline period. Two different post-baseline scenarios are simulated, called the stable case (no shift) and the unstable case (slow shift).

For the stable case, we generate the sequence of states as a realization of a Markov process with state transition probabilities that are empirically matched to those extracted from the HVC data described in the next section. We fix the number of state transitions T and the initial state is set to $y_0 = 0$. The duration $\Delta\tau_1$, i.e., the time to the first transition, is drawn from $TN(\mu_0, \sigma_0)$, a truncated normal (TN) distribution with location and scale parameters μ_0 and σ_0. The duration $\Delta\tau_2$ of the second event is independently drawn from a TN distribution with mean $\mu_1 = 24 - \mu_0$ and variance $\sigma_1^2 = \sigma_0^2$. This process is repeated until the T-th transition variable $\Delta\tau_T$ has been drawn. The discretized transition times $\tau_m = \sum_{i=1}^{m} \Delta\tau_i$, $m = 1, \ldots, T$, specify the state sequence $\{y_t, t = 1, \ldots, N\}$. A consecutively occurring pair of (wake, sleep) periods is called a session. Over a given session m, the data channels $x_t^{(i)}$, $i = 1, 2$, were drawn independently from a truncated normal distribution $p(x_t^{(1)}|y_t = k)$ of the form $TN(u_k^{(1)}, \sigma_k^{(1)})$ and a log-normal distribution $p(x_t^{(2)}|y_t = k)$

of the form $LN(u_k^{(2)}, \sigma_k^{(2)})$, whose translation and scale parameters u and σ depend on the state $k = 0, 1$. These two distributions emulate the median heart rate feature and the accelerometer standard deviation feature used in Section 3.4.2.2 for the full time course HVC data.

For the unstable case, the parameters of the distributions of y_t and $\boldsymbol{x}_t$ were changed from session to session to reflect perturbations during the post-baseline period. Specifically, the truncated normal parameters (μ_{mk}, σ_{mk}) of the event durations $\Delta\tau_m$ were made to be session dependent by matching these parameters to the empirical distributions of the HMM-FLDA estimated duration for the m-th session, where the empirical distribution was constructed over the sub-population of symptomatic subjects in the HVC study. Furthermore, we introduced a time-varying post-baseline conditional mean $u_{mk}^{(i)} = E[x_t^{(i)}|y_t = k]$ of the i-th channel over the m-th session for the k-th event state. In particular, the post-baseline conditional mean was modeled as a quadratically varying function of m: $u_{mk}^{(i)} = u_{1k}^{(i)} - b_k^{(i)} (m - m_o)^2 / (1 - m_o)^2 + b_k^{(i)}$, for $m = 1, \ldots M,\ i = 1, 2,\ k \in \{0, 1\}$. Here m_o is the index of the session where the mean $u_{mk}^{(i)}$ reaches a positive or negative apex, and $b_k^{(i)} = u_{m_o j}^{(i)} - u_{1k}^{(i)}$ is the difference between the apex value $u_{m_o k}^{(i)}$ and the initial value $u_{mk}^{(i)}$. The pair $(m_o, b_k^{(i)})$ controls how the mean values of $(x_t^{(1)}, x_t^{(2)})$ change over sessions. In the simulation, m_o was randomly drawn from $\{5,\ 6,\ 7\}$ with equal probability, and we considered 2 sets of values for the $b_k^{(i)}$'s. To achieve a close facsimile to the experimental data analyzed in Section 3.4, we simulated $M = 11$ sessions, and the time between samples was set to $\delta_t = 1/6$ hr, i.e., the 10 minute aggregated sample period used in that section. Furthermore, all of the translation and scale parameters used in the simulation were matched to summary statistics obtained from the HMM-FLDA analysis described in Section 3.4.2.2. Several realizations of the simulated data are shown in Fig. 3.5 in the Appendix.

Table 3.1: Out of sample performance of the proposed adaptive transfer learning algorithm as compared with standard HMM operating on the original data (HMM) and operating on LOESS detrended data (dHMM).

Setting	Methods	Accuracy	F1	Cosine dis.	Onset diff.	Duration diff.
stable	HMM	**0.9981**	**0.9987**	**0.9968**	**0.0165**	**0.0307**
	dHMM	0.9981	0.9986	0.9967	0.0253	0.0373
	Proposed	0.9980	0.9986	0.9964	0.0168	0.0313
unstable++	HMM	0.8959	0.9209	0.8559	1.6704	2.9169
	dHMM	0.9127	0.9231	**0.8957**	2.4709	2.2653
	Proposed	**0.9356**	**0.9544**	0.8904	**1.0679**	**1.5756**
unstable+-	HMM	0.9369	0.9510	0.9187	**0.2763**	1.9750
	dHMM	0.9271	0.9370	0.9114	1.0170	2.0131
	Proposed	**0.9758**	**0.9833**	**0.9564**	0.3712	**0.3817**

Table 3.1 compares the empirical performance (a total of 1000 independent trials on 100 different realizations of the model) achieved by the proposed HMM-FLDA procedure to other procedures (HMM and dHMM) for the stable and unstable cases described above. Each of the stable and unstable cases is characterized by the value of the four coefficients $(b_0^{(1)}, b_1^{(1)}, b_0^{(2)}, b_1^{(2)})$ that define the trend in the post-baseline mean $u_{mk}^{(i)}$. For a stable subject there is no mean trend for either event class and all of these coefficients are zero. Two unstable cases are considered: a case called *unstable*++, where the trends for both event classes are concave with a randomly located peak, and a case called *unstable*+-, where for each event class the trends go in opposite directions, i.e., one trend is concave while the other is convex. The table shows the value of 5 performance criteria (averaged over 1000 independent trials) when the method is trained on a single realization of the two signals and tested on an independent out-of-sample realization drawn from the same distribution. The performance criteria are:

- `Accuracy`: the number of samples with correctly classified event states divided by the total number of samples,
- `F1`: the harmonic mean of the precision and recall classification criteria,
- `Cosine dis.`: the cosine distance between the vector of classified event states and the ground truth states,
- `Onset diff.` the absolute difference between estimated onset time of a sleep session and the nearest onset time of a true sleep period, averaged over sessions,
- `Duration diff.` the absolute difference between estimated duration of a sleep session and the duration of a true sleep period that overlaps with the predicted one, averaged over sessions. (If the predicted period doesn't overlap with any true period, the difference is taken as the length of the predicted period)

Table 3.1 compares the mean performance of the proposed method to a standard off-the-shelf two-state HMM state estimator and a variant of HMM, called dHMM, that uses locally estimated scatterplot smoothing (LOESS) [Cle79] to detrend each of the two data channels as a preprocessing step before training the HMM. The HMM and LOESS were implemented with R packages `depmixS4` [VS$^+$10] and `fANCOVA`. The results shown in Table 3.1 establish the benefit of our proposed HMM-FLDA adaptive procedure. In terms of event classification performance (Accuracy, F1, and Cosine distance), except for cosine distance in the unstable++ case, where the performances are statistically equivalent, the proposed method significantly outperforms the competing HMM and dHMM procedures for both unstable++ and unstable+- models (p-value $\ll$0.01 according to one-sided paired t-test). It is also worth noting that, in the unstable++ case,

LOESS detrending of dHMM improves on HMM across all criteria except for onset diff., which is the r.m.s. estimation error for sleep onset time. The Supplementary Appendix 3.6.3 provides additional details on this simulation.

Table 3.2: Parameter settings for the HVC data processing pipeline

Symbol	Value	Description	Loc. in pipeline
δ_t	10 minutes	epoch length	local feature extraction (pre-processing)
t_1	Hour 0	starting time of BL segment	Baseline HMM training
t_{N_0}	Hour 36	end time of BL segment	Baseline HMM training
Δ_t	3 hours	test window length	Domain adaptive FLDA training
d_l	$\{12, 13, \ldots, 59, 60\}$ hours	training window lengths	Domain adaptive FLDA training
γ	1	prior odds ratio	Domain adaptive FLDA training

3.4 Application to HVC experimental data

We apply the proposed adaptive transfer learning method to sleep/wake event classification for an experimental dataset undergoing a perturbation after baseline. This dataset of wearable data was collected as part of a human viral challenge (HVC) study where data from a cohort of participants was collected before and after exposure to a viral pathogen. More details on the HVC study are provided in Appendix 3.6.1. We will show that the proposed FLDA-HMM algorithm, trained individually on each participant without clinical outcome data, is capable of segmenting sleep events and extracting features that can be used to accurately predict clinical outcomes, specifically whether a subject is infected or not, and the onset time of infection.

Sleep has repeatedly been found to be associated with immune, cardiovascular, and neurocognitive function [Mil15, Irw15], among other functions. Many studies have revealed that changes in sleep pattern can be an important modulator of human response to diseases. For example, in a human viral challenge (HVC) study [DRR+00], researchers found that nasal inoculation with rhinovirus type 23 significantly reduced total sleep time among symptomatic individuals during the initial active phase of the illness. In another study [CDA+09], shorter sleep duration in the weeks preceding an exposure to a rhinovirus was found associated with lower resistance to illness. Since physiological signals such as instantaneous heart rate, physical activity, and skin temperature differ substantially during wake and sleep periods, automated sleep/wake labeling is possible. Therefore, the development of effective sleep monitoring methods has been of increasing interest.

While polysomnography (PSG) is the gold standard for sleep monitoring in sleep-related studies [KHH+17, DRS18], it is often cumbersome outside of a lab setting as it uses electroencephalogram (EEG), electrocardiogram (ECG), and electromyogram (EMG), and requires a registered technician to perform sleep scoring [RK]. Recently much effort has been made to remedy the inconvenience

of manual scoring and, alternatively, to provide automated analysis of the PSG signals [FDSR02, MYK+13, KDM+18]. Thus there has been growing interest in low-cost alternatives to PSG using small package wearable devices. Signals captured by wearable sensors have included movement induced accelerometer signals, circulating blood volume pulse (BVP), heart rate (HR), electrodermal activity (EDA), temperature, respiration effort (RSP), ambient light and sound. These devices can be wrist-worn, ankle-worn, arm-worn, lapel-worn, chest-worn or embedded in mobile phones [KMF08],[SCS+17], [DRS18]. The Empatica E4 device (Empatica Inc. USA), used in the HVC study discussed below, captures 4 such signals.

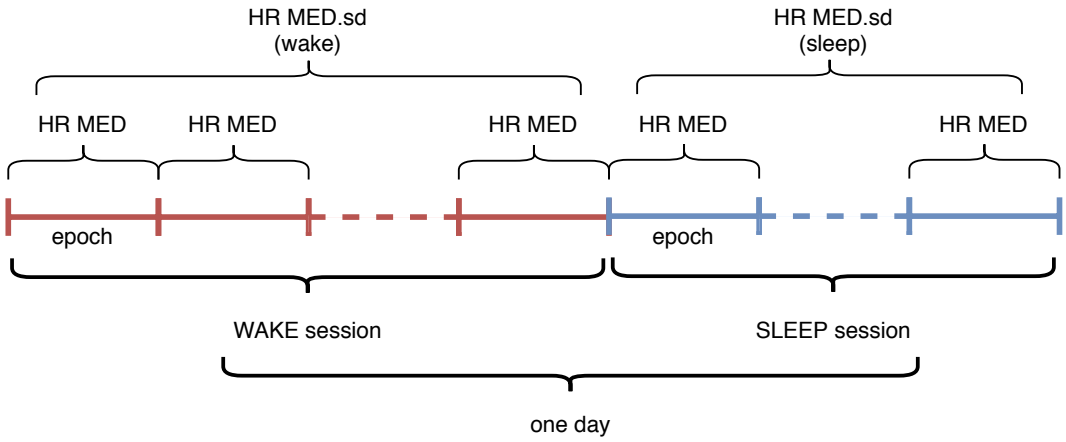


Figure 3.2: Labeling of events and event sessions from wearable data, illustrated for one of the features (HR MED = heart rate median) and one of the meta-features (HR MED.sd = standard deviation of HR MED) for Empatica E4 device in the Human Viral Challenge (HVC) study. In the first stage of the pipeline in Fig. 3.1b, epochs of 10 min duration (δ_t in Table 3.2) are used as windows over which temporally localized statistical features (mean, median, standard deviation) of the E4 device signals are extracted. These features are used in the second stage of the pipeline that extracts event labels, wake and sleep in the case of the HVC, and organizes them into contiguous labeled time segments, called sessions. In the third stage of the pipeline event-labeled meta-features (HR MED.sd (wake) and HR MED.sd (sleep)) are extracted as statistical summaries over the wake and sleep sessions.

3.4.1 HVC data and processing pipeline

Data was collected from E4 wristband devices worn by 25 participants enrolled in a longitudinal human viral challenge (HVC) study over the time period: 0 hours to 270 hours (11 days). The study participants were exposed to a pathogen (Inf H1N2) on the second day of the study at 36 hours (noon on day 2). Over subsequent days some participants developed symptomatic infection, confirmed by viral shedding from PCR assay on nasal lavage once per day. Such tests are often used to detect or confirm viral infection and to identify asymptomatic spreaders

[PCC+03, HZR+11, ZCV+09b, HLW+20]. In the HVC study, shedding in infected subjects was first detected either on day 4 at 81 hours, designating early onset shedders, or on day 5 at 106 hours, designating late onset shedders.

Four channels of physiological signals, including 3-axis accelerometry (ACC sampled at 32 sa/sec), heart rate (HR at 1 sa/sec), skin temperature (TEMP at 4 sa/sec) and electrodermal activity (EDA at 4 sa/sec), were measured by the E4 devices. See Fig. 3.4a in Appendix 3.6.2 for representative signals from an infected and a non-infected subject.

We implemented the three-stage data processing pipeline in Fig. 3.1b for sleep segmentation and high dimensional feature extraction from the HVC data. The first stage (top branch in Fig. 3.1b) consists of a pre-processing module, performing data conditioning, fine grain temporally-localized feature extraction and abnormality filtering. The second stage (middle branch) implements the HMM-FLDA module, performing adaptive sleep segmentation on non-abnormal data. The third stage (bottom branch) is a post-processing module, aggregating the data into sleep and wake sessions and performing coarse grain session-localized feature extraction.

Table 3.2 summarizes the parameter settings we used to implement the pipeline for the HVC data. Figure 3.2 depicts how the pipeline constructs sleep/wake sessions and session-localized features from the fine grained features.

Table 3.3: Features (**196** in total) extracted from sleep and wake sessions for each day from subjects in the HVC study.

Name	Number	Description
Duration	2	total sleep, night sleep
Onset/offset	2	night sleep only
HR summary	9×2	3 (mean, median, s.d.) × 3 (mean, median, sd within session) × 2 (sleep, wake)
HR linear coef.	6×2	3 (mean, median, sd) × 2 (coef.0, coef.1) × 2 (sleep, wake)
HR quadratic coef.	9×2	3 (mean, median, sd) × 3 (coef.0, coef.1, coef.2) × 2 (sleep, wake)
TEMP	24×2	same as HR
ACC	24×2	same as HR
EDA	24×2	same as HR

3.4.1.1 First stage: local features and abnormality filtering

Like other wearable sensors, the Empatica E4 captures data at high frequencies and, because of the fact that they are worn by subjects in non-laboratory situations, the raw data collected are often voluminous and noisy. To mitigate the impact of occasional poor readings and reduce computational burden, the pre-processing stage performs data conditioning, local feature extraction and abnormality filtering. First the missingness of a subject's available data is evaluated, resulting in

rejection of any subject with more than 40% missing time points. For each of the remaining subjects, their data is segmented into non-overlapping time intervals, which we call *epochs*, of equal length δ_t secs. The epoch length was set to $\delta_t = 10$ mins to achieve a tradeoff between oversmoothing the sleep/wake transitions (excessive length) and noise sensitivity (insufficient length). See Appendix 3.6.4 for additional discussion. Shorter epoch lengths may be more appropriate for classifying other types of event states, e.g., for capturing quality of sleep or detecting transitions between different sleep stages. For each of the 4 E4 signals, the module in Fig. 3.1b labeled "Local feature extraction" computes statistical features over each epoch corresponding to the signal mean (MEAN), signal median (MED) and signal standard deviation (SD). If a particular epoch has fewer than 90% available samples, e.g., due to the E4 not being worn, the epoch is discarded.

These local features are then processed by the "Abnormality filtering" module to identify time points at which outliers occur, labeled as non-normal data in Fig. 3.1b. Similarly to methods used in network intrusion detection [BBK13], the module uses marginal k-means clustering and quantile filtering to label time points as outliers (See Appendix 3.6.4.1). Subjects having less than 60% normal data cannot be reliably segmented and are omitted from subsequent analysis.

The last pre-processing step consists of selecting a subset of the 12 local features to train the HMM-FLDA sleep segmentation procedure. For this we select putative wake and sleep time intervals in the early evening around 21:00 and in the twilight hours around 4:00. These time periods were chosen since it is expected that most people would be awake at 9pm and sleeping at 4am. The sleep/wake discrimination capability of each feature over these two intervals was measured by a sleep/wake separability index (SWSI) computed over all non-abnormal subjects and over all available days. The SWSI is defined similar to the geometric separability index (3.5):

$$\text{SWSI}(k) = \frac{\sum_{t \in \mathcal{T}_{wake} \cup \mathcal{T}_{sleep}} \{\hat{y}_t + \tilde{y}_t(k) + 1\} \textbf{ MOD } 2}{|\mathcal{T}_{wake}| + |\mathcal{T}_{sleep}|}, \tag{3.6}$$

where $k = 1, \ldots, p$ indexes the features, $\mathcal{T}_{wake}$ and $\mathcal{T}_{sleep}$ are respectively the time intervals around 21:00 and 4:00 over all available days of data, $\hat{y}_t$ is the putative label assigned to the k-th feature $x_t(k)$, defined as $\hat{y}_t = 1$ if $t \in \mathcal{T}_{sleep}$ and $\hat{y}_t = 0$ otherwise, and $\tilde{y}_t(k)$ is the label of the nearest neighbor of $x_t(k)$ among $\{x_\tau(k) : \tau \in \mathcal{T}_{wake} \cup \mathcal{T}_{sleep}\}$. The criterion for including the k-th feature among those that are used to train the HMM-FLDA is that SWSI(k) be greater than 0.7 for at least 75% of the subjects.

3.4.1.2 Second stage: adaptive sleep detection

Adaptive sleep segmentation was implemented using the HMM-FLDA procedure described in Section 3.2. The initial training window (baseline) for training the HMM-FLDA (Algorithm 1) was defined as the period from 0 to 36 hours, the time of viral inoculation. The test window length was

chosen to be $\Delta_t = 3$ hours, which is based on our expectations about the time scale of immune response induced changes in the E4 signal distributions subsequent to pathogen exposure. The most rapid time scales of immune response are on the order of a few hours, e.g., inflammatory monocyte recruitment [HHMH03]. Such a choice of Δt will minimize the likelihood that the HMM-FLDA algorithm loses track of covariate shifts from batch to batch. The range of training window lengths $d_1, \ldots, d_L$ used by the optimization loop in Algorithm 1 was restricted to $d_1 = 12$ to $d_L = 60$ hours to guarantee a good balance of wake and sleep samples over the window.

After the HMM-FLDA procedure labels the normal data, an "Abnormality classifier" module identifies any abnormal samples (identified in Stage 1) that is physiologically meaningful and reinserts it into the data stream, labeling it with the label of the sample immediately preceding it (See Fig. 3.12 in Appendix 3.6.4.1). Finally, a modified median filter with a 90-min smoothing window is applied to remove short bursty sleep periods of less than 60 mins. Such short sleep periods are likely to lack deep sleep stages 3 and 4, which typically start 30 mins. after sleep onset and can last 20 to 40 mins. [CD$^+$05].

3.4.1.3 Third stage: session-level feature set

In the third stage of the pipeline, the sleep/wake sessions produced by the second stage are aligned to particular calendar days. A full day session is composed of wake sessions and sleep sessions. Any sleep session overlapping with a given day whose onset occurs before 5am (5:00) is associated with the previous day session. The binary sleep/wake label is then appended to the local features generated in the the first stage of the pipeline, doubling the total number of local features. Then a set of session-level features are extracted by computing several statistical summaries over daily wake and sleep sessions. This results in 196 session-level features that include the mean, median, and standard deviation (sd) of the event labeled local features over a sleep or wake session. All features related to standard deviation are log-transformed. The session-level features also include the coefficients of linear and quadratic fits to the time course of these features over a sleep or wake session. In addition session timing features, such as, sleep duration, sleep onset, and sleep offset are included in the session-level feature set. See Table 3.3.

3.4.2 Online sleep segmentation

We emulate an online implementation of the pipeline in Fig. 3.1b by reapplying it successively to all available data at the end of each day. Daily updating of the segmentation corresponds to the real-world scenario where data is uploaded from the E4 device once per day, corresponding to the data acquisition rate in the HVC study. Continuous (real-time) updating will only become practical when devices have sufficient power for continuous data transfer or onboard processing. Below we

Table 3.4: Top 4 E4 wake/sleep features for predicting clinical outcome within 24 hours (later than day 3) of inoculation day (day 2) based on device data collected over the range 0-60 hours (day 1 and 2).

logistic regression model			*continuation-ratio regression model*				
Feature	Coef.	AUC	Feature	Coef.	AUC		
					Early	Late	No onset
HR MED.sd (sleep)	-3.921	0.758	HR MED.sd (sleep)	-5.073	0.882	0.718	0.864
ACC SD.linear.coef1 (wake)	-14.468	0.737	HR MEAN.sd (sleep)	-4.466	0.706	0.628	0.773
HR MED.quad.coef2 (sleep)	4.318	0.707	Total duration	-0.599	0.676	0.551	0.750
Offset	-1.452	0.697	Night duration	-0.616	0.647	0.500	0.705

illustrate the online implementation for a case of pre-infection segmentation before shedding occurs, and a full time course segmentation at the end of the study. For additional details see Appendix 3.6.5.

3.4.2.1 Illustration: pre-infection segmentation

Here only the first 60 hours (up to day 3 at 12pm) of the data are available to the pipeline. The first 60 hours include the inoculation time (36 hours) and the first two nights of sleep. Note that no infection (shedding) is detected before day 4 at 81 hours. The preprocessing stage of the pipeline removed 5 subjects with excessive missing or abnormal data: 2 of these subjects had more than 40% of their time points missing and 3 subjects had more than 40% abnormal time points. For the remaining 20 subjects local feature selection was accomplished by applying SWSI (3.6) to all 12 variables to contrast early evening periods (19:00-22:00) to twilight periods (2:00-5:00). Only three features were found to have SWSI above 0.7 for at least 75% of the subjects: the mean and median heart rate (HR MEAN and HR MEDIAN) and the standard deviation of the magnitude accelerometer (ACC SD). As the mean and median heart rate are highly correlated and the SWSI of HR MEAN has lower 25% quantile than does the HR MEDIAN, only the HR MEDIAN and ACC SD were selected. Three hour long sleep and wake periods were necessary in order to obtain a sufficient number of representative time samples since only two days and nights are available.

3.4.2.2 Illustration: full time course segmentation

The full time course (0-270 hours) of a subject's data was made available to the pipeline. No subjects had more than 40% missing time points over this period. Stage 1 abnormality filtering resulted in elimination of 5 subjects as excessively abnormal, who were removed. To compute SWSI for variable selection on each of the remaining 20 subjects, we used shorter putative wake periods (21:00-22:00) and sleep periods (4:00-5:00) since there are many more available days and

nights than in the 60 hour pre-infection time period. Applying the same SWSI variable selection criterion three features were selected to train the HMM-FLDA: HR MED, HR SD, and ACC SD.

3.4.2.3 Clinical outcome prediction

We used the feature set generated by the pre-infection segmentation to perform early detection of infection. Two clinical outcomes were considered: 1) a binary infection state (whether or not the subject will shed virus anytime after 60 hours) designating the subject as infected or non-infected; and 2) the ternary infection state corresponding to onset time of viral shedding. The three onset times are defined as: early onset shedding (first detected by PCR at 81 hours), late onset shedding (first detected by PCR at 106 hours), and no onset shedding (subject never sheds virus). Among the 20 subjects whose 0-60 hour data passed the abnormality filtering test, 9 of them were infected and 11 were non-infected. Among those infected, 3 were early shedders and 6 were late shedders.

Two predictor models were used for each of these outcomes, which were trained only on the sleep/wake features collected on inoculation day (wake/sleep sessions for day 2). For prediction of whether a subject will shed or not we applied logistic regression (LR) [FHT01] and for prediction of onset time we used continuation ratio (CR) regression [Agr10], which can be interpreted as a discrete version of the Cox regression model [HvdAS01].

For both the logistic and the continuation-ratio models the classification capability of individual features is evaluated by leave-one-out cross-validation. The top 4 discriminative features are reported in Table 3.4, together with their corresponding regression coefficients and a detection performance criterion: the area under the ROC curve (AUC). AUC is a measure of the accuracy of a binary classifier and a value of AUC near 1 means very high accuracy. For CR regression, since there exist three classes, three different AUCs are calculated using the "one-versus-the-rest" strategy [Bis06]. The CR performance is rank ordered in decreasing order of the minimum of the AUC's for Early, Late and No onset. See Appendix 3.6.5 for boxplots of these top 4 features.

The top feature for both LR and CR models is a sleep feature: the standard deviation of the median heart rate (HR MED.sd (sleep)). Interestingly, most of the 4 top ranked features come from sleep sessions, with the exception of the linear coefficient to a linear fit to the time course of the standard deviation of ACC, denoted ACC SD.linear.coef1 (wake). Features related to sleep heart rate variation (HR MED.sd, HR MEAN.sd) and sleep duration (total duration, offset, night duration) are most discriminative.

The results of running full timecourse sleep/wake segmentation reveals striking temporal differences between shedders and non-shedder features. Figure 3.3 indicates qualitatively different total sleep duration and sleep offset behaviors in these two groups over the full 11 day duration of the study. The non-shedder group sleeps on the average 2.041 hours longer than does the shedder group (p-value < 0.005 using two-sided t-test) over this time period. This is consistent with studies

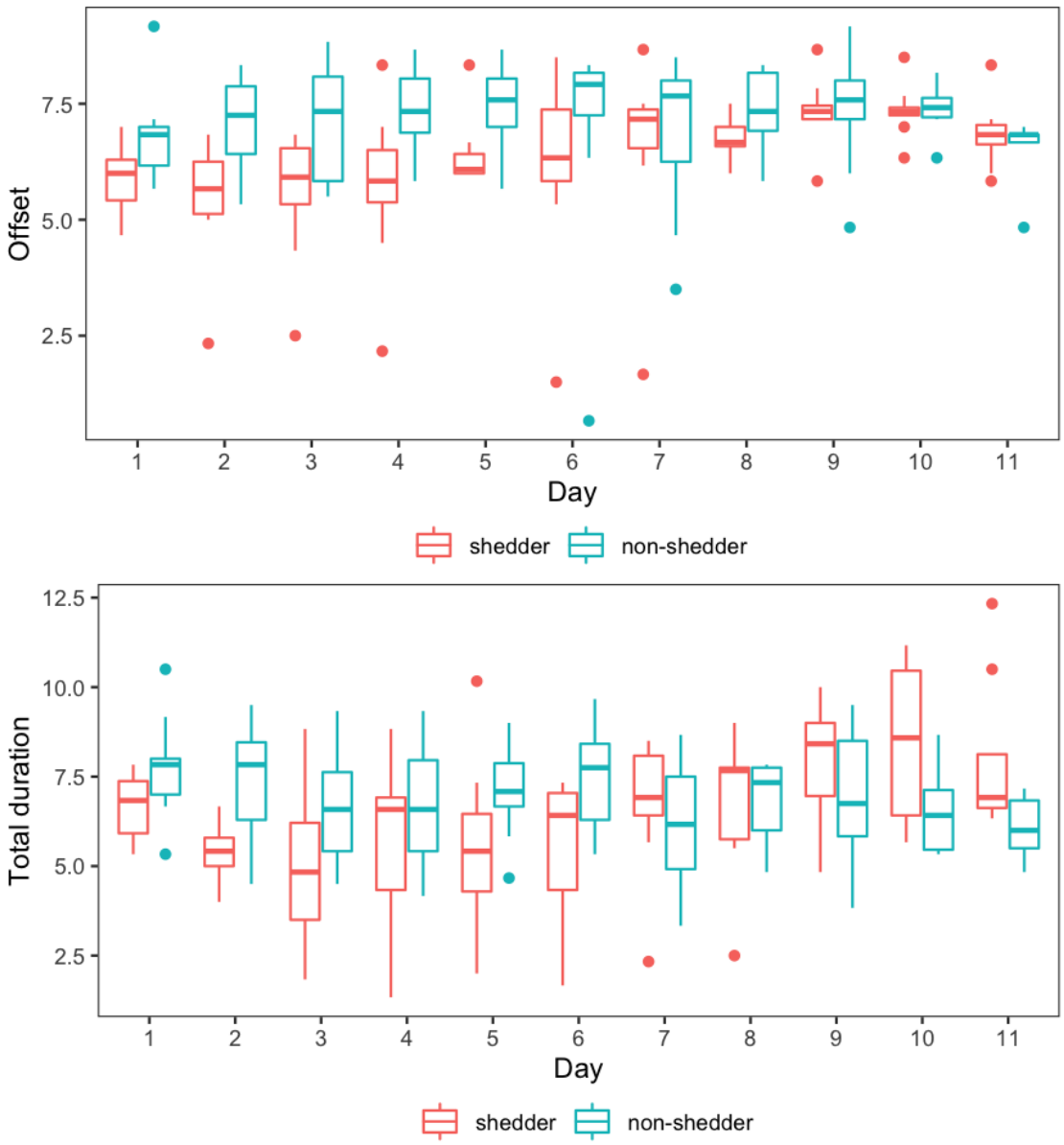


Figure 3.3: Offset (wake-up time) and total sleep duration (night and day sleeping) features extracted by the proposed HMM-FLDA pipeline applied to the experimental human viral challenge study (HVC) when all time points (0-270 hours) are available. Viral inoculation took place on day 2 and all infected people (shedders) started shedding virus on day 4 or day 5.

of the effects of sleep on the course of respiratory infection [DRR+00]. The mean sleep duration deficit among the shedders gradually decreases over time, and the trend reverses after day 6. The offset feature illustrates that non-shedders tend to wake-up later than the shedders until day 9 of the study.

3.5 Discussion and Conclusions

We have developed an unsupervised adaptive algorithm for classification of latent event states from multivariate physiological data collected from a wearable device. The algorithm adapts to perturbations of the initial training distributions using a sequential transfer learning model to mitigate covariate shift. The proposed algorithm operates without requiring *a priori* information about true sleep/wake states and is capable of automatically detecting anomalies and abnormal data records. Numerical simulations established significant advantages of our model relative to hidden Markov approaches to hidden event classification.

The results presented in this paper are not without limitations. The HVC experiment has the limitation of small sample size and the classes are imbalanced. The negative effect of class imbalance can be compensated, to an extent, using methods such as the synthetic minority over-sampling technique (SMOTE) [CBHK02]. While no substitute for increasing actual sample size, we have demonstrated that SMOTE can improve clinical outcome prediction accuracy as measured by AUC (Table 3.11 in Appendix E). Another limitation is that the HVC experiment lacks ground truth information about the true sleep/wake states of the subjects. The simulation study we presented emulating a similar HVC experiment with ground truth is an in-silico validation but a controlled experiment, e.g., performed in a sleep lab, would provide better confirmation. It would also be worthwhile to test the algorithm in a larger scale experiment that collects self-reported sleep diaries in addition to clinical data.

There are also limitations of the proposed HMM-FLDA adaptive event segmentation algorithm. First, the algorithm may fail if there is an abrupt and excessively large shift in the event class distributions from time to time. More generally, loss of track due to abrupt changes is a limitation of the gradual self-training approach [KML20],[CWB11] commonly adopted in transfer learning. In extreme cases this limitation may be insurmountable as there are fundamental theoretical limits that limit tracking ability of any adaptive algorithm [HL94]. If the abrupt shift in distributions persists over time, a possible remedy would be to episodically re-initialize the HMM-FLDA algorithm during the adaptation phase. Secondly, the classification accuracy of FLDA may be poor if the event classes are not linearly separable. At the possible cost of reduced simplicity of implementation, use of a non-linear classifier in place of FLDA would overcome this limitation, e.g., using kernelized FLDA or a Support Vector Machine (SVM) [FHT01].

We conclude by pointing out that the proposed framework for adaptive multi-channel event segmentation and feature extraction easily generalizes beyond the setting of the binary sleep/wake segmentation illustrated in this paper. Monitoring different stages of sleep or different wake activity types would be a natural non-binary extension. With the continuing advances in the capabilities of wearable devices for digital health, many new applications will be enabled by continuous multi-event

tracking. These could include accurate behavioral and health assessment tools that will advance personalized health care.

The R code and HVC data used for this paper has been made publicly available at GitLab (`gitlab.eecs.umich.edu/yayazhai/shezhai_bme2020`).

3.6 Supplementary Materials

3.6.1 Human Viral Challenge (HVC) Study

A human viral challenge (HVC) study was conducted in 2018 as a collaborative effort between Duke University and University College London under a grant from the Defense Advanced Research Projects Agency (DARPA) under the PROMETHEUS program. Thirty nine healthy volunteers between the ages of 18 and 55 were enrolled as participants in the study, which took place in the United Kingdom.

The HVC was divided into outpatient and confinement phases. During the confinement phase participants stayed overnight for a period of 8-11 days in total from the morning before the day the viral challenge was administered to the end of confinement. During the outpatient phase subjects were evaluated for health conditions by tests including ear, nose and throat (ENT), respiratory and cardiac assessment. In addition to other data types not relevant to this paper, wearable device data (Empatica E4) and clinical infection status data (viral shedding) were collected from the participants. All data was anonymized prior to transfer to Duke and Michigan for the analysis described in this paper.

Participant exclusion criteria: Chronic respiratory disease (asthma, COPD, rhinitis, sinusitis) in adulthood. Inhaled bronchodilator or steroid use within the last 12 months. Use of any medication or other product (prescription or over-the-counter) for symptoms of rhinitis or nasal congestion within the last 3 months. Acute upper respiratory infection (URI or sinusitis) in the past 6 weeks. Smoking in the past 6 months or $>$5 pack-year lifetime history. Subjects with allergic symptoms present at baseline. Clinically relevant abnormality on chest X-ray. Any ECG abnormality. Those in close domestic contact (i.e. sharing a household with, caring for, or daily face to face contact) with children under 3 years, the elderly (>65 years), immunosuppressed persons, or those with chronic respiratory disease. Subjects with known or suspected immune deficiency. Receipt of systemic glucocorticoids (in a dose $\geq$ 5 mg prednisone daily or equivalent) within one month, or any other cytotoxic or immunosuppressive drug within 6 months prior to challenge. Known IgA deficiency, immotile cilia syndrome, or Kartagener's syndrome. History of frequent nose bleeds. Any significant medical condition or prescribed drug deemed by the study doctor to make the participant unsuitable for the study. Pregnant or breastfeeding women. Positive urine drug screen. Detectable baseline antibody titres against influenza challenge strains. History of hypersensitivity to eggs, egg proteins, gentamicin, gelatin or arginine, or with life-threatening reactions to previous influenza vaccinations.

Confinement phase study: The eight days of the confinement phase consisted of a 36 hour healthy reference time period (baseline), inoculation at 36 hours (exposure), and a post-baseline time period. The E4 data of the participants was collected over the entire period and viral shedding was measured

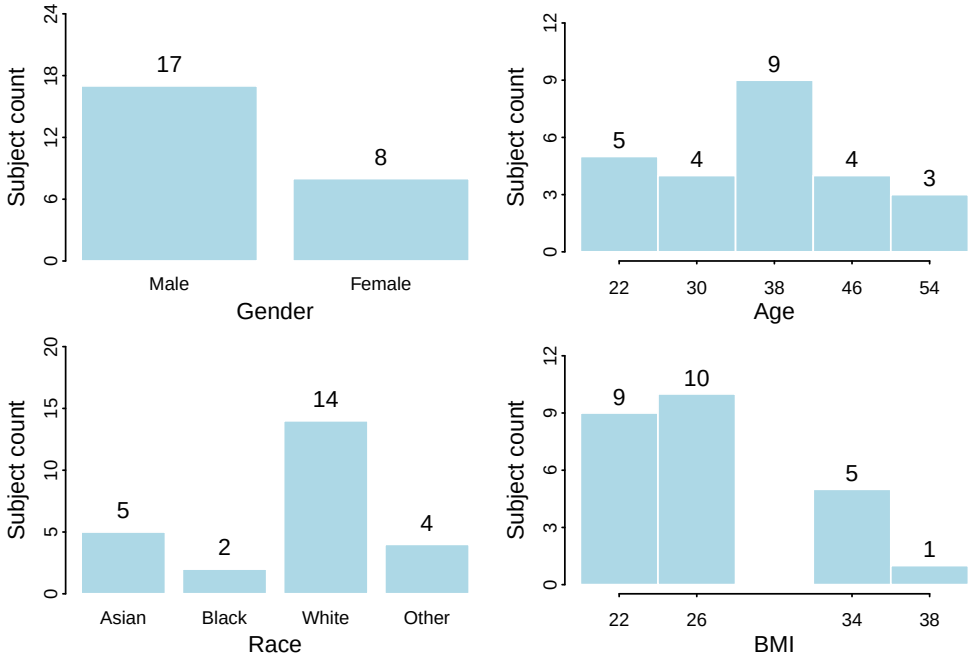
once per day in the morning. On the second day (day 2) at approximately noon (36 hours from start of study) each participant was challenged by a GMP influenza A/California/04/2009-like (H1N1) virus strain. The inoculation was administered by inserting intra-nasal drops on a single occasion with diluted inoculum with an average dose of 10^6 TCID50 in 1mL PBS divided equally between the two nostrils. This resulted in an average attack rate of 44%. Following inoculation, advice regarding hand hygiene was given and subjects were provided with alcohol hand gel and face-masks if they moved between the inoculation room and the quarantine ward.

Viral shedding assay: Over the confinment phase of the study, viral shedding was measured once per day through a nasal lavage. The collected fluid was aliquoted into sterile microfuge tubes and centrifuged for analysis of cells, and lavage fluid was later analysed by singleplex PCR to quantify the degree of viral shedding of the inoculated strain. Multiplex PCR was performed on the pre-inoculation lavage and post-inoculation lavage collected during the study to exclude the presence of other respiratory viruses.

Wearable device and protocol: Over the confinement phase of the study, subjects agreed to comply with the following wearable device protocol: 1) they wear the Empatica E4 device properly, i.e., comfortably tight on the wrist of their dominant hand; 2) they take care to maintain the device and protect it from shocks, water immersion, and other damage; 3) they wear the Empatica continuously without interruption except for periods that they were showering, recharging, or uploading data. The E4 has several sensors that measure physiological parameters including blood volume pulse, skin conductance, temperature, and movement. The E4 was recharged once per day during which time each participant's data were uploaded to a cloud server for processing using Empatica proprietary software.

The result of this processing was reported at sub-second temporal resolution as the following variables: heart rate (1sa/sec), skin temperature (4sa/sec), electrodermal activity (4sa/sec), and 3 axis accelerometer (32sa/sec). These were mapped to a vector of four variables at each time point: HR, TEMP, EDA and ACC, respectively, where ACC was computed as the Euclidean norm (magnitude) of the 3 dimensional acceleration vector. Subjects were trained on best practices for wearing and maintaining the E4 devices over the course of the confinement phase of the study.

Of the 39 participants enrolled in the HVC, only 25 had sufficient quality E4 data to be included in the analysis presented in Section 3.4. A histogram of the demographic data of these 25 participants is shown in the figure below. The other participants had wearable data that suffered from factors such as device failure, excessive missingness or data corruption, making their data unusable for our analysis. Non-compliance with wearable device protocols, data upload errors, and device malfunction were the cause of most of these problems.



Demographics of 25 participants in HVC study.

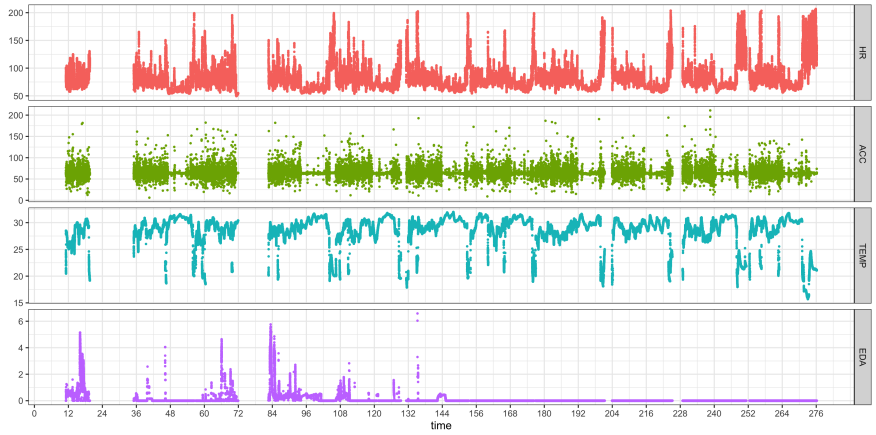
3.6.2 HVC data

The raw time course E4 data of two representative subjects in the HVC study are shown in Fig. 3.4. Subject 1 (Fig. 3.4a) did not become infected after inoculation on the second day of the study (day 2), i.e., this subject had no detectable level of viral shedding at any time over the 8-11 days of the study, while Subject 2 (Fig. 3.4b) became infected. The four E4 signals shown in the figure are heart rate (HR), accelerometer (ACC), temperature (TEMP), and electrodermal activity (EDA). The data clearly shows diurnal differences in signal behavior, corresponding to the cycling of sleep and wake states of these subjects. A quadratic trend in heart rate and temperature is clearly visible in Subject 2, the trend peaking at around 144 hours. Robustness to this trend is desirable and motivated the proposed HMM-FLDA sequential adaptive sleep/wake segmentation algorithm.

3.6.3 Simulation study

In this section we include additional figures supporting the simulation study reported in Table 3.1 and described in Section 3.3 of the main text. We also report on in-sample performance comparisons between the proposed HMM-FLDA and the competing HMM and dHMM algorithms.

As mentioned in Sec. III these simulations were intended to emulate the experimental HVC data used in Sec IV. In Fig. 3.5 we show realizations of two simulated device channels, $X1$ and $X2$, emulating the HR MED and ACC SD features in the HVC study, under the unstable++, and unstable+- models for mean trends occurring after 36 hours. See Figs. 3.4a and 3.4b for comparison



(a)

(b)

Figure 3.4: Four channel device data (Empatica E4) from human viral challenge (HVC) study for two subjects (Subjects 1 and 2). The four channels are: HR (red), ACC (green), TEMP (blue) and EDA (purple). Time 0hrs corresponds to 12am local time on the first day (day 1) of the study. Viral inoculations were administered to all subjects on the morning of the second day (day 2), i.e., between 32hrs and 36hrs. **(a)** data from a Subject 1 who had no detected shedding (Non-infected class); **(b)** data from Subject 2 for whom shedding was detected (Infected class).

to real E4 data.

Table 3.5 provides a more comprehensive report of the results of our simulations of the proposed HMM-FLDA event segmentation than does Table 3.1 in the main text. Three segmentation algorithms are compared: the proposed HMM-FLDA, the HMM and a detrended HMM (dHMM) using LOESS pre-filter to detrend prior to applying HMM. The simulation parameters for the random wake/sleep event transitions, event conditioned data distributions, and the covariate shift are the same as were used for Table 3.1.

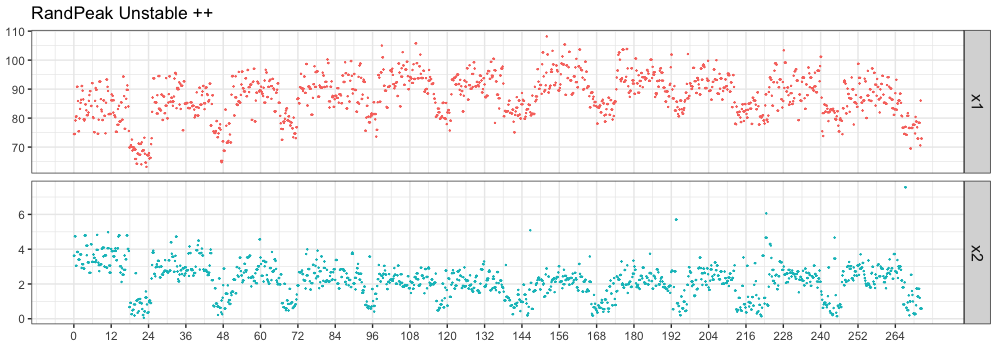
The top part of the table shows in-sample performance and the bottom part shows out-of-sample performance showing both mean and standard deviation (SD) of the performance measured over

100 simulation trials. Only the means of the OoS part of this table were shown in Table 3.1. The performance of the best performing method in terms of the MEAN performance is bolded.

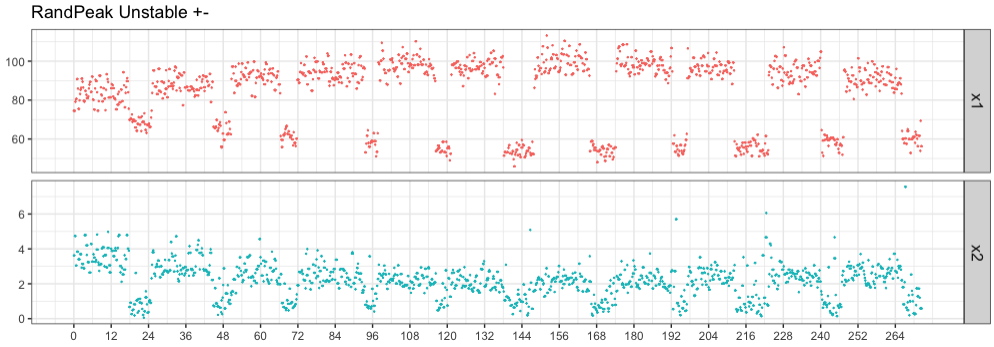
For the in-sample results (top part of table), event segmentation performance on the training set is shown. For the out-of-sample results (bottom part of table), an hold-out dataset was created from an independent simulation using the same model as the training set. The performance of each segmentation algorithm, trained only on the training set, was evaluated on the hold-out set. Note that the proposed HMM-FLDA would never be implemented in this way as such an implementation would turn off the adaptation mechanism on future data. The bottom part (OoS) of Table 3.5 provides evidence for the accuracy and robustness of the proposed approach.

The bolded entries in Table 3.5 denote the best mean performance per model (stable, unstable++, unstable+-) and performance criterion (Accuracy, F1, Cosine dist, Onset diff, and Duration diff). The asterisk on an MEAN entry in the OoS part of Table 3.5 indicates the best performing method for each of the five criteria where best was determined using a one-sided paired t-test of significance applied to the set of 1000 simulations thresholded at a 0.01 level of significance. The OoS part of Table 3.5 has an additional row, P-VALUE, that is the p-value of the one-sided paired t-test that the proposed HMM-FLDA has better performance than both the HMM and dHMM. Only two P-VALUE entries are not significant, the case unstable++ for cosine distance and the case unstable+- for onset diff. For the former case, dHMM is better than HMM but dHMM cannot be characterized as better than the proposed HMM-FLDA (0.01 level of significance). For the latter, HMM is better than the proposed for onset diff estimation. Hence, except for the onset diff criterion in the unstable+- case, the proposed method is equivalent or better than the other methods at a 0.01 level of significance.

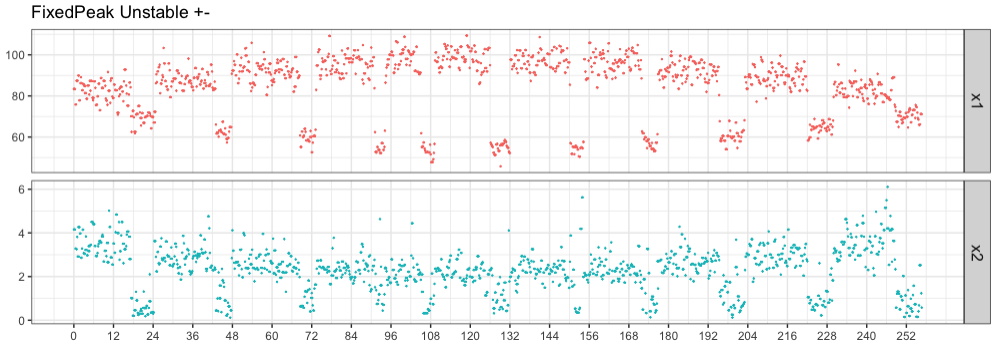
In Fig. 3.6 is shown boxplots and density plots of the out-of-sample (OoS) simulation of all three scenarios, whose mean is shown in Table 3.1. These plots show the distribution of errors committed by the various methods compared in terms of event labeling accuracy, onset estimation error, and duration estimation error when the trained methods are applied to independent sample trajectories drawn from the same distribution. As compared to the others the proposed method has a distribution that is more highly concentrated and has fewer outliers.



(a)



(b)

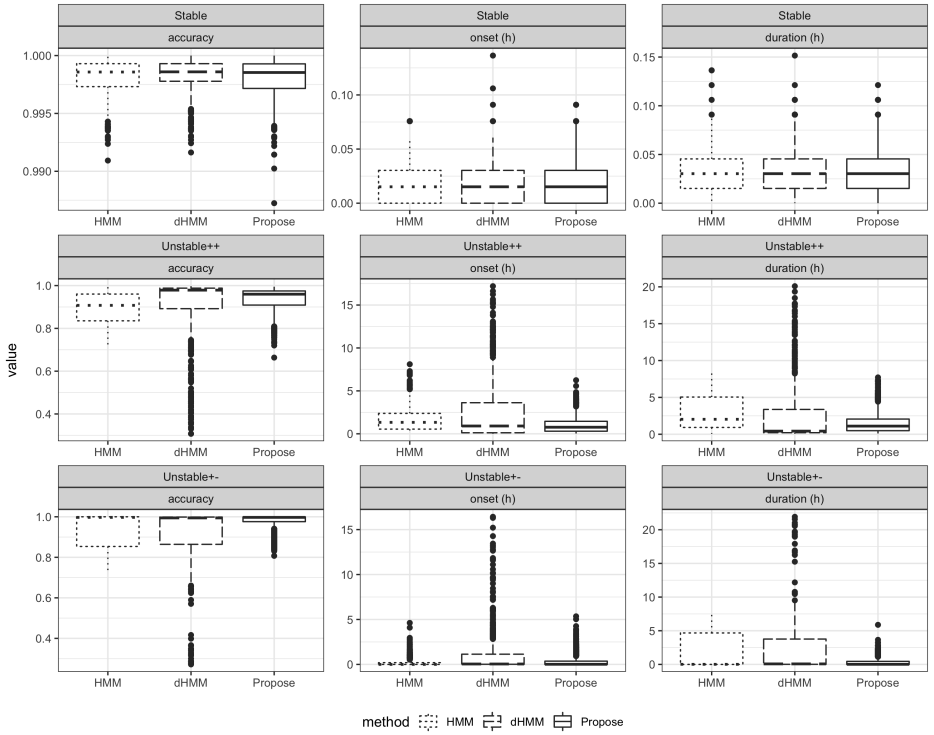


(c)

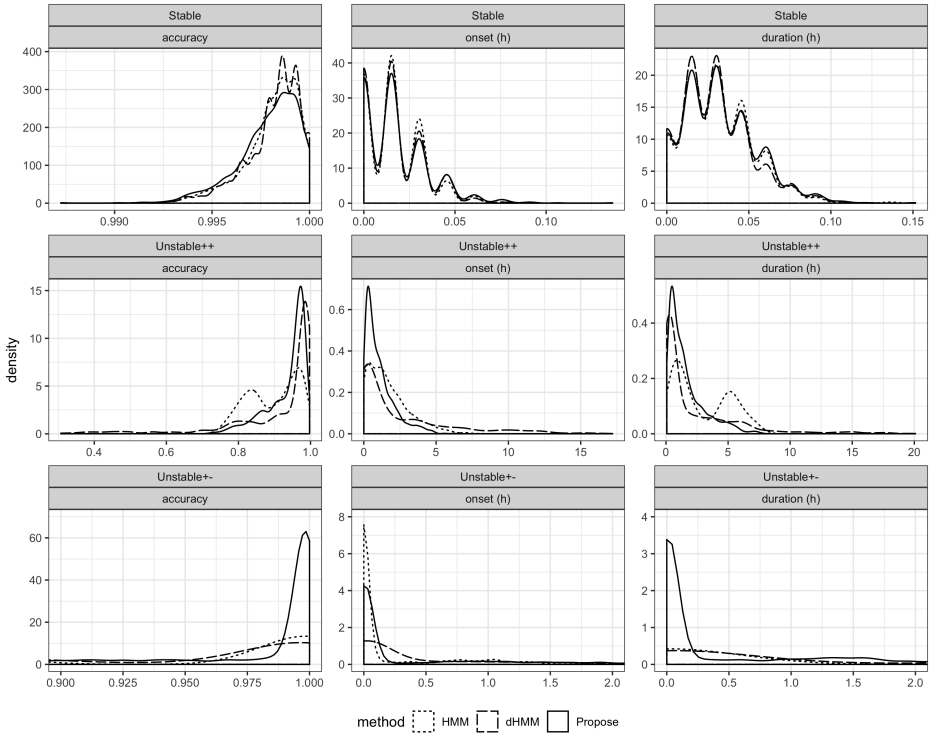
Figure 3.5: Realizations of simulated signals for: (**a**) Unstable++ case for which the sleep and wake session both have concave trends after 36 hours in channel $X1$; (**b**) Unstable+- case where the sleep and wake sessions have convex and concave trends, respectively, in channel $X1$; (**c**) another realization of Unstable+- case with sleep and wake sessions whose trends hit their apogee at a different time.

Table 3.5: Performance of the proposed adaptive transfer learning algorithm as compared with standard HMM operating on the original data (HMM) and operating on LOESS detrended data (dHMM).

Setting	Methods		Accuracy	F1	Cosine dis.	Onset diff.	Duration diff.
(0, 0, 0, 0)	HMM	MEAN	**0.9982**	**0.9987**	**0.9968**	**0.0236**	0.0427
stable		SD	**0.0012**	**0.0009**	**0.0021**	**0.0199**	0.0287
	dHMM	MEAN	0.9981	0.9986	0.9966	0.0271	0.0450
		SD	0.0012	0.0009	0.0022	0.0209	0.0288
	Proposed	MEAN	0.9981	0.9987	0.9967	0.0259	**0.0420**
		SD	0.0012	0.0009	0.0021	0.0218	**0.0278**
(15, 10, 0.5, -0.5)	HMM	MEAN	0.8744	0.9052	0.8234	2.4786	3.6780
unstable++		SD	0.0642	0.0520	0.0754	1.5541	2.2121
	dHMM	MEAN	**0.9558**	**0.9680**	**0.9306**	0.9912	**1.1574**
		SD	**0.0502**	**0.0375**	**0.0713**	1.6795	**1.4117**
	Proposed	MEAN	0.9487	0.9636	0.9146	**0.9813**	1.4605
		SD	0.0498	0.0363	0.0781	**0.9889**	1.3812
(-15, 15, 0.5, -0.5)	HMM	MEAN	0.9371	0.9513	0.9185	0.4329	2.0992
unstable+-		SD	0.0836	0.0660	0.1043	0.7836	2.7722
	dHMM	MEAN	0.9483	0.9615	0.9253	0.4909	1.4544
		SD	0.0666	0.0505	0.0928	0.9295	1.8854
	Proposed	MEAN	**0.9923**	**0.9946**	**0.9868**	**0.1631**	**0.1699**
		SD	**0.0221**	**0.0157**	**0.0365**	**0.4648**	**0.4651**
stable	HMM (OoS)	MEAN	**0.9981**	**0.9987**	**0.9968**	**0.0165**	**0.0307**
		SD	**0.0005**	**0.0004**	**0.0009**	**0.0051**	**0.0079**
	dHMM (OoS)	MEAN	0.9981	0.9986	0.9967	0.0253	0.0373
		SD	0.0015	0.0011	0.0027	0.0163	0.0211
	Proposed (OoS)	MEAN	0.9980	0.9986	0.9964	0.0168	0.0313
		SD	0.0015	0.0011	0.0027	0.0163	0.0211
		P-VALUE	1.0000	1.0000	1.0000	0.7850	0.8735
unstable++	HMM (OoS)	MEAN	0.8959	0.9209	0.8559	1.6704	2.9169
		SD	0.0679	0.0547	0.0792	1.0972	2.2120
	dHMM (OoS)	MEAN	0.9127	0.9231	**0.8957***	2.4709	2.2653
		SD	0.1284	0.1377	**0.1089**	2.6125	3.1967
	Proposed (OoS)	MEAN	**0.9356***	**0.9544***	0.8904*	**1.0679***	**1.5756***
		SD	**0.0353**	**0.0263**	0.0533	**0.7614**	**1.0664**
		P-VALUE	**9.94E-07**	**1.07E-10**	0.9521	**4.04E-28**	**3.73E-08**
unstable+-	HMM (OoS)	MEAN	0.9369	0.9510	0.9187	**0.2763***	1.9750
		SD	0.0810	0.0635	0.1019	**0.4079**	2.5778
	dHMM (OoS)	MEAN	0.9271	0.9370	0.9114	1.0170	2.0131
		SD	0.1151	0.1298	0.1112	2.1413	3.2464
	Proposed (OoS)	MEAN	**0.9758***	**0.9833***	**0.9564***	0.3712	**0.3817***
		SD	**0.0326**	**0.0226**	**0.0591**	0.6225	**0.5472**
		P-VALUE	**1.63E-31**	**9.69E-37**	**5.80E-16**	0.9999	**1.34E-61**



(a)



(b)

Figure 3.6: (**a**) Boxplots for label accuracy, onset estimation error, duration estimation error for simulations shown in Table 3.1 for out-of-sample (OoS) simulation performance (excluding two outliers by dHMM method in the stable case). (**b**) corresponding interpolated densities associated with (**a**) (for unstable+- case, the density plots are shown for a zoomed in region to better compare density concentration).

Table 3.6 shows runtime comparisons. The proposed method runs slower than HMM and dHMM. However, 1 minute run time of HMM-FLDA is still small as compared to the 8-11 days time period of HVC data. Furthermore, the method is not optimized in terms of runtime and can likely be accelerated.

Table 3.6: Average run time (seconds) per replication of the proposed adaptive transfer learning algorithm as compared with standard HMM operating on the original data (HMM) and operating on LOESS detrended data (dHMM) on a 2.3 GHz Dual-Core Intel Core i5 processor with 16 GB RAM.

	stable	unstable++	unstable+-
HMM	0.4347	0.5612	0.4913
dHMM	0.7396	0.8351	0.8347
Proposed	53.5126	54.3654	55.2601

3.6.4 Analysis pipeline implementation for HVC study

Here we provide additional details on our implementation of the pipeline of Fig. 3.1b, in the context of the Human Viral Challenge Study Empatica E4 data. We illustrate the implementation for both the pre-infection time-line (0-60 hours) and the full time-line (0-270 hours). First we describe the features used for abnormality detection.

Wearable devices are subject to outliers, anomalies and other abnormal sensor readings. Some types of outliers are physiological and are important to include in the final event-segmented data stream. Other types of outliers are technical and can be due to device malfunction or improper wearing of the device. These technical outliers must be removed early in the analysis pipeline so as to not compromise downstream event detection and labeling performance. We use a two stage procedure for isolating such anomalies in the pre-processing stage (Stage 1) of the pipeline in Fig. 3.1b and classifying them as physiological vs technical outliers for possible re-insertion in the transfer learning stage (Stage 2). The first stage of the procedure is called *abnormality filtering* and the second stage is called *abnormality classification*, which will be discussed below. Both procedures apply standard outlier detection methods to a set of predefined features. However, since the purposes of these procedures are different, the outlier detection thresholds are different.

Features for abnormality filtering (Stage 1) and classification (Stage 2): The best features to use for detection of abnormal samples will be experiment dependent and device dependent. Here we explain how the abnormality filtering features were selected for the HVC experiment where participants wore the Empatica E4 device, as discussed in Section 3.4. Based on experiments on an Empatica E4 performed in our laboratory we determined that there are three principal causes for

abnormal measurements, each manifesting abnormality in different combinations of channels. See Fig. 3.7. For abnormality filtering and classification physical intuition motivated us to select three temporally localized features, HR MED, TEMP MED and ACC SD, as they are especially affected by the types of abnormalities described below.

- Device not worn (NW). *Effect*: Skin temperature sensor (TEMP) reads ambient temperature and activity sensor (ACC) records little or no physical movement. *Primary features affected*: median of TEMP and standard deviation of ACC magnitudes are abnormal.

- Device loss of contact (LOC). *Effect*: Intermittent skin contact causes spurious signal dropout. *Primary features affected*: median of TEMP and HR and standard deviation of ACC magnitudes are abnormal.

- Subject engages in intense activity (Active). *Effect*: heart rate (HR) and activity sensor (ACC) readings increase significantly over burst of physical activity. *Features primarily affected*: median of HR and standard deviation of ACC magnitudes are abnormal.

In the HVC study we used HR MED and TEMP MED features for abnormality filtering while we used HR MED, TEMP MED and ACC MED for abnormality classification. The combination of HR MED and ACC MED is especially important for classifying abnormalities that are due to physiological causes,e.g., when a subject is engaged in intense exercise which is legitimate wake session activity that should be reinserted in the final segmented data stream in Stage 2 of the pipeline.

Figure 3.8 illustrates three simulated cases with different levels of separability. Two variables (X_1, X_2) are generated from BVN$(\mu_1, \mu_2, 1, 1, 0)$. For samples in Class 0 (indicated by red dots), $\mu_1 = \mu_2 = 0$, while for samples in Class 1 (indicated by green triangles), we considered three settings: (1) $\mu_1 = \mu_2 = 0$; (2) $\mu_1 = \mu_2 = 1.5$; (3) $\mu_1 = \mu_2 = 3$, corresponding to non-, weakly and strongly separable scenarios. SI values based on both projection distance and Euclidean distance are reported for each case. We observe that SI is indeed able to effectively characterize separability, and that the projection distance is preferred since it gives value closer to 0.5 in the non-separable case and value closer to 1 in the strongly separable case.

3.6.4.1 Online feature extraction pipeline: applied to 0-60 hours of data

Here we describe details of the proposed pipeline, illustrating with the pre-infection (0-60 hours) data. The result of applying the pipeline to the full data (0-270 hours) is described afterward.

Stage 1: pre-processing Pre-processing in stage 1 of the pipeline accomplishes three tasks: temporal windowing and conditioning; local feature extraction, and abnormality filtering. Here we

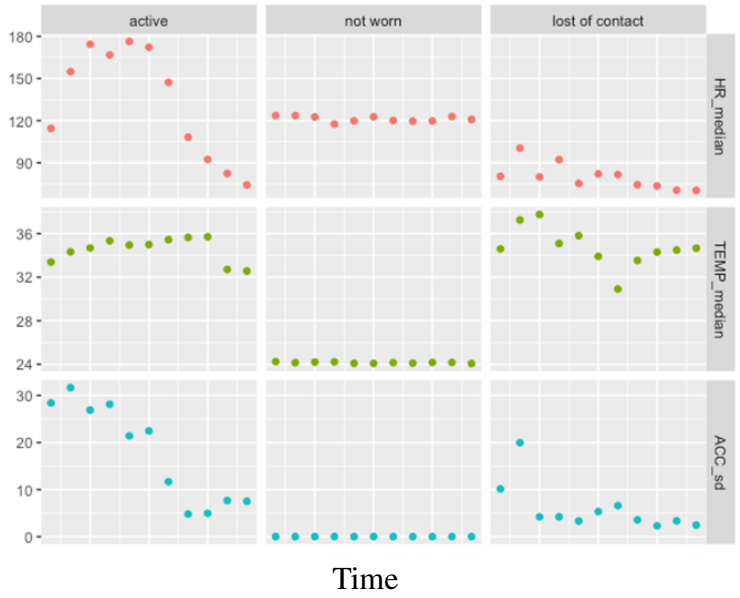


Figure 3.7: Experimental Empatica E4 data collected from one of the co-authors of this paper under three different conditions: active, not worn and loss of contact. Shown are measured values of the variables used in our selected abnormality feature set: HR median, TEMP median and ACC sd. Note the very different values of these three parameters between the active, not worn and loss of contact classes.

illustrate each of these tasks when the pipeline is applied to the pre-infection data, i.e., only the first 60 hours (2 days and two nights) are available. Results of applying the pipeline to the full data from 0 to 270 hours are presented for comparison in Section 3.6.4.2.

Stage 1: temporal windowing and conditioning

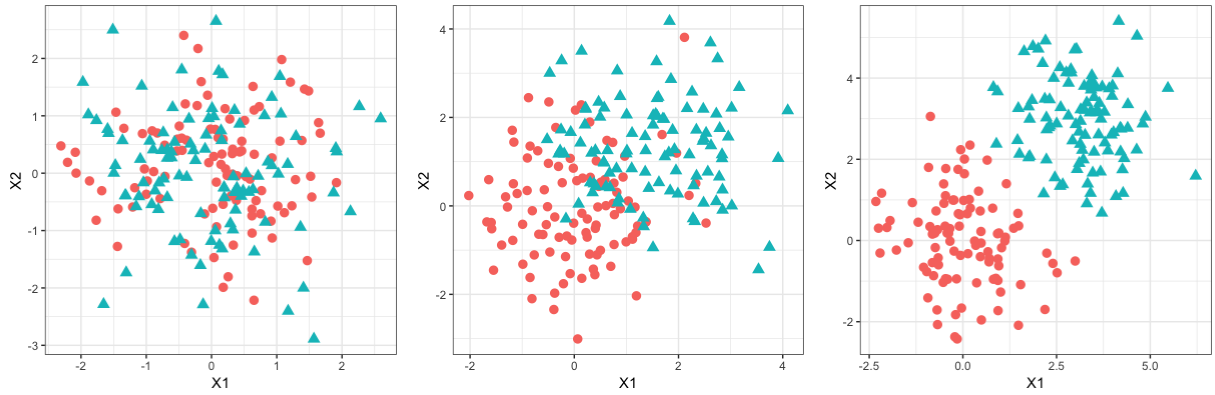
The tuning parameters and their settings for Stage 1 of the pipeline are

- Temporal window (epoch) length: 10 mins
- Subject availability threshold: 60%
- Subject abnormality threshold: 40%

The subject availability threshold of 60% is applied to filter out subjects that have E4 data missingness of greater than 40%. Subject availability is quantified the proportion of time points not missing among the full set of sampling times (number of seconds) over the 0-60 hour period. Subjects with less than 60% data availability are not further processed. Figure 3.9 shows that 2 subjects in the HVC cohort (Subjects with identified 13 and 17) have insufficient data availability over this time period.

Stage 1: abnormality filtering

The abnormality filtering module operates as follows. After temporally localized summary statistic features are extracted from the device for each subject, these features are tested for



(a) $SI_1 = 0.470$, $SI_2 = 0.605$ (b) $SI_1 = 0.750$, $SI_2 = 0.785$ (c) $SI_1 = 1.000$, $SI_2 = 0.990$

Figure 3.8: Separability indices under three simulated scenarios: **(a)** non-separable, **(b)** weakly separable, **(c)** strongly separable, where SI_1 is based on projection distance and SI_2 is based on Euclidean distance.

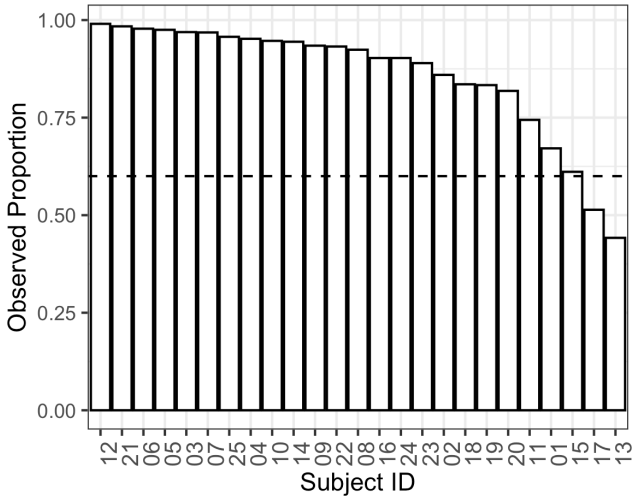


Figure 3.9: Subject availability for the sleep/wake segmentation when the first 60 hours of data is available for subjects in the HVC study. Two subjects (13 and 17) have excessive ($> 40\%$) missingness. The observed proportion is defined as the ratio of the number of time samples in the subject's data record and the total number of sampling times that should be available over the 60 hour period.

abnormality using a clustering-based anomaly detection procedure. The procedure labels the abnormal samples as *non-normal data* and these are temporarily removed from the data stream.

The clustering-based anomaly detection procedure is as follows. Define the vector-valued local features for the i-th subject at the t-th time instant $\tilde{\boldsymbol{X}}_{i,t} = (\tilde{X}_{1i,t}, \ldots, \tilde{X}_{Mi,t})$, where here the feature dimension is $M = 12$. As explained above, only two of these features was used for abnormality filtering: HR MED and TEMP MED. For each subject i, a combination of k-means clustering and quantile thresholding is used to determined a normal region $\mathcal{C}_i$ in this two dimensional feature

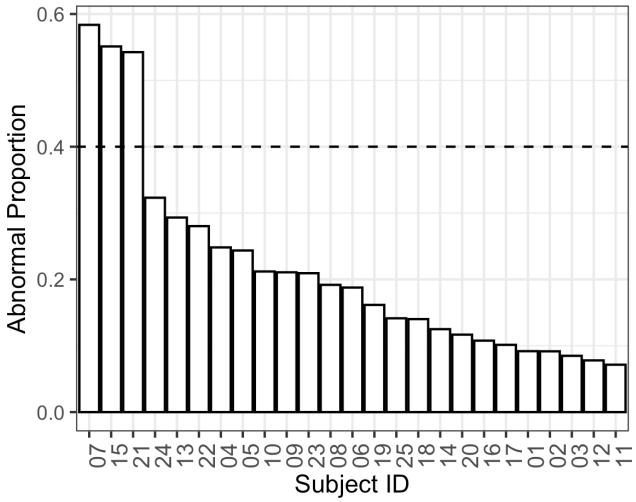


Figure 3.10: Abnormal proportion for the sleep/wake segmentation over 0-60 hours for subjects in the HVC study. Three subjects have an excessive number of abnormal time samples (greater than 40%).

space. Subject i's feature instance at time t is declared non-normal if it is not in $\mathcal{C}_i$. While we also investigated abnormality using k-means clustering in the full 2 dimensional feature space of HR MED and TEMP MED (See Fig. 3.15b), the abnormality filtering method we adopted in the pipeline of Fig. 3.1b performs k-means clustering separately on each each of the feature dimensions.

For the HVC data, where the events of interest are sleep and wake, the k-means algorithm is set to extract 3 clusters corresponding to sleep and wake (normal) and non-normal classes. Let $\{S_{im1}, S_{im2}, S_{im3}\}$ be the resulting clusters with centers (centroids) $\{\mu_{im1}, \mu_{im2}, \mu_{im3}\}$. When the normal and non-normal cluster classes are well separated it is easy to construct a normal region $\mathcal{C}_i$, e.g., all points in the feature space having a majority of k-nearest neighbors in $\{(\tilde{X})_{i,t}\}_t$ outside of the non-normal cluster. More often, however, the separation between the normal and non-normal cless is not sufficient and a different method is needed. We determine $\mathcal{C}_i$ as the set complement of the hyper-rectangle of minimal volume whose empirical coverage probability is 95%. The rectangular edge lengths and position are thus determined by the marginal sample quantiles along each feature dimension.

Specifically, we define a sequence of subject-dependent cutoff values $\{c_{i1}, \ldots, c_{iM}\}$ for abnormality. The rectangular normal region is designated as the Cartesian product $S_i^{nor} = S_{i1}^{nor} \times \ldots \times S_{iM}^{nor}$, where

$$S_{im}^{nor} = \begin{cases} S_{im1} \cup S_{im2} & \text{if } |\mu_{im2} - \mu_{im1}| < |\mu_{im3} - \mu_{im2}|, \\ S_{im1} & \text{o.w..} \end{cases}$$

Cutoff values defining the normal set are defined as

$$c_{im} = \begin{cases} Q_{0.025}(\{\tilde{X}_{mi,t} \in S_{im}^{nor}\}) & \text{if } \mu_{i,nor} > \mu_{i,abn}, \\ Q_{0.975}(\{\tilde{X}_{mi,t} \in S_{im}^{nor}\}) & \text{o.w.}, \end{cases}$$

where $Q_q(\cdot)$ is the $q-$th quantile of the samples, and $\mu_{i,nor}$, $\mu_{i,abn}$ are the centroids of the normal and non-normal set, respectively. For the experimental HVC data we used the 95% outlier rule to define the lower and upper quantiles as 2.5% and 97.5%, but less stringent threshold values may be adequate for other datasets.

For the subjects in HVC study with the first 60 hours of data available, Figure 3.10 shows a bar-plot of the resulting non-normal data proportion for each subject. Three subjects (7, 15 and 21) had excessive abnormal data and were eliminated from further processing.

Thus, when taken with subjects who had excessive missingness, a total of five subjects were excluded from further analyses.Some of these subjects may come back into the analysis when a greater amount of data is available (see subsection C for the case where 0-270 hours are available).
Feature selection for training adaptive segmentation algorithm: As explained in the main text, we extracted a subset of the 12 local features using the sleep/wake separability index (SWSI) (3.6). The SWSI is computed over all subjects, to contrast each feature over the putative sleep periods 2:00-5:00 and the wake periods 19:00-22:00 in the 0-60 hour time interval of the pre-infection data window. Figure 3.11 shows the resulting SWSI for each of the 12 features, in descending order of median. Applying the selection rule that a selected feature must be greater than 0.7 for at least 75% of the subjects, three features are above threshold: HR MEAN, HR MED and HR SD and, as HR MEAN and HR MED are highly correlated, we chose eliminated HR MEAN since it has a lower 75% quantile than does HR MED.

Stage 2: adaptive segmentation The HMM-FLDA procedure is implemented for a given subject on all the time points that pass through the abnormality filter. The details of the HMM-FLDA are given in Sec 3.4.1.2.

Stage 2: Separability index

By definition (3.5), SI is the proportion of samples that share the same label with their nearest neighbors, and hence SI $\in [0, 1]$. Intuitively, when samples from two classes form two tight, well-separated clusters with little overlap, the nearest neighbor of one sample from, say, Class 0, will most likely belong to Class 0 as well, which will result in a large SI value close to 1. In contrast, when samples from two classes follow exactly the same distribution, i.e., completely non-separable, then the nearest neighbor of one sample will have equal probability of being Class 0 or Class 1, and

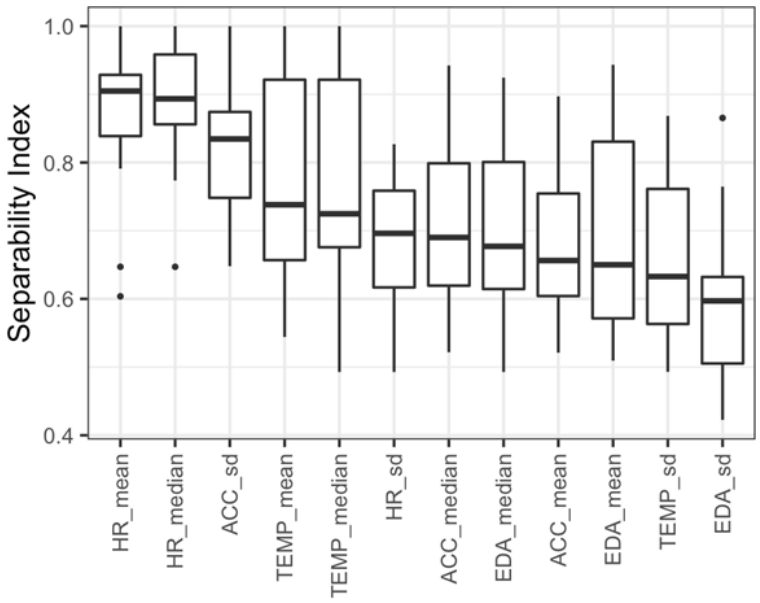


Figure 3.11: Box-plot of sleep/wake separability index (3.6) for pre-infection (0-60 hour) data, showing the spread of the 12 local features extracted from wearable data in the first stage of the processing pipeline. Excluding HR MEAN, which is highly correlated with HR MED and has a smaller 25% quantile, two local features (HR MED, ACC SD) are selected.

thus, the SI of these samples is close to 0.5. A large SI value implies strong separability of classes, and is usually an indication of reliable prediction. SI also depends on the measure of distance used to determine the nearest neighbors. The projection distance captures the difference among samples on the optimal direction $\boldsymbol{w}$ that is most relevant to distinguishing between the two classes, and thus is better than Euclidean distance in regard to characterizing separability, as demonstrated by Figure 3.8.

Once the HMM-FLDA event classification procedure terminates the initial sleep/wake segmentation will have missing time points that have been removed by the abnormality filter. Some of these abnormal time points, e.g., those due to rare physiological events like an exercise session, are reinserted into the data stream. The method to do this is based on an abnormality classification procedure discussed in the next subsection.

Stage 2: Abnormality classification and reinsertion

The purpose of the abnormality classification module is to identify, re-insert and assign an event label to physiologically meaningful non-normal data, i.e., abnormalities that are not due to technical issues associated with device failure or improper wearing of the device. Any such non-normal samples re-inserted into the data stream inherit the event class label of the session into which the sample's time stamp falls.

Using predefined features HR MED, TEMP MED, and ACC SD, abnormality classification

was performed using a decision tree using marginal quantile thresholding. For each non-normal sample, we applied a standard 1.5 interquartile range (IQR) test [Tuk77, HIT86] to each feature independently. If a feature falls outside of this range it is declared an outlier and further classified. The lower and upper endpoints of the 1.5 IQR interval are defined as:

$$\begin{aligned} \text{LOWER} &= Q_{0.25} - 1.5 * (Q_{0.75} - Q_{0.25}), \\ \text{UPPER} &= Q_{0.75} + 1.5 * (Q_{0.75} - Q_{0.25}). \end{aligned} \tag{3.7}$$

Here $Q_{0.25}$ and $Q_{0.75}$ are the 25% and 75% sample quantiles of the feature empirical distributions computed over the detected sessions. All samples that were categorized as "Wake" or "Active" were incorporated back into the corresponding session. Abnormal samples assigned to other categories were discarded. The decision tree for this procedure is shown in Figure 3.12.

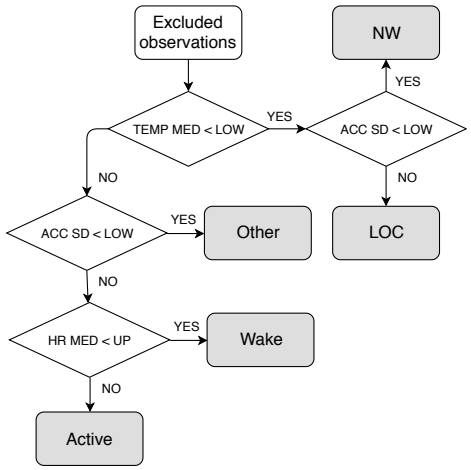


Figure 3.12: Decision tree for abnormality classification module for the HVC data. Non-normal samples (excluded observations) identified by the abnormality filter in first stage of the pipeline of Fig. 3.1b are excluded from the training set used by the second stage transfer learning algorithm. The abnormality classifier re-evaluates these samples for possible re-insertion and labeling of physiologically meaningful abnormalities. The classifier classifies the samples into final categories of: device not worn (NW), loss of contact (LOC), Wake, Active and Other. Samples that are classified as "Active" or "Wake" are re-inserted into the corresponding session.

Stage 2: Median filtering

After processing by HMM-FLDA and reinsertion of physiologically meaningful abnormalities, there commonly exists short bursts of sleep sessions. We expect that some of these are actually sleep while others correspond to resting without sleep. While such bursty behavior could possibly be directly incorporated into an HMM model, e.g., using a semi-Markov switching process [Ani08], we took a simpler approach that applies a modified median filter with 90 min smoothing window

(median filter of order 9) that has the effect of merging sleep sessions shorter than 60 mins into a wake session. We choose to eliminate such short sessions in order to eliminate disambiguate resting and light naps from sleep sessions, defined as a session having deep sleep stages, i.e., Stage 3 & 4 of non-rapid eye movement (NREM), usually starting 30 minutes after sleep onset and lasting approximately 20 to 40 minutes in the first sleep cycle [CD$^+$05]. Hence, the 60-minute threshold is designed to eliminate putative sleep sessions that had no deep sleep stages.

3.6.4.2 Offline feature extraction pipeline: applied to 0-270 hours of data

Stage 1: pre-processing When evaluated over the time period 0-270 hours, no subjects had missingness greater than 40% and thus none were filtered out due to inadequate data availability. However, five subjects were found to have greater than 40% abnormal samples and were filtered out. Figure 3.13 shows the bar-plot of the abnormal proportions for all 25 subjects, showing the 5 subjects exceeding the 40% threshold for inclusion.

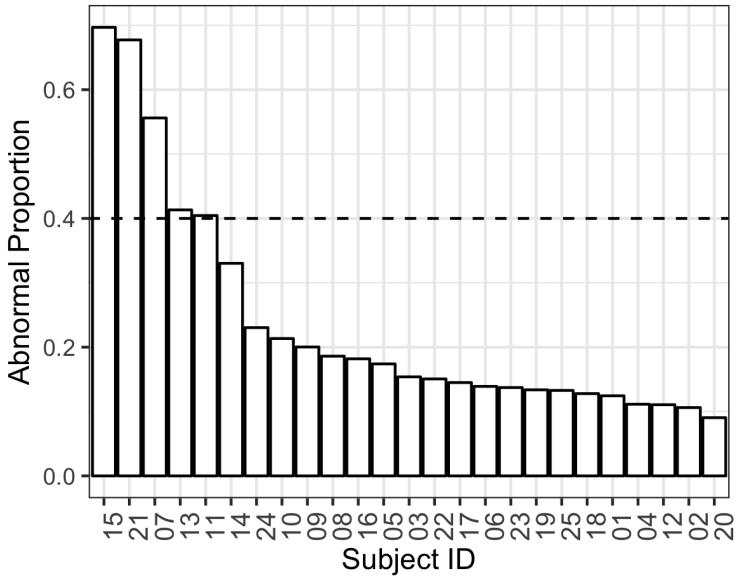


Figure 3.13: The bar-plot of abnormal data proportions over all time points (0-270) in HVC study data. Bar-plot of availability (analog to Fig. 3.9) is not shown since no subjects had more than 40% missing data over this 270 hour time period.

To select features for the HMM-FLDA sleep/wake segmentation algorithm over the full time course of 0-270 hours, we designated two one hour periods of the day, 3:00-4:00 and 21:00-22:00, respectively, as sleep and wake (resting). As contrasted to the pre-infection segmentation in which we used 3 hour periods in the 60 hours (2 nights) of available data, here we could take advantage of the higher specificity of one hour periods since 11 nights are available over the 270 hours. The sleep/wake separability indices (SI), defined in (3.6), of each of the 12 temporally localized features

were computed over all time points in order to evaluate their discrimination power. Figure 3.14 shows box-plots of the SWSI values for the 12 local features.

Four local features had SWSI greater than 0.7 for over over 75% of the subjects: HR MEAN, HR MED, HR SD, ACC SD. Since HR MEAN and HR MED are highly correlated, HR MEAN was excluded from the top 4 SI features as its 25% quantile is lower than that of HR MED. The remaining features were then used as feature variables in the second stage of the pipeline to identify sleep/wake sessions for each subject. Like in the 0-60 hour segmentation, we again note that the EDA features have lower sleep/wake separability indices than the other E4 variables.

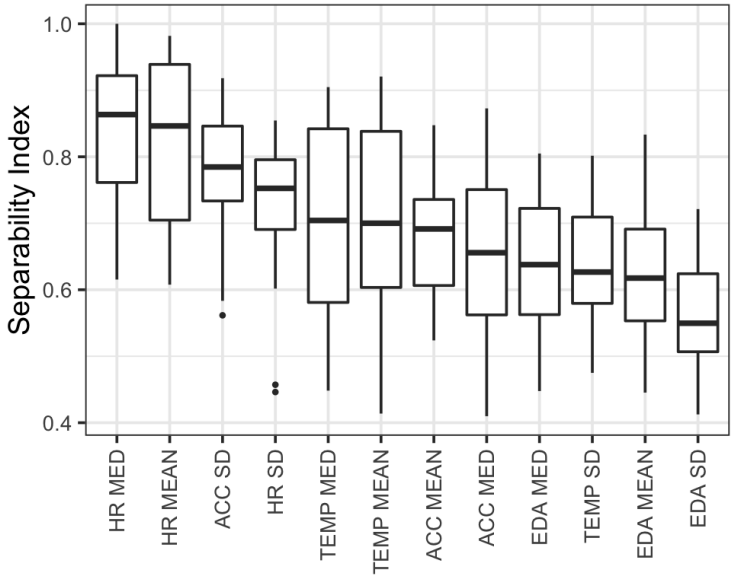


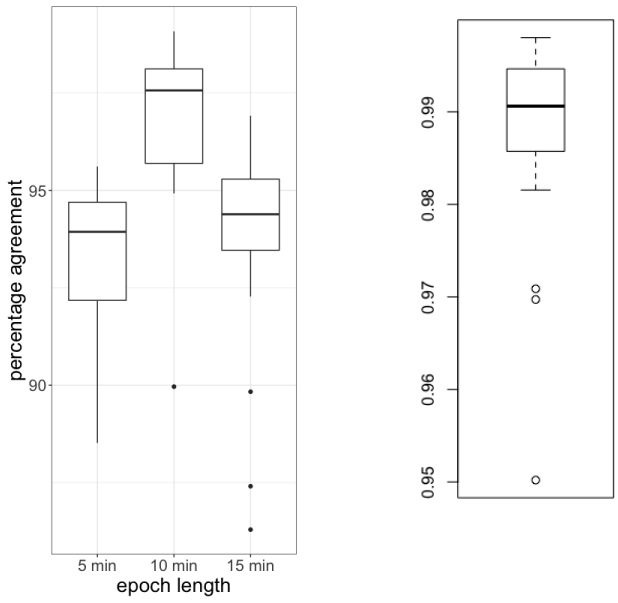
Figure 3.14: Box-plot of sleep/wake separability index (3.6) for full (0-270 hour) data, showing the spread of the 12 temporally localized features extracted from wearable data in the first stage (pre-processing) of the processing pipeline. Excluding HR MEAN, which is highly correlated with HR MED and has smaller 25% quantile, three features (HR MED, ACC SD, HR SD) are selected for training HMM-FLDA in the second stage of the pipeline.

Stage 2: adaptive segmentation The following table summarizes the results of applying abnormality classification (Fig. 3.12) in Stage 2 to the abnormal 10 min epochs identified by the abnormality filter in Stage 1. For the purposes of illustration, the table is restricted to epochs falling within the typical resting period (21:00 - 22:00) and the typical sleeping period (03:00 - 04:00). The counts in the table represent abnormality classification over the full 270 hours of the HVC experiment summed over the 20 subjects falling under the 0.4 abnormal proportion threshold in Fig. 3.14. The Normal category are the Wake and Active classes in Fig. 3.12 and represent epochs that were re-inserted into the segmented event stream after termination of the HMM-FLDA algorithm. No epochs in these one hour intervals were labeled as "Other."

	Total	NW	Loss-of-contact	Normal
Resting (21:00-22:00)	1194	0	122	1072
Sleep (03:00-04:00)	1194	0	46	1148

Sensitivity of pipeline to tuning parameters Here we illustrate the lack of sensitivity of the segmentation to our choice of two tuning parameters: epoch length δ_t and the number k of nearest neighbors in the k-means classifier used for abnormality filtering.

Local optimality of selected epoch length for 0-270 hour data: To check the sensitivity of our choice of 10 min epoch length δ_t we compared three different epoch lengths 5, 10 and 15mins, on the full data (0-270 hours). Figure 3.15a shows a consensus criterion that is defined as the *agreement rate* between the sleep/wake classifications of the HMM-FLDA pipeline using the 5, 10 and 15 minute epoch lengths, respectively. Figure 3.15a shows that the 10 min epoch length results in sleep/wake labels that achieve the highest agreement rate. In this sense, $\delta_t = 10$ mins. is locally optimal for segmenting the 0-270 hour data. If other types of events are of primary interest, e.g., higher temporal resolution sleep stage analysis or sleep quality analysis, for which the presence of outliers may be useful information, shorter epoch lengths might work better.



(a) (b)

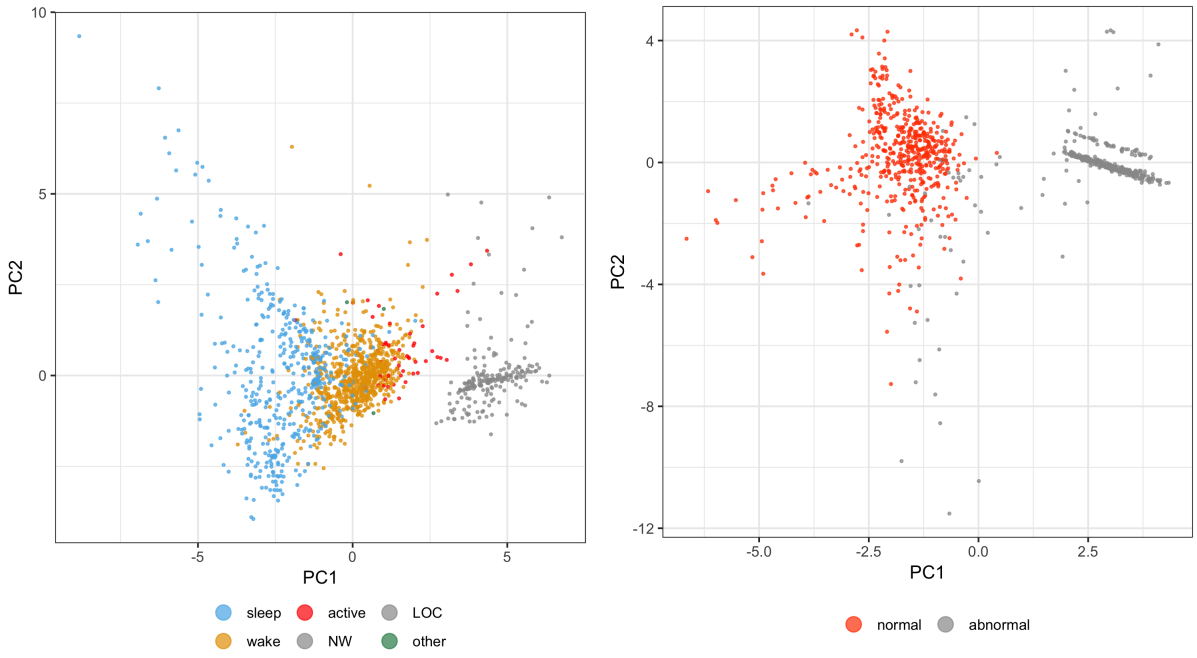
Figure 3.15: **(a)** Box-plot of HMM-FLDA event label agreement rates between three different epoch lengths showing that the choice of $\delta_t = 10$ min comes closest to consensus. **(b)** Box-plotshowing high agreement rates between abnormality detection based on marginal and multivariate k-means clustering among all subjects.

Choice of dimension in the k-means clustering used for abnormality filtering (0-270 hours): In the context of the HVC data with the full 270 hours of data available, Fig. 3.15b compares k-means clustering for different feature dimensions in the context of abnormality filtering described in Sec. 3.6.4.1. Specifically, the case of multivariate k-means, where k-means is applied to local feature vectors in their full 12 dimensions, is compared to the case of marginal k-means, where it is applied to each feature dimension independently. For each subject we calculated the agreement rate (percentage agreement) between the multivariate and marginal k-means implementation, where agreement is defined as the number of times they agree that a time sample is abnormal over the subject's 270 hours of data. The box plot in Fig. 3.15b shows the distribution over the 25 subjects and indicates that there is greater than 0.98 agreement rate for all but 3 subjects whose agreement rates are all still greater than 0.95. Thus the simpler marginal k-means is virtually equivalent to the multivariate k-means.

Example: abnormality filtering for two of the subjects (0-270 hours) For the purposes of visualization of the differences between the features of subjects with usable and unusable data, we show the results of principal components analysis (PCA) of two subjects in Figures 3.16a and 3.16b. These figures show scatter plots of the first two principal components of all 12 temporally localized features over the full 0-270 hour data. Both participants have clusters of outliers (grey colored points falling on the right side of the figures) that were identified as non-normal by the abnormality filter in our processing pipeline of Fig. 3.1b. For the subject in Figures 3.16b the non-normal data corresponded to a HR and TEMP features that are highly abnormal (mean TEMP $< 20^o$C and mean HR $> 160bpm$) over long time periods, suggestive of a device that is either malfunctioning, not being worn properly, or not being worn. This latter subject was one of the 5 subjects that was eliminated from the analysis since more than 40% of all his time points were detected as abnormal. For the subject in Figure 3.16a the detected non-normal data corresponded to a short time period where the HR and ACC features exhibited high median and high standard deviation, suggestive of a short session of intense physical exercise. This subject was included in the analysis since less than 40% of the time points were detected as abnormal. For this subject, the abnormality classifier described in Appendix 3.6.4.1 classified these non-normal data points as "Active," and, based on their time stamps, inserted them into back into the corresponding wake session determined by the HMM-FLDA procedure.

Example: segmentation results for two subjects (0-270 hours)

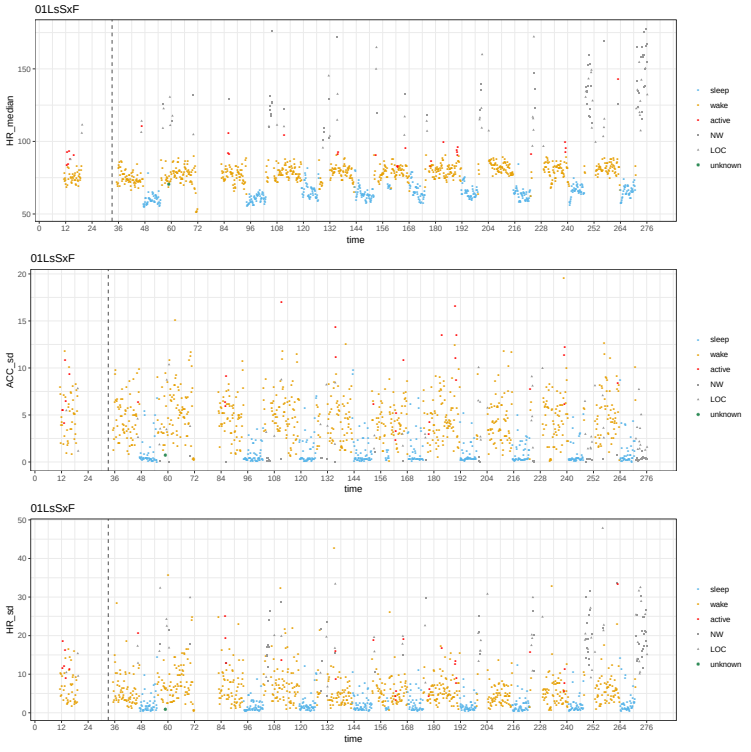
Figure 3.17 shows the sleep segmented data for two subjects with usable data (Subject 1 and Subject 2 shown in Fig. 3.4) at the output of the pipeline of Fig. 3.1b when all time points (0-270 hours) of all subjects are available for processing.



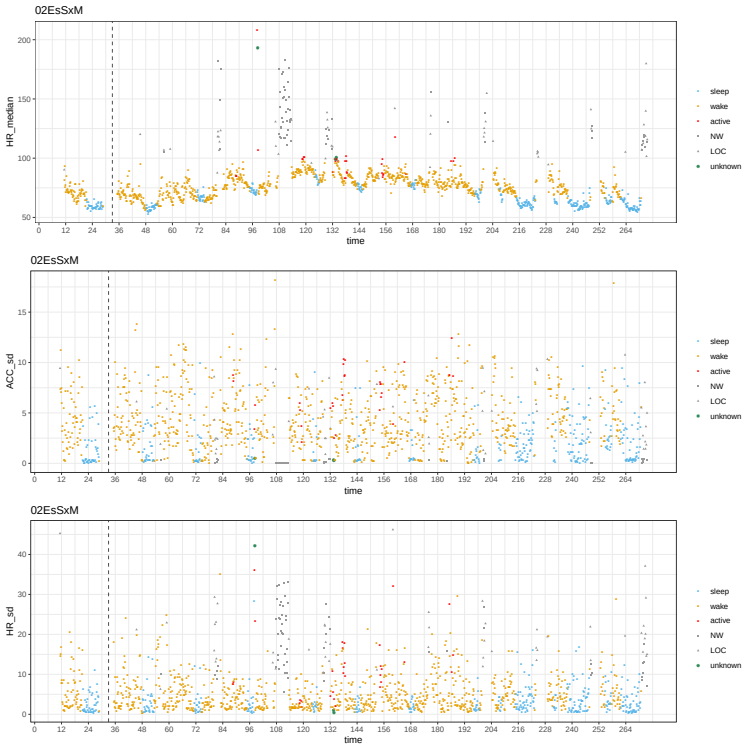
(a) (b)

Figure 3.16: (**a**) scatter plot of the first two PCs for a subject with usable data (abnormal proportion < 40%); (**b**) scatter plot of the first two PCs for a subject with unusable data (abnormal proportion > 40%)

.



(a)



(b)

Figure 3.17: Processed segmented time courses for Subjects 1 and 2 shown in Fig. 3.4 for the 3 event classification features, HR median, ACC sd and HR.sd, color coded according to the abnormality classification categoires: sleep, wake, active, not-worn (NW), loss-of-contact (LOC), and other (Unknown). **(a)** segmented data from non-infected Subject 1; **(b)** segmented data from infected Subject 2.

3.6.5 Clinical outcome prediction

Here we give additional information relevant to prediction of clinical outcomes, infection/non-infection status and shedding onset time, as discussed in Section 3.4.

Recall that the pipeline produces high dimensional sleep/wake labeled features obtained as statistical summarizations of the local E4 features. Two implementations are evaluated: offline and online feature extraction. The offline implementation accesses the subject's full data (all sampling times, 0-270 hrs), using it to perform anomaly detection, HMM-FLDA sleep/wake segmentation and feature extraction. The online implementation only has access to the pre-infection data (sample times over 0-60 hours) to perform these operations. The offline and online versions of the pipeline both ended up rejecting 5 of the subjects with excessive missingness and abnormality, and for each non-rejected subject generates a set of features described in Table 3.3.

In Table 3.7 are shown the distribution of subjects over the clinical outcomes, in the separate categories of infection status and infection onset time. Here infection status is the binary outcome that a subject will or will not shed virus over the course of 270 hours of the confinement phase of the challenge study. Infection onset time is a ternary response variable indicating the categories "Early onset" (shedding begins on day 4), "Late onset" (shedding begins on day 5), and "No onset" (shedding not detected on any day).

The 1st row of Table 3.7 shows numbers of subjects in each infection category among all 25 subjects having E4 data records. The 2nd row shows the same statistics for the 20 subjects with sufficient quantity and quality data as determined by the missingness and abnormality filtering implemented by the offline feature extraction pipeline having access to the full time course (0-270 hrs). The 3rd row shows these statistics for the 20 subjects with sufficient quantity and quality data determined by the online feature extraction pipeline (0-60hrs).

Note that the classes are moderately well balanced for the binary infection status class but are unbalanced for the ternary shedding onset time, with the early class having only 2 or 3 subjects. We address this class imbalance using a compensation method in Section 3.6.5.

To assess the value of each of the 196 features for predicting clinical outcomes in Table 3.7 to types of univariate classifiers were implemented on each feature and their performance was evaluated using the area under the receiver operating characteristic (ROC) curve.

For classifying the binary infection status the logistic regression (LR) classifier was implemented. The LR classifier fits the probability $Pr(Y = y)$, $y \in \{0, 1\}$, to the generalized linear model:

$$\text{logit}\big[Pr(Y = 1|X)\big] = \beta_0 + \beta_1 X\,,$$

where $\text{logit}(x) = \log[x/(1-x)]$, $y \in \{0, 1\}$ is the shedding status, and X is one of the 196 features produced by the pipeline.

Table 3.7: First row: infection status shedding onset time of all 25 subjects having E4 data records. Second row: infection status of the 20 subjects with sufficient non-anomalous data over 0-270 hours as determined by the offline feature extraction pipeline having access to the full time course data (0-270 hours). Third row: infection status of the 20 subjects with sufficient non-anomalous data as determined by the online feature extraction pipeline having access only to the pre-infection time course data (0-60 hours).

Subject group	Infection Status		Shedding onset time		
	Infected	Not infected	Early	Late	No onset
25 subjects w/ E4 data records	11	14	3	8	14
20 subjects w/ sufficient data (0-270 hrs)	8	12	2	6	12
20 subjects w/ sufficient data (0-60 hrs)	9	11	3	6	11

For classifying the onset time we fit a *continuation-ratio* (CR) regression model [Agr10] to the conditional probability $P(Y = y|Y \geq y)$, where $y = 1$ for early shedders, $y = 2$ for late shedders and $y = 3$ for no onset (non-) shedders). This becomes an ordinal censured response model when "no onset" is identified as onset at infinity. Mathematically, the continuation ratio is a logistic regression model for this conditional probability:

$$\text{logit}\big[Pr(Y = y \,|\, Y \geq y)\big] = \beta_0 + \beta_1 X \,,$$

for $y = 1, 2, 3$.

3.6.5.1 Clinically discriminating features: online pre-infection timecourse feature extraction

We emulate online implementation of the feature extraction pipeline for infection status prediction by applying the pipeline to only the pre-infection timecourse (0-60 hours). In this case, only the first 60 hours affect the sleep segmentation and the sleep features. Some of the 196 resulting sleep/wake features in Table 3.3 are useful for predicting the the clinical outcome, i.e., infection status, of exposure to the viral challenge. Table 3.8 shows the ten top ranking features for predicting infection vs. non-infection (left column) and for predicting infection onset (right column) where the predictors use features from inoculation day only. To obtain the rankings on the left column of the table, the AUC attained by the univariate logistic regression (LR) was computed using leave-one-out cross validation (LOOCV) on the 20 subjects who were not filtered out by Stage 1 of the pipeline, i.e., these subjects do not have excessive missingness or abnormal samples. We also include the Accuracy (average classification error), computed using the same LOOCV procedure as used for computing AUC. Observe that the top 3 features ranked by accuracy are sleep features. Feature rankings by Accuracy and AUC are not the same since AUC is more stringent as it measures

accuracy over the full range of sensitivities and specificities of the LR model. The right column of the table is obtained by ranking the features in decreasing order of min {AUC(Early), AUC(Late), AUC(No onset)}.

Table 3.8: Top 10 E4 sleep/wake features capable of predicting clinical outcome (shedding) using only the extracted features from inoculation day (day 2) when these features are obtained from the online implementation of HMM-FLDA, for which only pre-infection data (0-60 hrs) is available to the HMM-FLDA procedure.

logistic regression model				*continuation-ratio regression model*				
Feature	Coef.	Accuracy	AUC	Feature	Coef.	AUC		
						Early	Late	No onset
HR MED.sd (sleep)	-3.921	0.842	0.758	HR MED.sd (sleep)	-5.073	0.882	0.718	0.864
ACC SD.linear.coef1 (wake)	-14.468	0.750	0.737	HR MEAN.sd (sleep)	-4.466	0.706	0.628	0.773
HR MED.quad.coef2 (sleep)	4.318	0.789	0.707	Total duration	-0.599	0.676	0.551	0.750
Offset	-1.452	0.789	0.697	Night duration	-0.616	0.647	0.500	0.705
EDA MED.linear.coef0 (wake)	0.802	0.600	0.697	ACC MEAN.sd (sleep)	-2.303	0.500	0.449	0.602
EDA MEAN.linear.coef0 (wake)	0.815	0.650	0.687	Offset	-0.774	0.618	0.436	0.750
HR MEAN.sd (sleep)	-4.077	0.632	0.677	ACC SD.linear.coef1 (wake)	-10.394	0.412	0.679	0.747
HR MED.linear.coef0 (wake)	0.117	0.650	0.677	HR MEAN.linear.coef0 (wake)	0.091	0.412	0.488	0.646
Total duration	-0.655	0.632	0.667	ACC MEAN.linear.coef1 (wake)	-13.907	0.412	0.512	0.636
HR SD.quad.coef2 (sleep)	3.949	0.737	0.646	HR SD.sd (sleep)	-1.831	0.471	0.410	0.568

Figures 3.18 and 3.19 show box-plots of the four top features in Table 3.8 for logistic regression and continuation ratio regression stratified over shedders and non-shedders.

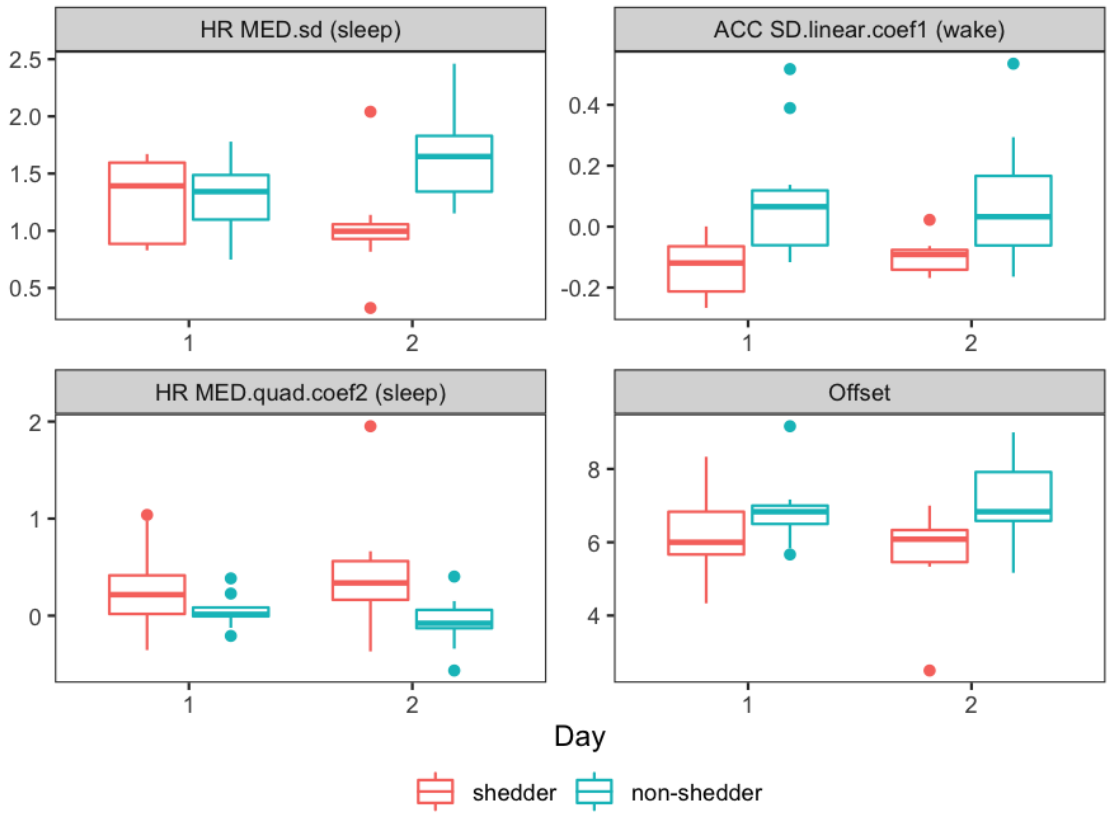


Figure 3.18: Boxplot of top 4 features for predicting binary infection status using univariate logistic regression on the 196 sleep/wake features on the day of inoculation (day 2) obtained from online HMM-FLDA, which only has access to pre-infection data (0-60 hrs).

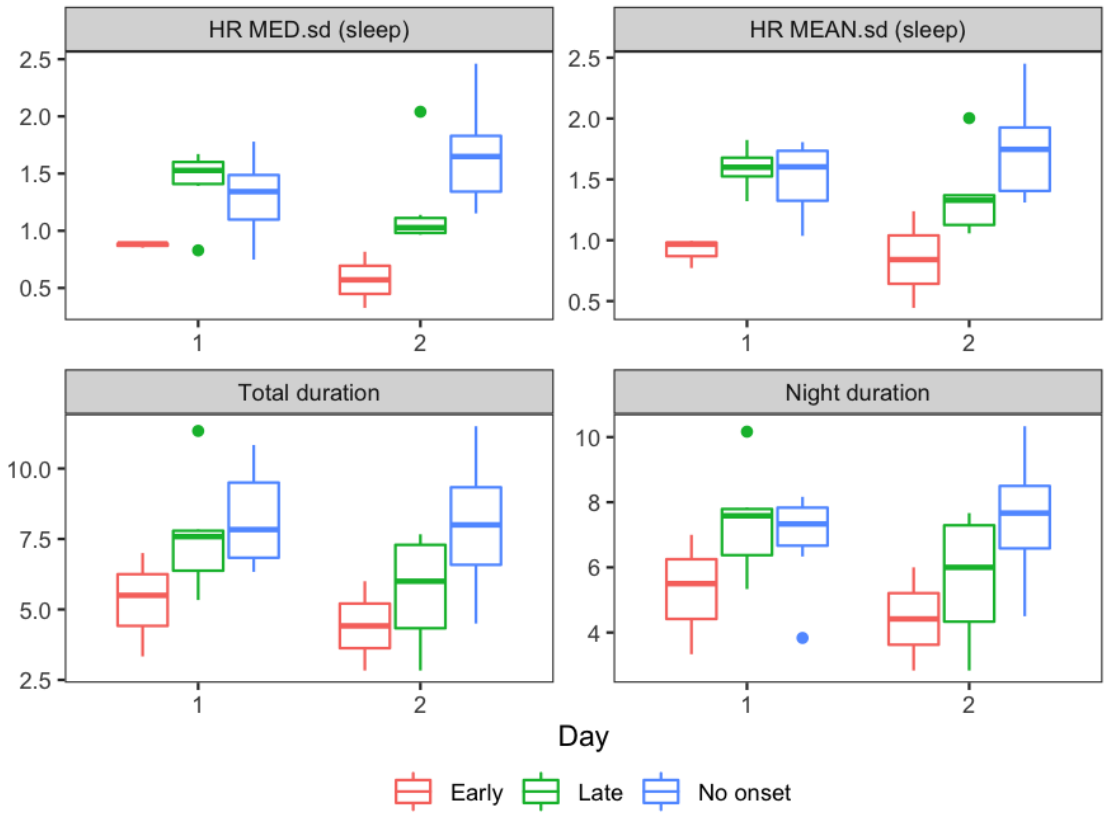


Figure 3.19: Boxplot of top 4 features for predicting ternary infection onset time using univariate continuation ratio regression on the 196 sleep/wake features on the day of inoculation (day 2) obtained from online HMM-FLDA, which only has access to pre-infection data (0-60 hrs).

3.6.5.2 Clinically discriminating features: offline full timecourse feature extraction

In the offline implementation, the full timecourse (0-270 hours) is provided to the pipeline for sleep/wake segmentation and the session-level feature set depends on all past and future time points. To these features we trained the same LR and CR infection status classification algorithms status as in the online implementation described above. Table 3.9 shows the top 10 features in terms of maximizing the AUC. Many of the features in this offline table are similar to the features found by the online implementation shown in Table 3.8. The offline AUC tends to be higher than the online AUC in Table 3.8, possibly due to improved wake/sleep segmentation when more data is available for training (270 hours of data instead of 60 hours of data).

Table 3.9: Top 10 E4 sleep/wake features capable of predicting clinical outcome (shedding) using only the extracted features from inoculation day (day 2) when these features are obtained from the offline implementation of HMM-FLDA, for which the full data (0-270 hrs) is available to the HMM-FLDA procedure.

logistic regression model				*continuation-ratio regression model*				
Feature	Coef.	Accuracy	AUC	Feature	Coef.	AUC		
						Early	Late	No onset
HR MED.sd (sleep)	-6.196	0.750	0.833	HR MED.sd (sleep)	-7.195	0.944	0.631	0.844
Offset	-1.629	0.650	0.802	Total duration	-1.139	0.750	0.619	0.813
Total duration	-1.211	0.750	0.781	HR MEAN.sd (sleep)	-4.597	0.667	0.595	0.750
HR MEAN.sd (sleep)	-4.958	0.650	0.750	Night duration	-0.854	0.667	0.560	0.698
Night duration	-0.896	0.600	0.688	Offset	-0.864	0.611	0.488	0.802
HR MED.mean (wake)	0.147	0.750	0.625	HR MED.mean (wake)	0.154	0.472	0.619	0.646
HR MED.median (wake)	0.149	0.750	0.625	HR SD.sd (sleep)	-1.877	0.500	0.464	0.573
HR MEAN.median (wake)	0.142	0.750	0.615	HR MED.median (wake)	0.150	0.444	0.595	0.625
ACC SD.linear.coef1 (wake)	-7.160	0.550	0.615	HR MEAN.mean (wake)	0.151	0.417	0.607	0.635
HR MEAN.mean (wake)	0.141	0.750	0.604	HR MEAN.median (wake)	0.142	0.417	0.607	0.625

Figures 3.20-3.24 show box-plots of the top 5 top logistic regression features (first column) in Table 3.9 stratified over shedders and non-shedders.

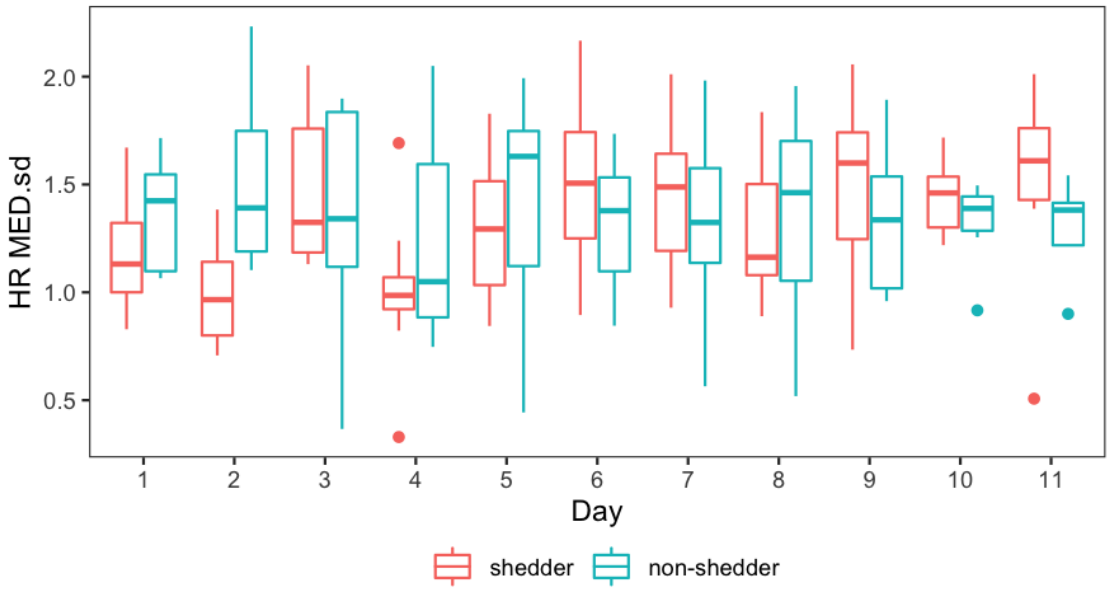


Figure 3.20: HR MED.sd (Standard deviation of median heart rate) sleep feature generated by the proposed HMM-FLDA pipeline when implemented offline, i.e., the HMM-FLDA sleep/wake segmentation is computed assuming availability of the full time course (0-270 hours) human viral challenge study (HVC) data. Viral inoculation took place on day 2 and all viral shedders started shedding on day 4 or day 5. Note that, as compared to the non-shedders, the shedders tend to have lower sleeping heart rate variation in the first two days.

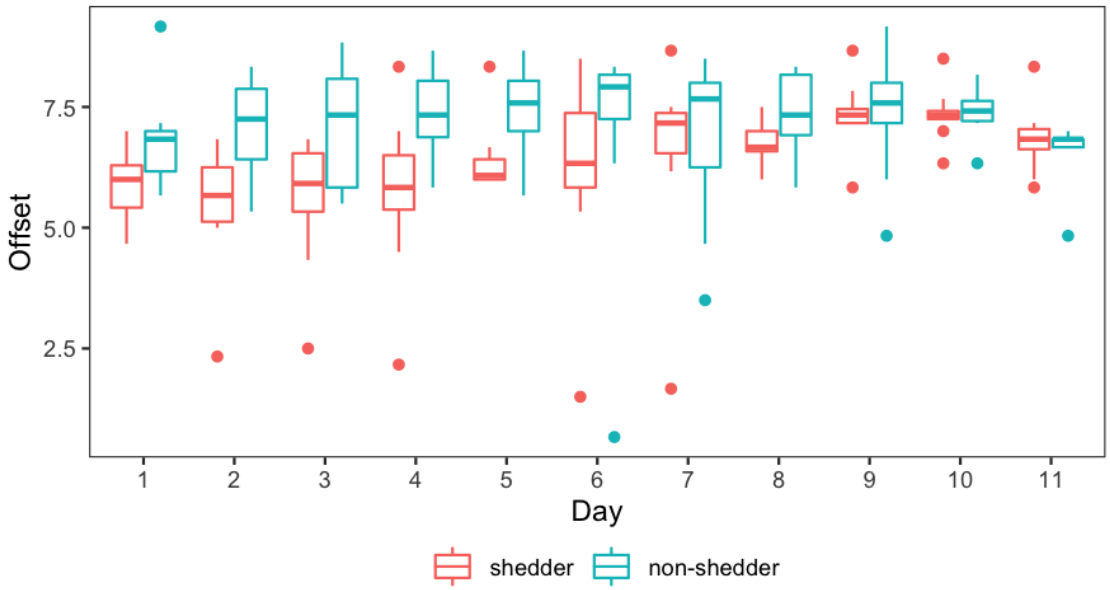


Figure 3.21: Sleep offset (wakeup time from night sleep) feature generated by the proposed HMM-FLDA pipeline when implemented offline, i.e., the HMM-FLDA sleep/wake segmentation is computed assuming availability of the full time course (0-270 hours) human viral challenge study (HVC) data. Viral inoculation took place on day 2 and all viral shedders started shedding on day 4 or day 5. Note that, as compared to the non-shedders, during the first 8 days of the study the shedders tend to wake up earlier than the non-shedders.

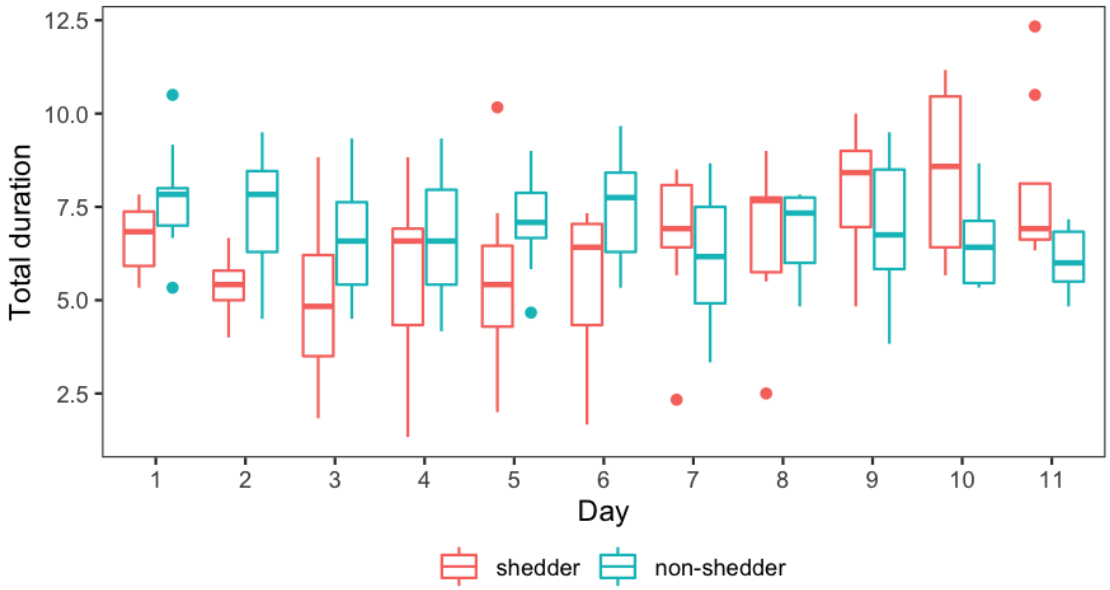


Figure 3.22: Total duration (the sum of all night and day sleep hours) sleep feature generated by the proposed HMM-FLDA pipeline when implemented offline, i.e., the HMM-FLDA sleep/wake segmentation is computed assuming availability of the full time course (0-270 hours) human viral challenge study (HVC) data. Viral inoculation took place on day 2 and all viral shedders started shedding on day 4 or day 5. As compared to the non-shedders, the shedders tend to have a sleep deficit until the 7th day.

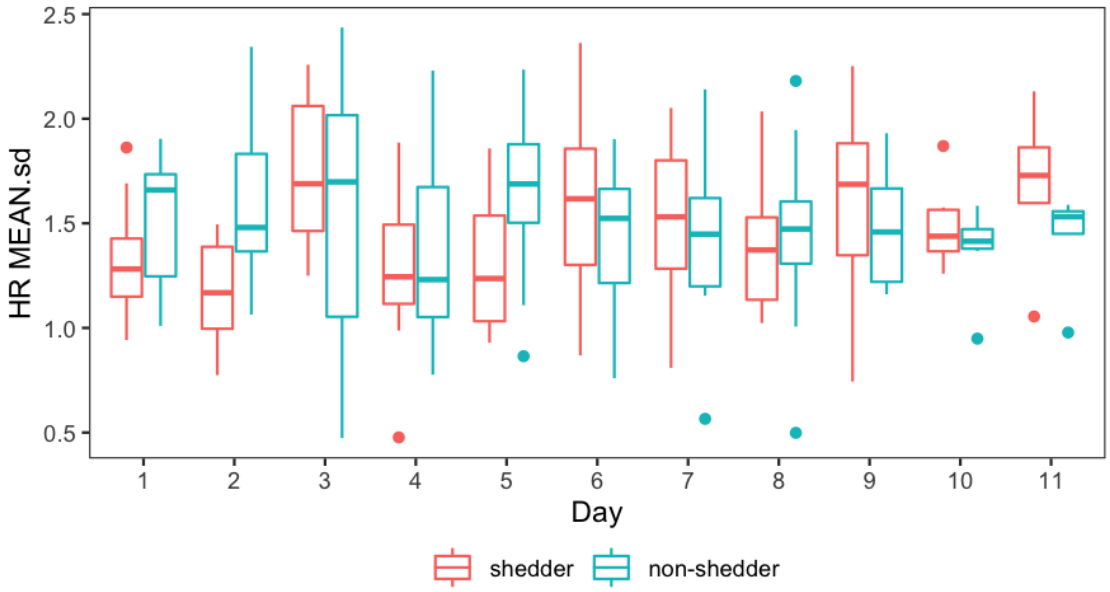


Figure 3.23: HR MEAN.sd (standard deviation of the heart rate mean) sleep feature generated by the proposed HMM-FLDA pipeline when implemented offline, i.e., the HMM-FLDA sleep/wake segmentation is computed assuming availability of the full time course (0-270 hours) human viral challenge study (HVC) data. Viral inoculation took place on day 2 and all viral shedders started shedding on day 4 or day 5.

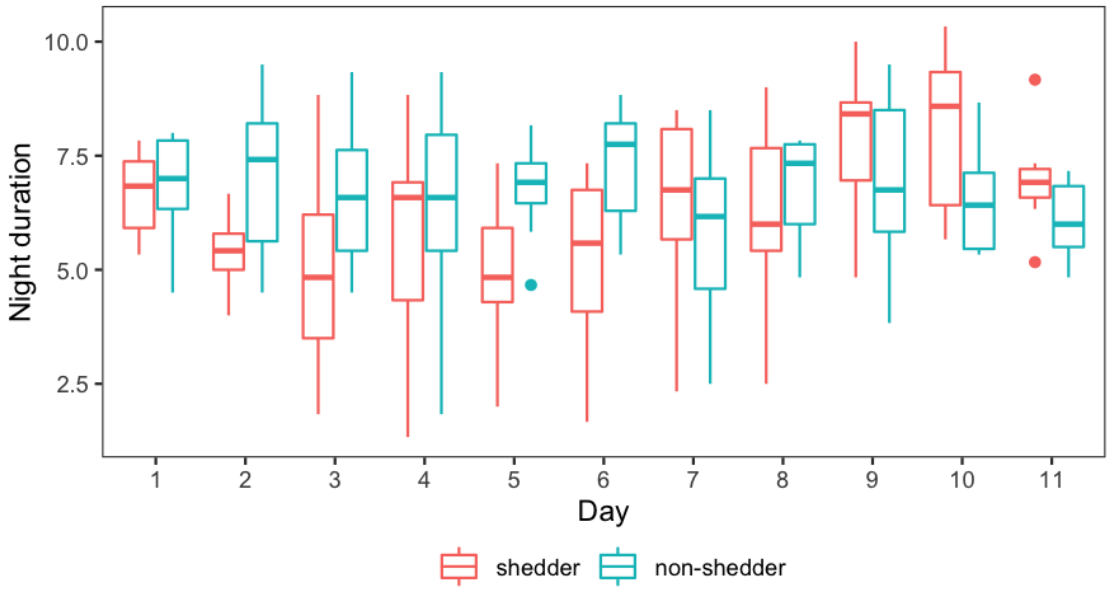


Figure 3.24: Night sleep duration feature generated by the proposed HMM-FLDA pipeline when implemented offline, i.e., the HMM-FLDA sleep/wake segmentation is computed assuming availability of the full time course (0-270 hours) human viral challenge study (HVC) data. Viral inoculation took place on day 2 and all viral shedders started shedding on day 4 or day 5. As compared to the non-shedders, the shedders tend to have a night sleep deficit until the 7th day. Compare to the total duration of sleep feature shown in 3.22 that includes duration of both night-time and daytime sleep.

3.6.5.3 Extensions of marginal clinical outcome predictor

Two extensions to the univariate regression framework for clinical outcome prediction are presented here. First a multivariate logistic regression and continuation ratio classifier are implemented on the top three univariate features for the full timecourse segmentation data. Then we apply a method that improves (univariate or multivariate) classifier performance for data with imbalanced class sizes.

Multivariate classifiers When the top three univariate logistic regression features in Table 3.9 (HR MED.sd, offset and total duration) are combined into a single model, the multivariate logistic model achieves AUC $= 0.844$, and the continuation-ratio model achieves AUC $= 0.885$ for classifying "Late shedder versus others." This represents only a moderate improvement relative to the corresponding marginal AUC's listed in the table.

Table 3.10 shows the Pearson correlation between the 3 features having maximum marginal area-under-the-curve (AUC) of the receiver operating characteristic curve (ROC) for classifying clinical outcome, presented in Section 3.4. The presence of significant correlation between feature over the population explains why multivariate LR and CR classifiers of clinical outcome can only attain marginally better AUC than do the univariate LR and CR classifiers.

Table 3.10: Pearson correlation coefficients (p-values) between top 3 features selected using AUC criterion for univariate logistic regression and continuation ratio classifiers of clinical outcome.

	Total duration	Offset	HR MED.sd
Total duration	1	0.778 (< 0.001)	0.434 (0.056)
Offset		1	0.485 (0.030)
HR MED.sd			1

Imbalanced class sizes : It is well known that imbalanced class size in the training data can negatively affect performance, especially for the under-represented classes. To mitigate the impact of these imbalanced shedding class sizes, we applied the synthetic minority over-sampling technique (SMOTE) [CBHK02] to rebalance the class sizes. SMOTE is a method for equalizing minority class sizes by introducing synthetic data samples that rebalances to match the majority class size. It does this rebalancing as follows. First a feature instance x is randomly selected from the minority class. Then for this point x the k minority class nearest neighbors $\{x_i\}_{i=1}^{k}$ are identified. One of these nearest neighbors x_R is selected as random and a random variable U is drawn from the uniform distribution over $[0, 1]$. Finally, the new SMOTE point in the minority class is defined as

$s = x + U * (x_R - x)$. This process is repeated until the minority class has as many samples as the majority class.

SMOTE can be applied to any classifier but can be expected to work best when the classification regions are convex in feature space. SMOTE is commonly applied to cases where the class imbalance is moderately small, e.g., the ratio of the smallest to largest class sizes is greater than 0.5. For the HVC data, Table 3.7 indicates that for the binary infection status outcome the imbalance is moderate (class size ratio $>$ 2/3) but for the ternery infection onset outcome the imbalance is severe (class size ratio can be as low as $1/3$).

First we considered compensation of class imbalance in the univariate logistic and continuation ratio regression predictors of shedding and shedding onset 24 hours before any subject starts to shed virus. SMOTE was applied using $k = 3$ nearest neighbors for each minority class except for the early onset shedder group where only 1 nearest neighbor was used since this group only contains 2 or 3 samples. Since SMOTE is a randomized algorithm, 100 SMOTE runs were performed. Recall that HR MED.sd was the top ranked feature for both LR and CR and for both the full data and the pre-infection data.

We implemented SMOTE for LR and CR on the univariate predictor using HR.MED.sd and on the multivariate predictor using HR.MED.sd, Total Duration and Offset. We considered both the offline case (sleep/wake segmentation was performed by FLDA using the full data, 0-270hrs) and the online case (sleep/wake segmentation was performed by FLDA using only the pre-infected data, 0-60hrs).

The first 3 rows and the last 3 rows of Table 3.11 shows the Accuracy (average classification error rate) and the AUC for logistic regression (LR), and the AUC for continuation ratio regression (CR) in the univariate case. The right columns of the table indicate significant AUC performance improvement for CR, for which class size is severely imbalanced (P-VALUE is the result of double sided paired t-test). In the case of the LR, on left side of table, the class imbalance is not as severe and SMOTE leads to no improvement in AUC or Accuracy for the univariate case. Indeed, SMOTE compensated univariate LR has worse Accuracy (p-value $<$ 0.01 according to double sided t-test) when either offline or online sleep/wake segmentation is used.

The middle 3 rows of the table show the results of applying SMOTE to multivariate logistic and continuation ratio regression for the offline case. For multivariate CR, SMOTE improves the AUC (p-value $<$ 0.01 according to double sided t-test). For multivariate LR, SMOTE improves the AUC and the Accuracy only in the offline case.

Table 3.11: SMOTE compensation for imbalanced sample size for predictors of shedding and shedding onset time 24 hours before first shedding occurs.

	Feature(s)			LR		CR AUC		
				Accuracy	AUC	Early	Late	No onset
Offline training (FLDA trained on 0-270hrs)	HR MED.sd (sleep)	w/o SMOTE		0.750	0.833	0.944	0.631	0.844
		SMOTE	MEAN	0.703	0.846	0.951	0.656	0.907
			P-VALUE	4.68E-11	0.015	1.38E-06	6.68E-07	2.80E-56
	Total duration + Offset + HR MED.sd (sleep)	w/o SMOTE		0.800	0.844	0.889	0.655	0.885
		SMOTE	MEAN	0.838	0.865	0.974	0.811	0.937
			P-VALUE	2.21E-21	6.78E-35	1.66E-71	3.28E-83	1.38E-75
Online training (FLDA trained on 0-60hrs)	HR MED.sd (sleep)	w/o SMOTE		0.842	0.758	0.882	0.718	0.864
		SMOTE	MEAN	0.781	0.727	1.000	0.751	0.899
			P-VALUE	8.23E-13	3.00E-03	-	5.98E-06	1.69E-25

CHAPTER 4

Cognitive Performance Variability is Associated with Susceptibility to Infection

4.1 Introduction

Cognitive function and other psychological factors (e.g. stress) have long been associated with physiological health. In particular, reaction time, vigilance and processing speed are central to the human ability to perform optimally. Accumulating evidence suggests that intra-individual variability in reaction time (and other cognitive domains) may reflect neurobiological disturbance and have valuable prognostic significance [Sal07]. Higher variability of reaction time has been associated with greater mortality over 19-years of follow up in both younger and older adults as well as risk for falls and neurodegenerative disorders [SDTD06, HBB17].

Cognitive function is also closely linked to immune health and there is increasing recognition that immune cells play a physiological role in cognition and stress response [KGD12]. For example, T-cells have been reported to have a pro-cognitive effect and neurotransmitters involved in the immune response, such as acetylcholine, dopamine and noradrenaline, also play a key role in cognition [KGD12]. In healthy aging adults, elevated concentrations of pro-inflammatory cytokines has been linked to worse cognition [SMRS$^+$20]. This relationship is further reflected by the fact that many of the same factors that impair immune response (e.g. sleep deprivation, stress, alcohol consumption, depression, infections) also impair cognitive performance.

We therefore tested the hypothesis that intra-individual variability in vigilance and reaction time measured over a 3 day baseline, reflecting subtle changes in immune and brain health, would predict vulnerability to a common infection.

More specifically, in the context of a longitudinal human viral challenge study we establish associations between pre-exposure cognitive function and post-exposure immune response, as measured by various markers, such as severity of symptoms and viral shedding. Among the many pre-exposure cognitive markers that we study we find that it is a new measure, called the cognitive performance variability (CPV), that is the most correlated to immune response. The CPV is extracted from a person's performance on a sequence of NeuroCognitive Performance Tests over several days

leading up to exposure. The NeuroCognitive Performance Test (NCPT) uses 4 interactive games to measure a person's cognitive performance along 18 dimensions. These 18 NCPT variables are shown in Table 4.1a). The CPV score is a measure of the person's cognitive dissonance over time along any dimension. Unlike other measures of variability, like the linear coefficient of variation (CoV), the CPV is a non-linear max-pooled measure of variability of the NCPT variables.

Our findings rely on analysis of a week-long viral challenge study in which cognitive function of non-quarantined volunteers were monitored three times a day over a one week period using NCPT and other cognitive tests. On the fourth day the participants were inoculated with a viral pathogen; the common cold (HRV). The CPV was computed from NCPT data collected three times a day over the pre-exposure time period and the amount of viral shedding was computed from PCR data collected daily over the post-exposure time period.

The main contribution of this paper is the demonstration that a certain kind of cognitive variability measure, the aforementioned CPV score computed from pre-exposure data, has an uncommonly strong association with a subject's amount of post-exposure viral shedding. Our findings are supported on several pillars. First, we find that the Pearson correlation of CPV and shedding is near 0.9, that the R^2 linear regression goodness of fit is near 0.8, and the AUC for detecting post-exposure shedding above the median is near 1. The NCPT scores that contribute the most to the CPV are the Posner Tutorial Time, DigSym Time, DigSym Correct, Trail Time, and Reaction Time. The subjects that have high variability in at least one of these scores tend to have high levels of viral shedding after exposure.

As a point of comparison, the proposed CPV score achieves significantly better performance than that attainable using the standard coefficient of variation (CV), whose correlation with shedding is less by factor of two (Pearson correlation -0.42 compared to 0.88). We find that there is little value in supplementing the NCPT variables with other cognitive variables in the CPV score. These variables included pre-exposure sleep duration and the visual analog fatigue score (VAFS). This finding is consistent with the fact that the NCPT-only CPV score is correlated with these other variables (Figure 4.10 in supplementary material).

To probe the effect of reducing the number and timing of the NCPT test sessions, we performed a combinatorial cross-validation study of the influence of both the number of sessions and session timing on the CPV. Our findings indicate that CPV vs shedding correlation above 0.69 can be maintained with as few as $T = 5$ sessions as long as there is at least one session per day. This suggests that cognitive performance variability markers may have practical clinical and epidemiological application.

In Section 4.2 we state our main findings, followed by a Materials and Methods Section 4.3 where we present details on our methods. We continue with a Discussion Section 4.4 where we discuss these results and methods. Finally in the Conclusions Section 4.5 we summarize the paper

and comment on future work. The Supplementary Material 4.6 accompanying this paper contains further data on our methods and findings, as discussed below.

4.2 Results

A longitudinal viral challenge study was performed in 2015 in which 18 human volunteers participated over a period of 9 days. On the fourth day of the study participants were inoculated with HRV virus (the common cold) and the participants' daily viral shedding was collected for the remainder of the study. The cognitive function of the volunteers was collected three time per day over the pre-exposure days and the time series of 18 NCPT variables listed in Table 4.1a was transformed to a CPV score for each participant (see Materials and Methods).

Our main results are shown in Figure 4.1 which shows each participant's CPV scores computed after omitting the initial NCPT session (screening session). Figure 4.1c shows a remarkably strong association between total amount of post-exposure shedding and cognitive performance variability, as measured by the CPV score, for each individual. The Pearson correlation is 0.88 with 95% confidence interval $[0.69, 0.95]$. A linear regression of total shedding onto the CPV score gave an R^2 of 0.77 . Furthermore, a logistic regression of total shedding onto the CPV score yielded a perfect discriminant between high and low shedders, respectively defined as those whose total shedding is below versus above the population median.

To illustrate the role of the 18 individual NCPT variables in the CPV, we plot in Fig. 4.1d the univariate CPV scores for the two lowest shedding and the two highest shedding participants. This figure is extracted from Figure 4.16 in the Supplementary Material that shows the sequence of univariate CPV scores for all 18 study participants. Superimposed on the plot of these variables is a boxplot indicating score sensitivity to session perturbation, determined by leave-one-out analysis where the univariate CPV was recomputed after successively leaving a single NCPT session out of each participant's sequence (sans screening session). Figure 4.1d clearly shows that certain NCPT variables have significantly higher variability for the high shedders (lower two panels) than for the low shedders (top two panels). Note that the NCPT variable with highest variability (variable achieving peak score in each panel of Figure 4.1d) differs across study participants.

To explore the effect of reducing the number of cognitive testing sessions and their timing, we performed a combinatorial study of the association between shedding and CPV as we vary both the number of NCPT sessions and their associated timing patterns over the baseline time period. As the number of sessions ranging from $T = 3$ to $T = 10$, Fig. 4.2 shows the top ten patterns, ordered in decreasing Pearson correlation, and their associations to infection severity as measured by correlation, R^2 and AUC. The results of the full combinatorial study are given in Fig. 4.17, and the zoomed-in figure, Fig. 4.18, in the Supplementary Materials.

The connection between NCPT variables and gene expression would be interesting to investigate. We performed a Pearson correlation test to screen the genes significantly correlated with NCPT variables. The number of significant genes and the name of gene with smallest p-value is included in Table 4.19. There are $3,726$ genes significantly correlated with digSym-error, far more than reaction-time, which ranked second with 155 correlated genes. 6 NCPT variables is not significantly correlated with any gene. ADGRG7 has been repeated found to be the most significantly correlated gene with 6 NCPT variables. ADGRG7, GLYATL2 and PNMAL1 are also correlated with 7 NCPT variables. A pathway enrichment analysis was performed to understand the genes significantly correlated with the NCPT variables. The result is shown in Table 4.20. The most significant pathway is Phagosome with FDR 3.99e-3. Many members in MAPK family are targeted in the enriched pathways.

4.3 Materials and Methods

4.3.1 Challenge study protocol

Viral shedding and cognitive performance data were collected from a human rhinovirus (HRV) challenge study. A total of 24 volunteers were recruited and 19 participated in the study. One of these participants had a failed inoculation and was omitted from our analysis. The age range of the remaining 18 participants was between 18 and 23, two thirds of these participants were male, and 4 were non-caucasian. For a more detailed demographic summary see Fig. S1 of the Supplementary Materials. The study was reviewed and approved by the Institutional Review Boards at Duke University and at University of Virginia . Written informed consent was obtained from all participants. Exclusion criteria included pregnancy, chronic respiratory illness, high blood pressure, tobacco/drug/alcohol history, and serum antibody levels above 1:4 titers. The participants were not isolated during the study.

The challenge study lasted 9 days over which various types of biomarkers were continuously collected from participants using wearable wristbands (Empatica E4), whole blood assays (mRNA, p180 metabolites, steroids), nasal-pharyngal washes (viral shedding), cognitive testing (Lumos), and self-reported clinical data (symptoms). On day four at 8am (80 hours) each participant was inoculated via intranasal drops of diluted Human Rhinovirus strain type 16 with a dose of 100TCID50 in 1mL Lactated Ringer's Solution.

After inoculation the participants underwent daily nasal lavage each morning to determine the amount of viral shedding. Figure 4.3a shows the total amount of shedding for each participant over all post-inoculation study days, ordered from maximum to minimum. Cognitive scores were collected from test sessions 3 times per day in the morning, mid-day, and in the evening (Fig. 4.3b). This data was also collected from a reference session prior to the start of the study. In each session

the participant answered web-based questionnaires and engaged with Lumos brain testing software using computer tablets that were provided to them.

4.3.2 NeuroCognitive Performance Test

The NeuroCognitive Performance Test (NCPT) is a repeatable, web-based, computerized, cognitive assessment platform designed to measure subtle changes in performance across multiple cognitive domains [MSNH15]. It comprises of 18 subtests and the modular platform allows for customized subtest batteries for specific studies. It was formerly referred to as the Brain Performance Test. As such computerized tests may offer several advantages over traditional paper and pencil methods, such as greater consistency in administration and scoring, generation of alternate forms for repeated testing, precise stimulus control, ability to capture and analyze multiple components of a test taker's response, adaptation of difficulty levels, greater convenience and ability to administer at different settings. Test reliability and concurrent validity of the NCPT for unsupervised administration has been previously published. Specifically, the authors of [MSNH15] reported normative data for more than 130,000 individuals aged 13–89 years as well as data on the ability of NCPT to detect mild cognitive impairments.

The specific NCPT battery used in the study comprised of four subtests designed to measure attention, processing speed, response inhibition and cognitive load (task switching and executive function) – domains known to be sensitive to fatigue, stress and infections [Smi12]. The brief battery (15 minutes) was designed to be easy to complete and included the four subtests described below:

1. Attentional Cueing (Posner): A measure of selective attention and processing speed. An arrow cue is shown followed by a stimulus placed in one of 2 locations. Subjects pick the correct location of the stimulus.

2. Digital Symbol Coding: A measure of attention/vigilance, speed and immediate memory. Subjects enter the number corresponding to randomnly generated symbols using a key at the top of the screen in 90 seconds. The primary measure is number of correct responses minus number of incorrect responses.

3. Go/No-Go: A measure of response inhibition and processing speed. Participants were required to respond as quickly as possible to a target, but to avoid responding to distractions.

4. Trail Making B: A measure of executive function, speed and mental flexibility. Subjects connect the numbers from smallest to largest alternating between numbers and letters. The primary measure is completion time and there is no time limit.

These 4 subtests yield 18 scores related to speed, accuracy and congruency. The tests were administered at 10 time points across 3 days at baseline. Raw scores on all 18 tests across all 10 time points were used to compute the cognitive variability indices.

In addition to NCPT, several well established self-reported psychometric markers were measured at various times in the study. This included responses to fatigue related questions using two protocols: the Visual Analog Fatigue Scale (VAFS), measured 3 times per day; and the Fatigue Severity Scale (FFS), measured at screening and on the fourth day of the study. The VAFS is a response to a single question scoring fatigue from 10 (no fatigue) to 0 (severe fatigue), while FFS is comprised of responses to 9 fatigue-related questions. A large scale clinical validation study of these measures of fatigue was reported in [KLMNS89]. The Perceived Stress Scale (PSS) was used to measure stress at the initial screening session. The PSS is an instrument that measures a person's perceived stress over the past month consisting of 10 questions about stress on a scale of 0 to 4, which has been clinically validated in [CKM83]. Finally, the reduced Composite Scale of Morningness (rCSM) was used to measure an individual's chronotype. The rCSM consists of a subset of 7 questions from the set questions of the full CSM [SRM89] on the most productive part of the day. The rCSM has been clinically validated in [Ran09].

4.3.3 Cognitive Performance Score (CPV)

The CPV is computed as follows. For a particular participant i and cognitive variable j, e.g., NCPT *reaction-time*, let the value of the variable at the m-th session time be $x_{ij}(m)$, $m = 1, \ldots, N_{ij}$, where N_{ij} are the number of time samples, e.g, number of NCPT sessions, prior to inoculation time. In Eq. [4.1] we define the statistic $T_{ij} \in [0, 1]$. In this equation, $\overline{x}_{ij}^{+} = (N_{ij} - 1)^{-1} \sum_{m=2}^{N_{ij}} x_{ij}(m)$ and $\overline{x_{ij}^{-}} = (N_{ij} - 1)^{-1} \sum_{m=1}^{N_{ij}-1} x_{ij}(m)$ are sample means of left and right truncated versions of the sequence $\{x_{ij}(m)\}_{m=1}^{N_{ij}}$, respectively.

The univariate CPV for the i-th subject and j-th variable is defined as the quantity:

$$CPV_{ij} = -\log\left(1 - \frac{B\left(T_{ij}, \frac{N_{ij}-1}{2}, \frac{N_{ij}-2}{2}\right)}{B\left(\infty, \frac{N_{ij}-1}{2}, \frac{N_{ij}-2}{2}\right)}\right)$$

where $B(x; a, b)$ is the incomplete beta function

$$B(x; a, b) = \int_0^x t^{a-1}(1-t)^{b-1} dt.$$

The CPV for the i-th participant is defined as the maximum

$$CPV_i = \max_j CPV_{ij}.$$

The CPV can be interpreted as the $-log$ of the reported p-value of a significance test of change of mean in successive pairs of values $x_{ij}(m), x_{ij}(m-1)$, $m = 2, \ldots, N_{ij}$. Specifically, CPV_{ij} can be obtained by applying the one-sided ANOVA significance testing procedure [Sch99] to the columns of the matrix

$$\mathbf{X}_{ij} = \begin{bmatrix} x_{ij}(1) & \ldots & x_{ij}(N_{ij}-1) \\ x_{ij}(2) & \ldots & x_{ij}(N_{ij}) \end{bmatrix}.$$

This mathematical equivalence allows us to compute the CPV using standard ANOVA software (Matlab R2020a `anova1.m`).

4.4 Discussion

Empirically we see from the heatmap in Fig. 4.1 that five of the 18 cognitive scores contribute the most to the CPV: digSym-correct, digSym-time, reaction-time, trail-time, and posner-tutorialTime. It is interesting that 4 out of the 5 are time-to-completion features and only only one of them is a correctness feature.

The game-playing learning curves cannot alone explain the high associations between CPV and total shedding. A learning curve of each participant was extracted and it had no significant correlation to total shedding (Supplementary Materials).

The CPV was much more highly correlated with shedding titers than are other factors that have been previously related to susceptibility to infection, e.g., self-reported fatigue (VAFS), self-reported stress, or hormonal stress indicators (cortisol). See the correlation matrix Fig. 4.10 in Supplementary Materials. Furthermore, no significant correlation between CPV and time-of-day was observed.

The combinatorial study shown in Fig. 4.2 indicates interesting structure in the session patterns that guarantee high correlation between the CPV and viral shedding. As might be expected, the association tends decrease as the number of sessions decreases, and when the early initial screening session at time $T = 1$ is omitted. Furthermore, eliminating late sessions that are closer to exposure time (T closer to 10 than to 1) tend to degrade the high association more than eliminating early sessions. Interestingly, up to two successive sessions can be eliminated (roughly corresponding to a gap of 16 hours between tests) without reducing correlation below 0.69. In fact, for $T = 7$ the best way to omit sessions is to remove the session at $T = 1$ and eliminate the other two successively between $T = 3$ and $T = 9$. A correlation greater than 0.69 (AUC > 0.8) is attainable even for as few as $T = 5$ sessions as long as the they are distributed such that there is at least one test on each of the three pre-exposure days.

The association between NCPT variables and gene expression is intriguing. One of the most correlated genes, ADGRG7, encodes G protein-coupled receptor 128 (GPR128), a member of adhesion G protein-coupled receptor. Although there is no direct evidence connecting GPR128

and cognitive function, another member from the same family, GPR110, was reported to have a role in cognitive function [LHK$^+$16]. Many of the enriched pathways contain MAPKs and dysregulation of the RAS/MAPK signaling cascade has been reported to have a role in cognitive profile of disorders [CAP$^+$09]. These findings suggest an interesting temporal connection between NCPT variables and gene expression, and this observation might shed light on using non-invasive cognition measurements to replace invasive and expensive gene expression test in certain cases.

These findings raise the intriguing possibility that periodic cognitive testing for assessing susceptibility to severe infection may have clinical and/or epidemiological value. However, there are several factors that might impede translation of our results to the clinic or to public health. First, continuous testing over time would be necessary as the time of viral exposure cannot be anticipated. This could possibly be overcome if our results extended to a rolling window version of CPV and to more convenient cognitive stimuli than NCPT, e.g., derived from interactions with a cell-phone. Second, the effect of a person's adaptation (learning curve) to the cognitive stimuli (NCPT) could not be determined in the relatively short 3 day baseline of our study. Third, it is unknown whether our results would replicate for a different pathogen or for a cohort of participants more representative of the general population (age, occupation, medical history).

As in any observational study, there are limitations to our findings. These include: 1) The small sample size of the study, giving fairly wide (> 0.2) confidence intervals on the 0.88 correlation reported; 2) Deviations of individual participant behaviors from the cognitive testing protocol: 4 of the 18 participants completed fewer than 10 NCPT sessions and 2 participants did not participate in the NCPT early screening session. Furthermore, some participants did not abide strictly to the NCPT session timing (early morning, mid-day and evening). 3) As an observational study, we can't establish that CPV or any of its correlates are causal factors for increased viral shedding; but the reported associations suggest that cognitive performance variability deserves further study in the context of disease susceptibility.

4.5 Conclusions

Using data from a 9 day viral challenge study, this paper established a strong association between pre-exposure variability of cognitive function and severity of infection, as measured by total viral shedding after a person's exposure to the common cold. A person's cognitive variability over time was measured using thrice daily NCPT brain testing, quantified with a score function related to a one sided ANOVA test of significance. Our results suggest that regularly collected cognitive performance markers may be useful for predicting shedding susceptibility, with potential clinical and epidemiological application.

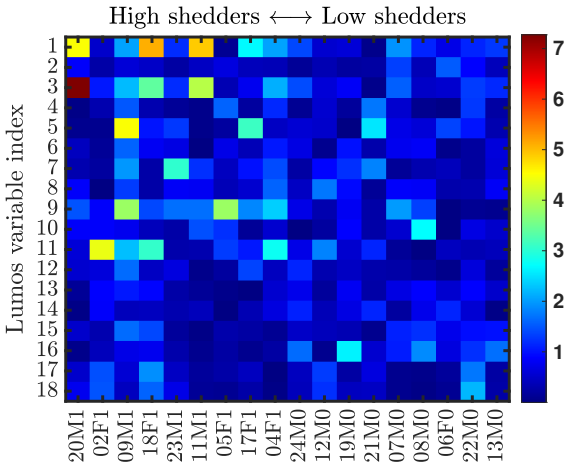
It is to be emphasized that the proposed cognitive performance variability (CPV) score is a fixed

function without any tunable parameters. Such a parameter-free score does not require fitting any model to population, unlike regression-based scores, and thus there is no overfitting error. However, if we had access to a larger sample population or a longer baseline for training, it is possible that we could improve on the CPV score by introducing some parameters. For example, we could fit a regression model with variable selection to the population, selecting the most important NCPT variables along with the regression coefficients. As another example, with a longer baseline, a temporal dependency weighted CPV model might be fitted to each subject, e.g., accounting for the effects of learning curves and circadian fluctuations.

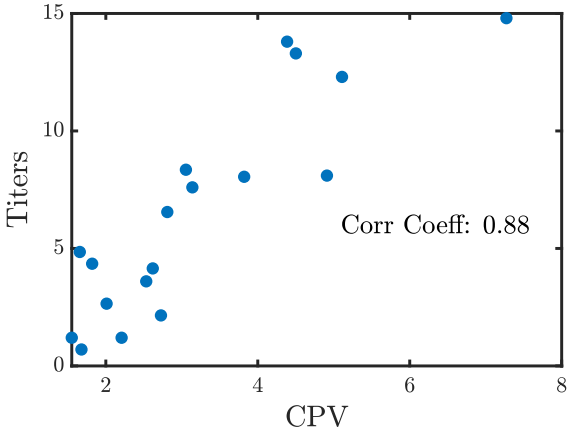
Index	Variable name	Index	Variable Name
1	digSym-correct	10	posner-tutorialError
2	digSym-error	11	posner-tutorialTime
3	digSym-time	12	posner-correct
4	reaction-error	13	posner-responseTime
5	reaction-time	14	posner-timeCongruent
6	trail-layoutNum	15	posner-timeIncongruent
7	trail-tutorialTime	16	posner-totalTime
8	trail-error	17	posner-Ncongruent
9	trail-time	18	posner-correctCongruent

20	Participant ID (1-24)
M	Participant gender (M/F)
1	Participant low/high shedding status (0/1)

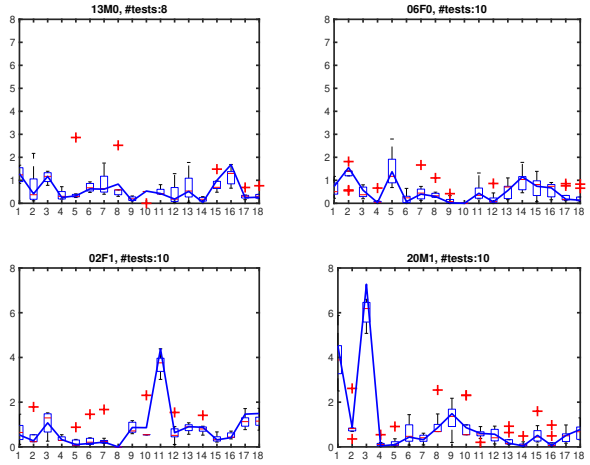
(a) Top: NCPT cognitive variables. Bottom: participant key.



(b) Heatmap of 18 univariate CPV scores vs 18 participants.

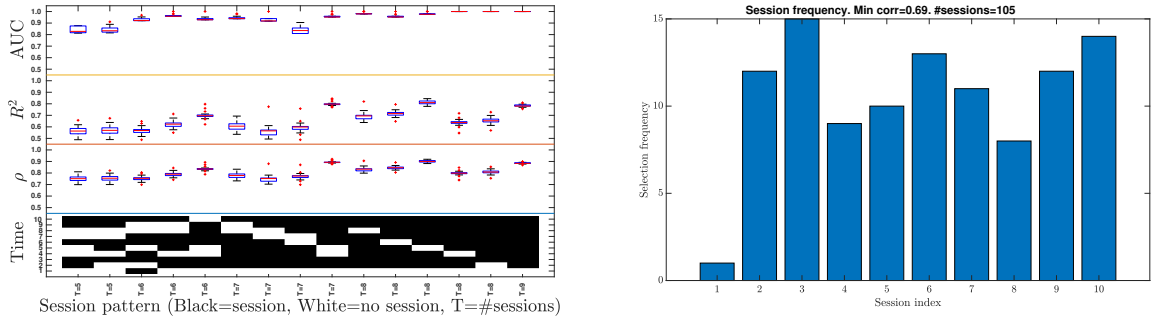


(c) Post-exposure shedding vs pre-exposure CPV score.



(d) Univariate CPV scores for low (top) and high (bottom) shedders.

Figure 4.1: Pre-exposure cognitive performance variability (CPV) is strongly associated with post-exposure viral shedding. (a) Definition of 18 NCPT variables measuring subject performance on 4 different games: Digital Symbol Coding (DigSym), Go/No-Go (Reaction), Trail Making (Trail) and Attention Cuing (Posner). (b) Heatmap of univariate CPV scores for each study participant (ordered from highest to lowest amounts of post-exposure shedding). (c) Scatterplot of CPV and shedding (Titer) for all 18 participants taking into account all post-screening and pre-exposure NCPT sessions. The high Pearson correlation of 0.88 suggests that the CPV score might be used as a marker of a person's susceptibility to severe infection and that periodic cognitive testing might be used for early prediction of infection severity. (d) Boxplots of the univariate Cognitive Performance Variability scores over the 18 NCPT variables in Table 4.1a for two of the lowest shedding (top row) and two of the highest shedding (bottom row) challenge study participants. The Solid blue curve denotes the CPV scores computed for each NCPT variable using the full session sequence, but omitting the reference session (earliest sessions Fig. 4.3b). Red boxplots show the distribution of the leave-one-out scores computed by successively dropping a single session from the sequence of post-screening sessions.

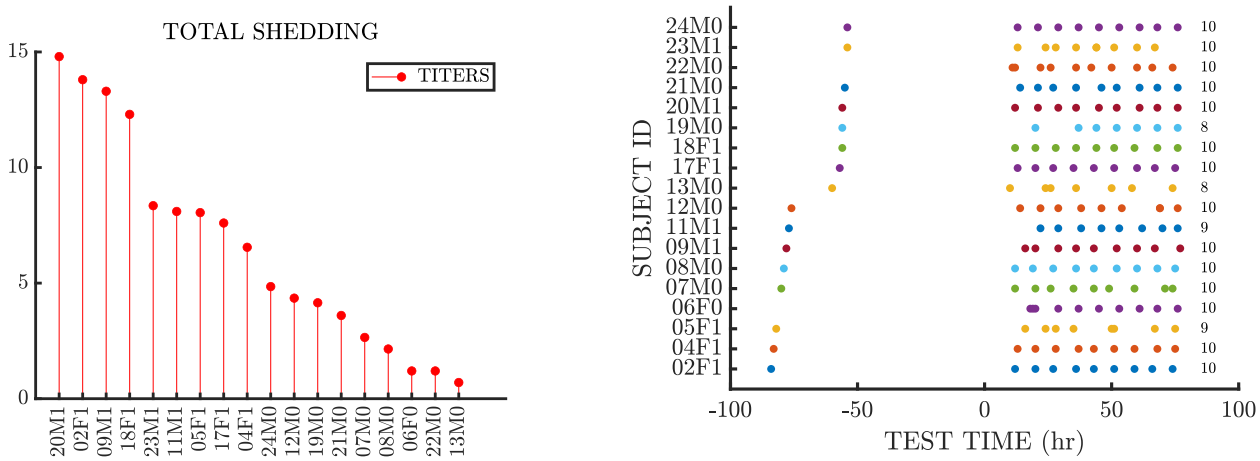


(a) CVP association vs. session pattern.

(b) Session selection frequency.

Figure 4.2: Lumos session timing patterns corresponding to the 15 highest CPV vs shedding correlations. Left: Boxplots of association measures between CPV and shedding for which the cross-validated (leave-one-subject-out) CPV Pearson correlation coefficients are at least 0.69 (the lower endpoint of 95% CI of correlation in Fig. 4.1c). The measures of association are the Pearson correlation coefficient ρ, the R^2 of linear regression, and the AUC of logistic regression of titers with respect to CPV. The AUC measures the association of CPV with Low vs High shedding (0,1) labels, where Low and High denote shedding below or above the population median, respectively. The pattern heatmap at bottom indicates the corresponding timing patterns of Lumos sessions with Time 1 corresponding to the initial screening and Time 10 corresponding to the test right before exposure. The number of sessions in each pattern is denoted by T, fewer than $T = 7$ sessions significantly reduces the association. Right: relative frequency of inclusion of particular session times attaining correlation > 0.69.

$$T_{ij} = \frac{(N_{ij}-2)\sum_{m=2}^{N_{ij}}(x_{ij}(m)-x_{ij}(m-1))^2}{(N_{ij}-2)\sum_{m=2}^{N_{ij}}(x_{ij}(m)-x_{ij}(m-1))^2+(N_{ij}-1)\sum_{m=2}^{N_{ij}}(x_{ij}(m)-\overline{x}^{+}_{ij})+(x_{ij}(m-1)-\overline{x}^{-}_{ij}))^2} \quad (4.1)$$



(a) Postexposure viral shedding.

(b) Timing of pre-exposure NCPT sessions.

Figure 4.3: Total post-exposure viral shedding and pre-exposure NCPT session timing. (a) Timing of the pre-exposure NCPT testing sessions for each of the 18 challenge study participants during the baseline part of the challenge study (0 to 80 hrs), which precedes exposure to the HRV pathogen. Appearing on the far left of the figure are the subjects' initial screening sessions which, for all but two participants, occurred several days before the start of the study. At far right of the figure is shown the total number of sessions, varying between 8 and 10, for each participant. (b) Total amount of viral shedding accumulated from the time of exposure to the end of the study, indexed over subjects ID's. The subject ID encodes the subject index (numeric first two characters from 1 to 20) the subject gender (M or F second character) and whether the participants's shedding is below or above (0 or 1) the population median of total viral shedding (5.7 titers).

4.6 Supplementary Materials

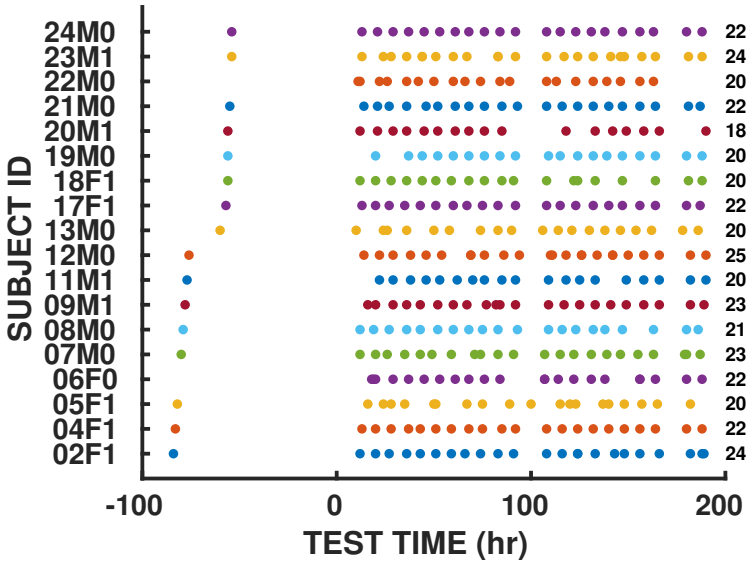


Figure 4.4: Timing of sessions over the entire study (pre- and post-exposure) for the 18 participants. The time axis is indexed over hours where 0h corresponds to 12AM of the first day of the challenge study and 80h corresponds approximately to the time of viral inoculation on the morning of the 4th day (8am). Figure 4.3b only shows the pre-exposure sessions.

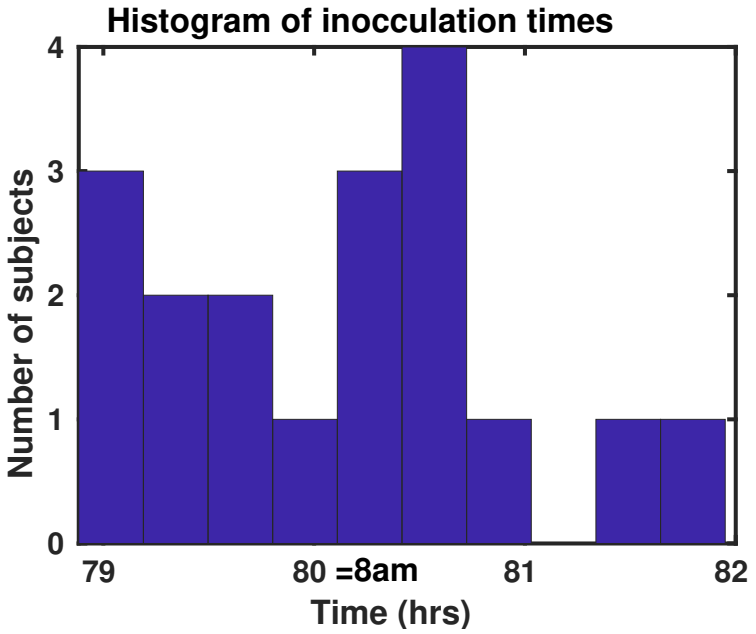


Figure 4.5: Histogram of the inoculation times for the 18 participants. Eighty hours corresponds to 8am on the morning of the 4th day of the challenge study. All but 2 participants are inoculated between of 7 and 9am.

		Low shedder	High shedder	Total
Median Age (range)		21 (20-24)	20 (19-33)	20(19-33)
Gender	Male	8	4	12
	Female	1	5	6
Ethnicity	White	7	7	14
	Black	2	0	2
	Asian	0	2	2
Dominant hand	Right	9	9	18
	Left	0	0	0
Median Systolic BP (range)		120 (100-130)	110(100-120)	112(100-130)
Median Diastolic BP (range)		70 (60-80)	70 (58-79)	70 (58-80)
Median Pulse (range)		64 (60-72)	72 (60-80)	67 (60-80)
Median Respiratory Rate (range)		20 (16-20)	18 (16-20)	19 (16-20)
Median Weight (range)		71 (61-89)	65 (55-73)	67 (55-89)
Getup time	before 6:30	0	1	1
	6:30-7:45	1	0	1
	7:45-9:45	5	6	11
	9:45-11	3	1	4
	11-12	0	1	1
Bed time	10:15-12:30	5	5	10
	12:30-1:45	4	2	6
	1:45-3	0	2	2
Productive time	Morning	4	3	7
	Afternoon	3	1	4
	Evening	2	3	5
	Night	0	2	2

Figure 4.6: Study demographics (upper 3 rows), physiological statistics, and self-reported biochronicity (morningness) statistics (lower 3 rows) collected during screening for the 18 subjects included in our analysis.

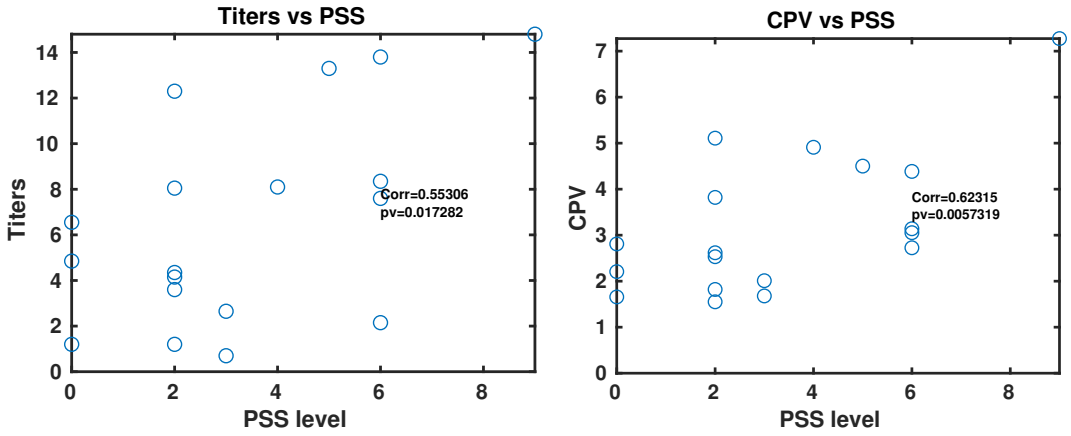


Figure 4.7: Scatter plots showing association between cognitive data collected at screening, the perceived stress scale (PSS) with the amount of viral shedding (left) and with the pre-exposure cognitive performance variability (CPV) over baseline.

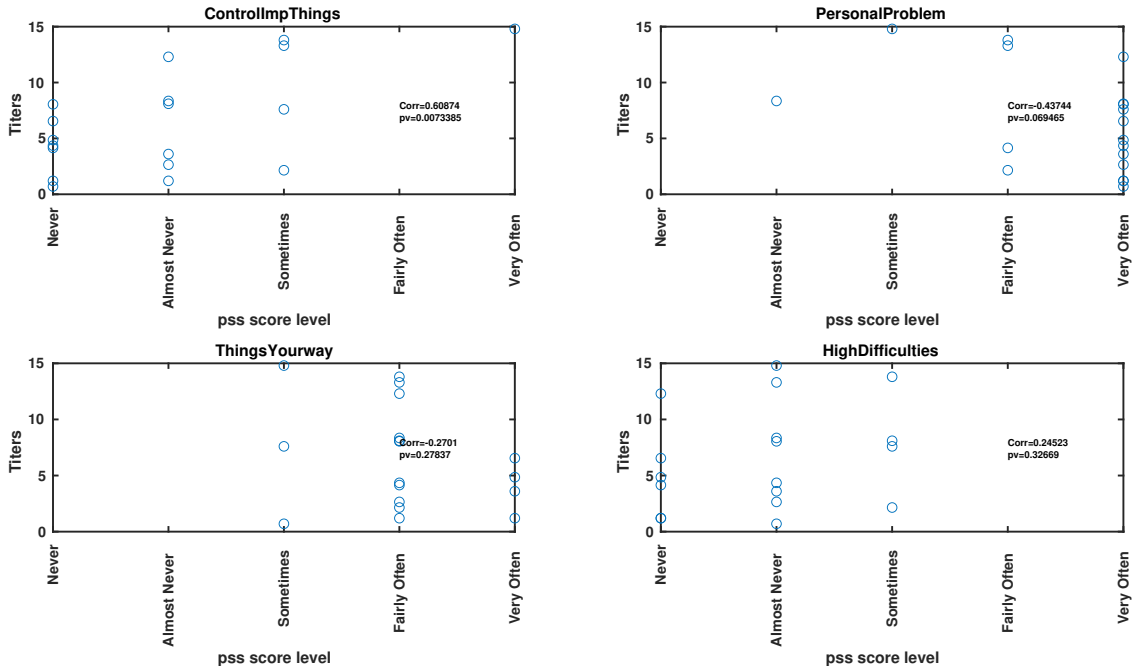


Figure 4.8: Scatter plots of viral shedding and four perceived stress variables self-reported by study participants during screening.

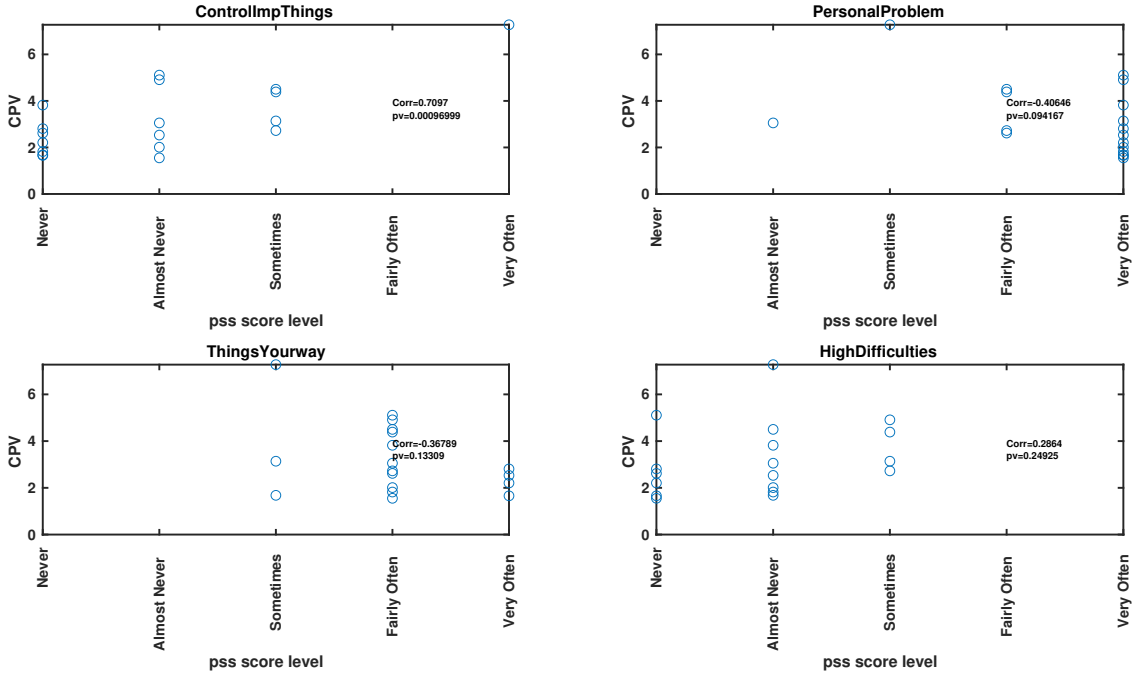


Figure 4.9: Scatter plots of proposed CPV score constructed over the pre-exposure time interval and four perceived stress variables self-reported by study participants during screening.

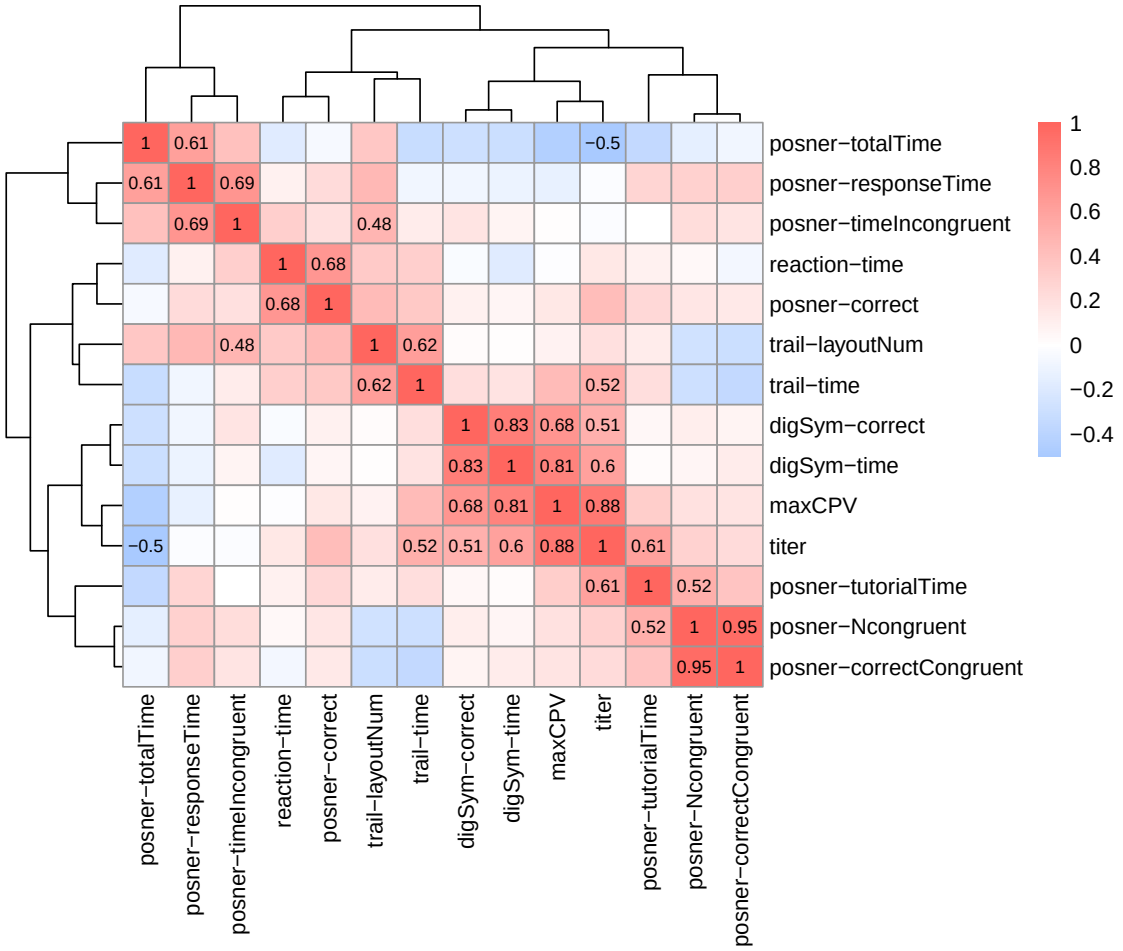


Figure 4.10: Heatmap of correlations between univariate CPV's of each NCPT variable, CPV (the maximum of the CPV's) and viral titer. Six NCPT variables do not have significant correlations with any other NCPT variable, titers, or CPV and are omitted. These are digSym-error, reaction-error, trail-tutorialTime, trail-error, posner-tutorialError, posner-timeCongruent. Statistically significant correlations (Fisher test at 0.05 level) are indicated. Besides CPV, 6 other univariate CPVs are also significantly correlated with titer, including the CPV of digSym-correct and digSym-time, which form a highly correlated block with CPV and titer. The significant pairs along with their confidence interval and p-values are included in the Table 4.1.

Variable 1	Variable 2	correlation	p_value	CI
CPV	titer	0.88	1.90E-06	[0.69,0.95]
CPV	digSym-time	0.81	5.18E-05	[0.55,0.93]
CPV	digSym-correct	0.68	1.75E-03	[0.32,0.87]
titer	posner-tutorialTime	0.61	7.78E-03	[0.19,0.84]
titer	digSym-time	0.6	8.09E-03	[0.19,0.83]
titer	trail-time	0.52	2.70E-02	[0.07,0.79]
titer	digSym-correct	0.51	3.22E-02	[0.05,0.79]
titer	posner-totalTime	-0.5	3.27E-02	[-0.79,-0.05]
posner-Ncongruent	posner-correctCongruent	0.95	9.55E-10	[0.88,0.98]
digSym-correct	digSym-time	0.83	1.85E-05	[0.6,0.94]
posner-responseTime	posner-timeIncongruent	0.69	1.51E-03	[0.33,0.88]
posner-responseTime	posner-totalTime	0.61	6.77E-03	[0.21,0.84]
posner-tutorialTime	posner-Ncongruent	0.52	2.84E-02	[0.06,0.79]
reaction-time	posner-correct	0.68	1.93E-03	[0.31,0.87]
trail-layoutNum	trail-time	0.62	5.72E-03	[0.22,0.84]
trail-layoutNum	posner-timeIncongruent	0.48	4.29E-02	[0.02,0.77]

Table 4.1: Significantly correlated CPV pairs. Titer is also added for analysis.

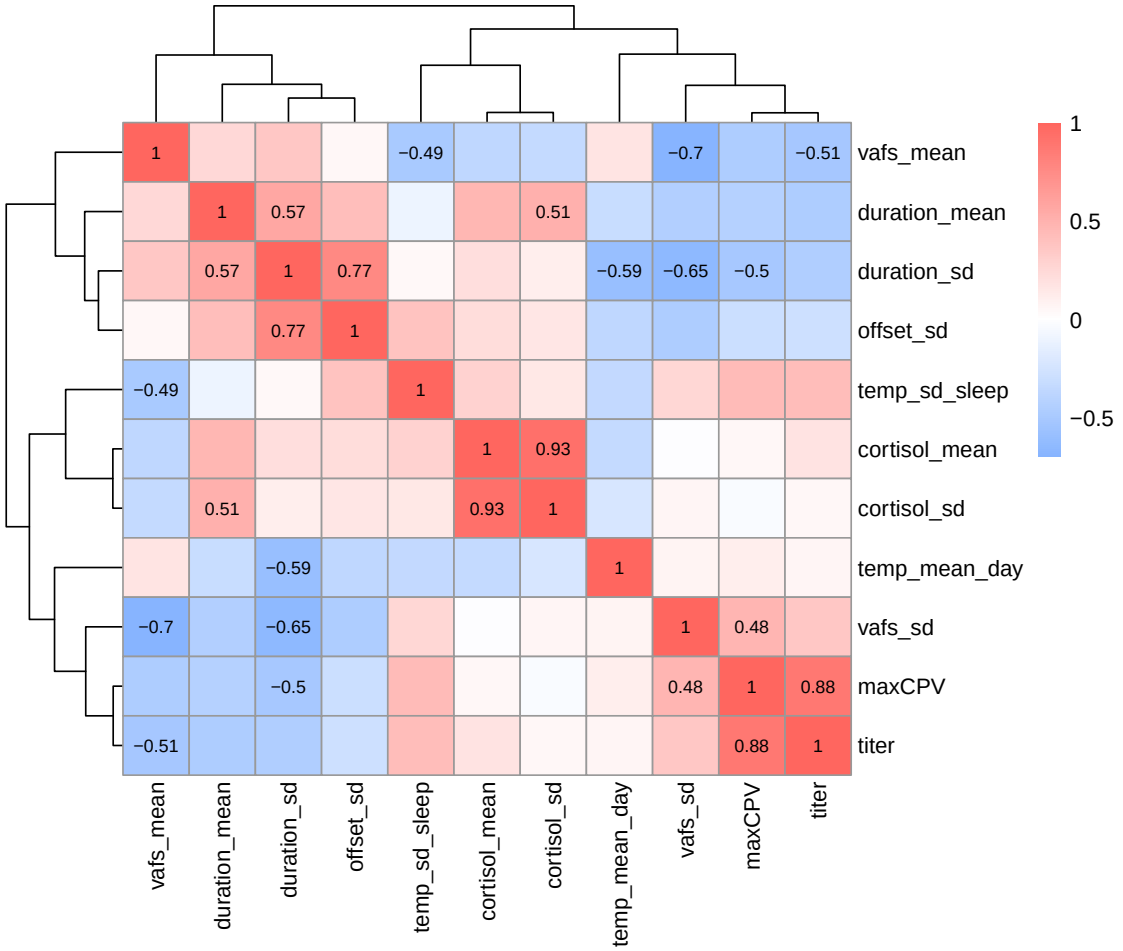


Figure 4.11: Heatmap of correlations between CPV and some of the other variables continuously collected during the challenge study. 5 variables are omitted in this figure as they are not significantly corrected with any other variable (digSym-error, reaction-error, trail-tutorialTime, trail-error, posner-tutorialError, posner-timeCongruent)(vafs_p, offset_mean, temp_mean_sleep, temp_sd_day, cortisol_JTK). Significant correlations are annotated with numbers. Observe that the CPV is not significantly correlated (Pearson p-value < 0.05) with the self-reported fatigue scores VAFS (VAFS_mean) or cortisol levels/periodicity. CPV is significantly correlated with the variation in sleep duration (duration_sd), which was collected from the Empatica E4 wearable device using sleep/wake segmentation algorithm [SZH+20]. Information on the statistically significant pairs or correlations are included in Table 4.2.

Variable 1	Variable 2	ρ	p_value	CI
CPV	titer	0.88	1.90E-06	[0.69,0.95]
CPV	PSS	0.62	5.73E-03	[0.22,0.82]
CPV	duration_sd	-0.5	3.91E-02	[-0.79,-0.03]
CPV	vafs_sd	0.48	4.27E-02	[0.02,0.77]
CPV	vafs_mean	-0.46	5.3E-02	[-?,-?]
titer	vafs_mean	-0.51	2.98E-02	[-0.79,-0.06]
cortisol_mean	cortisol_sd	0.93	3.52E-08	[0.81,0.97]
duration_sd	offset_sd	0.77	3.06E-04	[0.46,0.91]
vafs_mean	vafs_sd	-0.7	1.36E-03	[-0.88,-0.34]
vafs_sd	duration_sd	-0.65	5.00E-03	[-0.86,-0.24]
duration_sd	temp_mean_day	-0.59	1.23E-02	[-0.84,-0.16]
duration_mean	duration_sd	0.57	1.67E-02	[0.12,0.83]
duration_mean	cortisol_sd	0.51	3.84E-02	[0.03,0.79]
vafs_mean	temp_sd_sleep	-0.49	4.42E-02	[-0.79,-0.02]

Table 4.2: Table of significantly correlated biomarkers (Fisher test at level 0.05) including CPV, titers, and NCPT variables.

```
Linear regression model:
    Titers ~ [Linear formula with 4 terms in 3 predictors]

Estimated Coefficients:
                           Estimate       SE        tStat       pValue
                           ________    ________    ________    __________

    (Intercept)            -0.64837      2.5474    -0.25452       0.80279
    digSym_time             0.35812    0.078446      4.5652    0.00044069
    trail_time              0.30269     0.12179      2.4853      0.026199
    posner_tutorialTime     0.49132     0.11325      4.3384    0.00068112


Number of observations: 18, Error degrees of freedom: 14
Root Mean Squared Error: 5.51
R-squared: 0.805,  Adjusted R-Squared: 0.763
F-statistic vs. constant model: 19.2, p-value = 3.12e-05
```

Figure 4.12: Stepwise linear regression of viral titers on the NCPT covariates listed in Table 4.1a. The goodness of fit R^2 to titers is no better than obtainable using CPV as the sole predictor variable, which only involves fitting a single parameter.

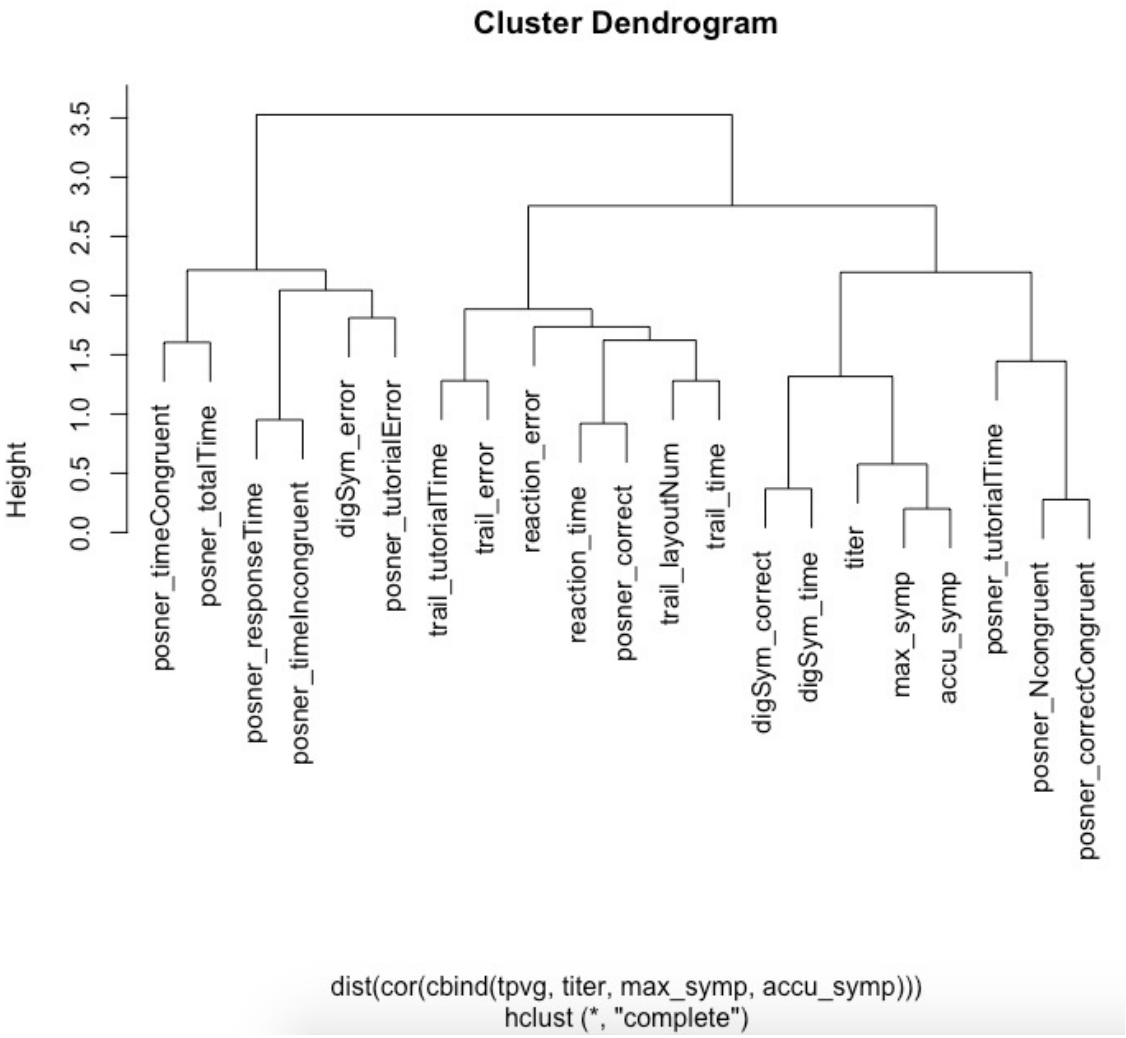


Figure 4.13: Correlation dendogram for viral titers, symptom and the NCPT covariates listed in Table 4.1a. The three NCPT variables included in the regression model are all on the right branch of depth two.

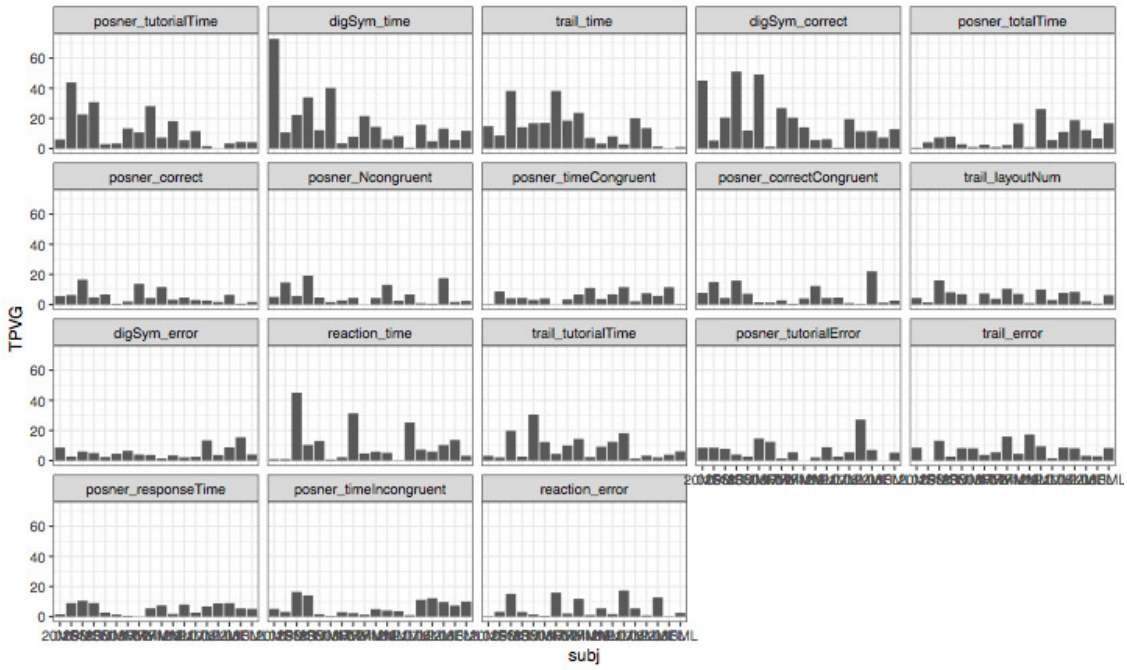


Figure 4.14: Panels indicate the distribution over subjects in the study of the univariate CPV of each of the NCPT variables in Table 4.1a ordered in decreasing correlation to CPV (maximum of the univariate CPVs).

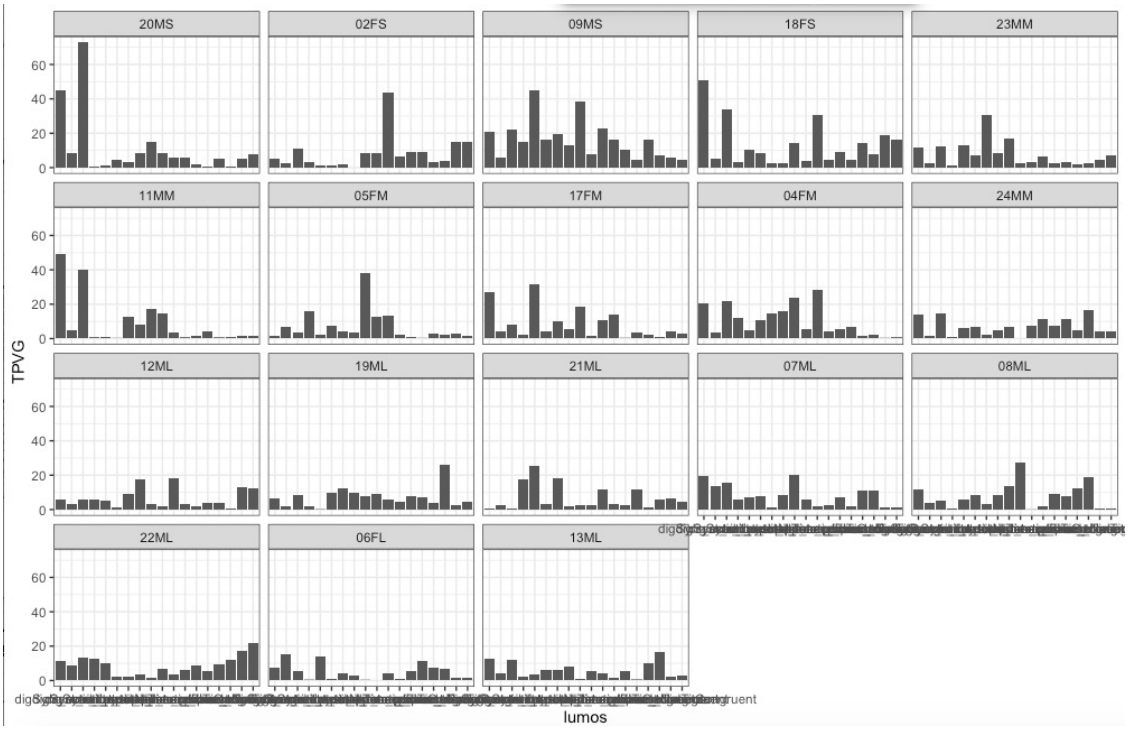


Figure 4.15: Panels indicate the distribution over NCPT variables in Table 4.1a in the study of the univariate CPV of each of the subjects, ordered from highest to lowest shedders.

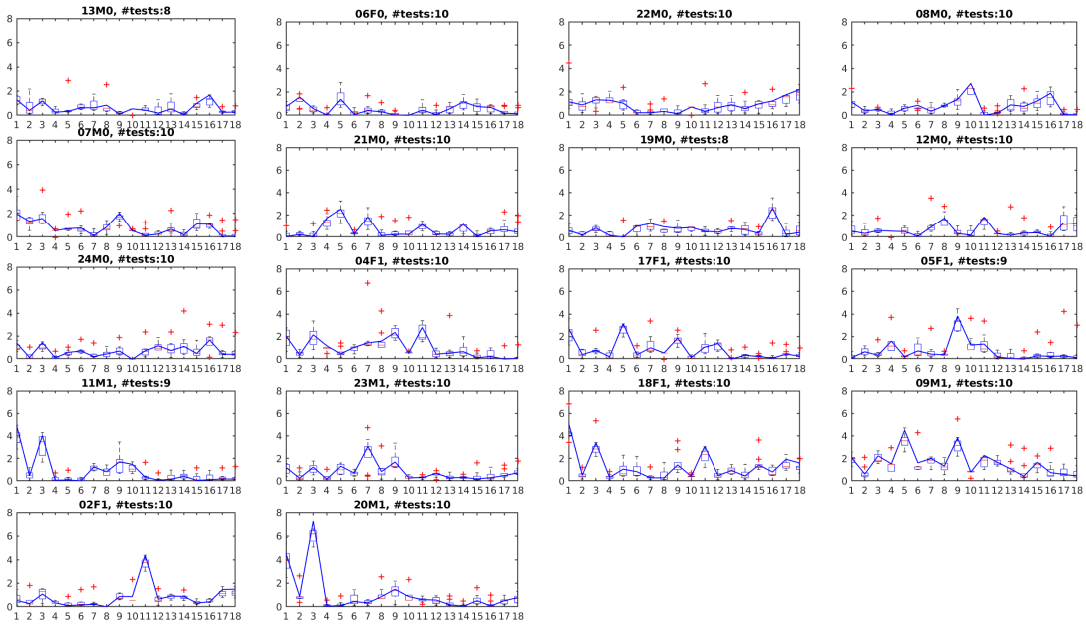


Figure 4.16: Boxplots of the univariate Cognitive Performance Variability scores over the 18 NCPT variables in Table 4.1a for all participants. Plots are arranged from lowest shedding (top left) to highest shedding (bottom right). challenge study participants. The Solid bluecurve denotes the CPV scores computed for each NCPT variable using the full session sequence, but omitting the initial screening session. Red boxplots show the distribution of the leave-one-out (loo) scores computed by successively dropping a single session from the sequence of post-screening sessions.

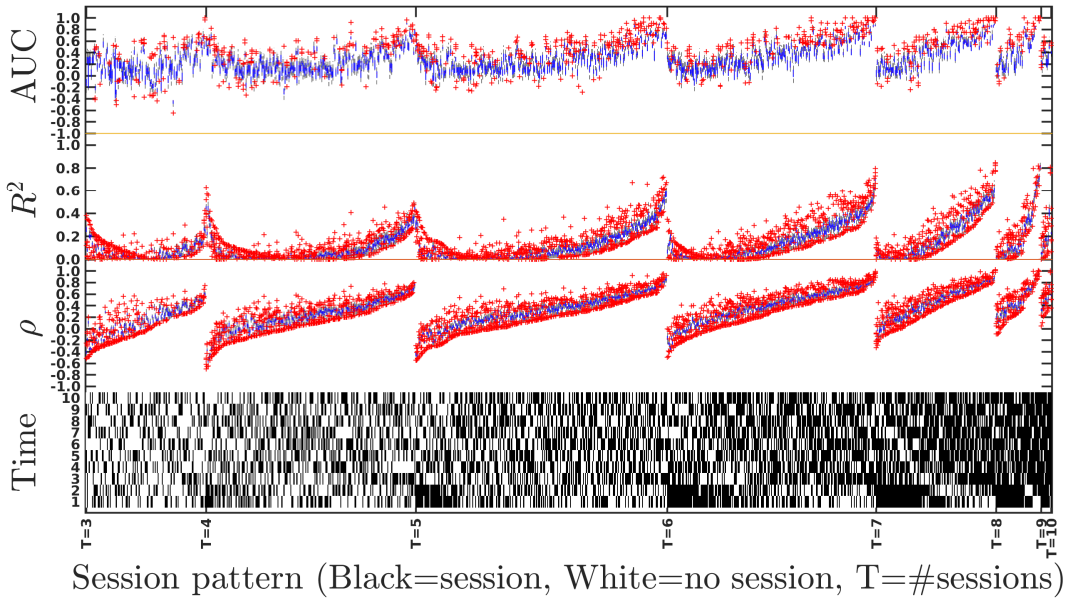


Figure 4.17: Boxplots of all $\sum_{k=3}^{10} \binom{10}{k} = 968$ cross-validated CPV Pearson correlation coefficients ρ, along with corresponding R squared and AUC for linear prediction of the amount of shedding and classification using logistic regression onto Low vs High shedding (0,1) labels, respectively, using the proposed maximum CPV score. The heatmap at bottom indicates the corresponding testing time patterns, out of the 968 possible patterns of all combinations of T=3,4,...,10 NCPT sessions administered during the baseline period. For each T=3,4,...,9, the top 10 patterns are rank ordered in decreasing order of the lowest cross-validated (leave-one-out) estimated value of ρ, excepting T=10 for which there is only one possible session pattern. The correlation coefficients, R^2 and AUC all degrade considerably when the reference session (1) is included in the calculation of the CPV.

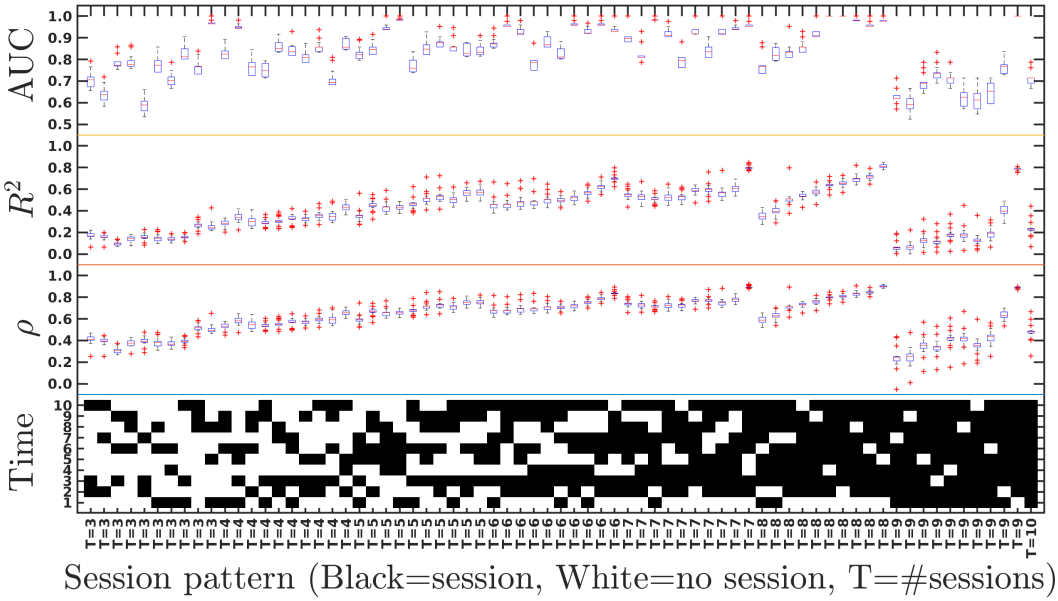


Figure 4.18: Zoomed version of previous figure showing the top 10 cross-validated CPV Pearson correlation coefficients ρ for each category of number of sessions $T = \{3, \ldots, 9\}$ and the single cross-validated CPV for $T = 10$ sessions.

NCPT variable	No. of correlated gene	Gene name	ρ	P value	FDR
digSym_correct	33	LTF	0.515	5.93E-12	1.08E-07
digSym_error	3726	NRG1	0.621	5.12E-18	9.34E-14
digSym_time	23	ADGRG7	0.527	1.57E-12	2.86E-08
reaction_error	0	CHAC1	0.333	2.16E-05	3.94E-01
reaction_time	155	ADGRG7	0.527	1.58E-12	2.88E-08
trail_layoutNum	0	CKAP2L	-0.333	2.20E-05	4.02E-01
trail_tutorialTime	1	KLRF2	0.409	1.19E-07	2.16E-03
trail_error	1	LHB	0.379	1.05E-06	1.91E-02
trail_time	0	MSRA	-0.331	2.48E-05	2.53E-01
posner_tutorialError	0	DEPDC4	0.316	5.92E-05	6.70E-01
posner_tutorialTime	9	MIR4760	0.654	2.14E-20	1.95E-16
posner_correct	14	MIR550A2	-0.369	2.13E-06	3.61E-02
posner_responseTime	83	ADGRG7	0.541	2.98E-13	5.43E-09
posner_timeCongruent	87	ADGRG7	0.494	5.37E-11	9.80E-07
posner_timeIncongruent	5	ADGRG7	0.465	9.23E-10	1.68E-05
posner_totalTime	19	ADGRG7	0.504	1.91E-11	3.48E-07
posner_Ncongruent	0	GNAL	0.270	6.51E-04	1.00
posner_correctCongruent	0	GNAL	0.270	6.50E-04	1.00

Figure 4.19: Pearson correlation between NCPT variables and genes.

Pathway Type	Name	FDR	Odds Ration	Pathway size	Shared size
Transport and catabolism	Phagosome	3.99E-03	2.269	132	48
Digestive system	Protein digestion and absorption	4.90E-03	2.671	77	31
Digestive system	Cholesterol metabolism	4.90E-03	3.770	41	20
Transport and catabolism	Lysosome	6.46E-03	2.159	122	43
Infectious diseases: Viral	Influenza A	7.59E-03	1.965	154	51
Infectious diseases: Parasitic	Leishmaniasis	1.69E-02	2.605	63	25
Cancers: Overview	Proteoglycans in cancer	2.18E-02	1.810	166	52
Immune system	Chemokine signaling pathway	2.23E-02	1.794	167	52
Infectious diseases: Bacterial	Salmonella infection	4.78E-02	2.189	73	26
Infectious diseases: Viral	Kaposi sarcoma-associated herpesvirus infection	4.78E-02	1.695	177	53
Cancers: Overview	Pathways in cancer	4.78E-02	1.387	464	120
Development	Osteoclast differentiation	4.78E-02	1.839	123	39

Figure 4.20: Enriched pathways of genes significantly correlated with NCPT variables.

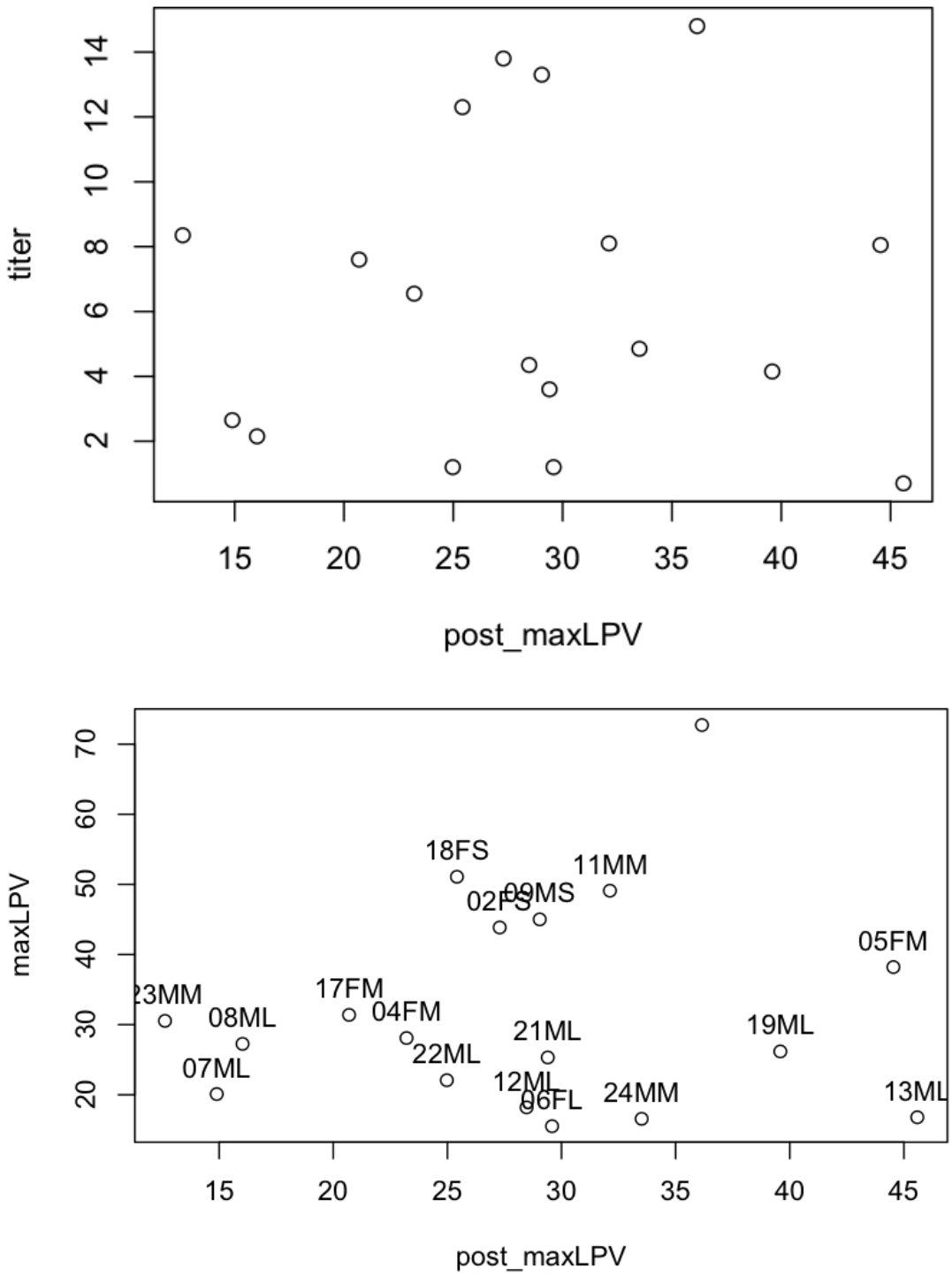


Figure 4.21: Top: Scatter of viral titers vs post-exposure CPV score. There is no no significant correlation. Bottom: Scatter of post-exposure CPV vs pre-exposure CPV.

CHAPTER 5

Conclusion and Future Directions

The significance of the contributions of this thesis might best be appreciated in the context of the widely acknowledged fact that some of the highest ARVI risk factors are poor sleep, stress, and fatigue. These risks reside both in the infected individual, i.e., personal morbidity associated with severe infection, and in the infectious individual, i.e., super-spreaders infecting the rest of the population, who may not be the same individuals. This thesis has exhibited new markers for susceptibility to ARVI infection and infectiousness, measured according to viral shedding and self-reported symptom levels, are related to stability of the cognitive, physiological and molecular states of the individual. For the physiological and cognitive states we have shown that these markers are associated with the temporal variations of the associated biomarkers over baseline, (sleep heart rate, sleep duration) for the physiological state and cognitive performance on games for the cognitive state. For the molecular states we have shown that it is the weakness and instability of the circadian cycle over baseline that differentiates the severely infected from the others. Markers and risk factors are not identical, however, so more work needs to be done to establish if these associations reflect causal relationships or not.

The contributions of this thesis thus fit into the landscape inhabited by future technologies for health maintenance, disease prevention, clinical medicine, and epidemiology. This would require much larger scale controlled studies that go beyond the small observational studies producing the data for this thesis. Such studies are strongly motivated by our results. This would permit validation of our results on larger cohorts, possible teasing out causal associations, and expanding the domain of applicability to demographics not well represented in the limited populations of the two available challenge studies, which were primarily college age healthy individuals.

A potential future vision might be constructed from the results of Chapters 3 and 4, which show that even a small duration segment of healthy baseline data suffices to learn digital biomarkers that index susceptibility markers and risk factors. This raises the intriguing question of whether variables can be found that, complementary to heart rate variation during sleep and cognitive performance variability, index a wide array of markers or behavioral risk factors for other diseases, e.g., Alzheimers, ALS, Parkinsons, addiction, etc. Could conveniently collectible, continuously

collected digital biomarkers derived from wearables and smart phones/tablet platforms play a significant role? With a much longer baseline it is conceivable that similar platforms might be used to infer more specific information on the healthy baseline, permitting better discrimination of risky deviations from baseline. With a sufficiently large corpus of labeled training data, perhaps collected continuously during a lifetime, algorithms similar to ours' could learn deviations from the baseline that are benign vs those that are potential risk factors. This algorithm could raise an alarm or silently cue complementary algorithms or measurements that might raise confidence on the benign nature of the deviation. With sufficient experimental resources and time, such an algorithm could in principle be derived and would truly be personalized, learning the range of healthy and non-healthy behaviors of the individual.

BIBLIOGRAPHY

[Ada98] Sudeshna Adak. Time-dependent spectral analysis of nonstationary time series. *Journal of the American Statistical Association*, 93(444):1488–1501, 1998.

[Agr10] Alan Agresti. *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons, 2010.

[ALML+14] Simon N Archer, Emma E Laing, Carla S Möller-Levet, Daan R van der Veen, Giselda Bucca, Alpar S Lazar, Nayantara Santhi, Ana Slak, Renata Kabiljo, Malcolm von Schantz, et al. Mistimed sleep disrupts circadian regulation of the human transcriptome. *Proceedings of the National Academy of Sciences*, 111(6):E682–E691, 2014.

[Ani08] Vladimir V Anisimov. *Switching processes in queueing models*. John Wiley & Sons, 2008.

[AOCC15] Sean T Anderson, Emma K O'Callaghan, Sean Commins, and Andrew N Coogan. Does prior sepsis alter subsequent circadian and sickness behaviour response to lipopolysaccharide treatment in mice? *Journal of Neural Transmission*, 122(1):63–73, 2015.

[AR07] Gidudu Anthony and Heinz Ruther. Comparison of feature selection techniques for svm classification. In *10th International Symposium on Physical Measurements and Signatures in Remote Sensing*, pages 1–6, 2007.

[AS05] Michael C Antle and Rae Silver. Orchestrating time: arrangements of the brain circadian clock. *Trends in neurosciences*, 28(3):145–151, 2005.

[BBFC08] Jimena Berni, Esteban J Beckwith, María Paz Fernández, and María Fernanda Ceriani. The axon-guidance roundabout gene alters the pace of the drosophila circadian clock. *European Journal of Neuroscience*, 27(2):396–407, 2008.

[BBK13] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials*, 16(1):303–336, 2013.

[BDTH12] Cécile Bazot, Nicolas Dobigeon, Jean-Yves Tourneret, and Alfred O Hero. Bayesian linear unmixing of time-evolving gene expression data using a hidden markov model. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 944–948. IEEE, 2012.

[BFS+09] Devin L Brown, Diane Feskanich, Brisa N Sánchez, Kathryn M Rexrode, Eva S Schernhammer, and Lynda D Lisabeth. Rotating night shift work and the risk of ischemic stroke. *American journal of epidemiology*, 169(11):1370–1377, 2009.

[BHPB17] Amber J Barton, Jennifer Hill, Andrew J Pollard, and Christoph J Blohmke. Transcriptomics in human challenge models. *Frontiers in immunology*, 8:1839, 2017.

[Bis06] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[BKN17] Reza Boostani, Foroozan Karimzadeh, and Mohammad Nami. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer methods and programs in biomedicine*, 140:77–91, 2017.

[BSB19] Sandra Bucci, Matthias Schwannauer, and Natalie Berry. The digital revolution and its impact on mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*, 92(2):277–297, 2019.

[BWV+01] Fiona C Baker, Jonathan I Waner, Elizabeth F Vieira, Sheila R Taylor, Helen S Driver, and Duncan Mitchell. Sleep and 24 hour body temperatures: a comparison in young men, naturally cycling women and women taking hormonal contraceptives. *The Journal of physiology*, 530(3):565–574, 2001.

[CAP+09] Laura Cesarini, Paolo Alfieri, Francesca Pantaleoni, Isabella Vasta, Marta Cerutti, Valentina Petrangeli, Paolo Mariotti, Chiara Leoni, Daniela Ricci, Stefano Vicari, et al. Cognitive profile of disorders associated with dysregulation of the ras/mapk signaling cascade. *American Journal of Medical Genetics Part A*, 149(2):140–146, 2009.

[CBHK02] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[CC08] Drew E Carlson and William C Chiu. The absence of circadian cues during recovery from sepsis modifies pituitary-adrenocortical function and impairs survival. *Shock*, 29(1):127–132, 2008.

[CD+05] Mary A Carskadon, William C Dement, et al. Normal human sleep: an overview. *Principles and practice of sleep medicine*, 4:13–23, 2005.

[CDA+09] Sheldon Cohen, William J Doyle, Cuneyt M Alper, Denise Janicki-Deverts, and Ronald B Turner. Sleep habits and susceptibility to the common cold. *Archives of internal medicine*, 169(1):62–67, 2009.

[CKM83] Sheldon Cohen, Tom Kamarck, and Robin Mermelstein. A global measure of perceived stress. *Journal of health and social behavior*, 24(4):385–396, 1983.

[CL14] André LV Coelho and Clodoaldo AM Lima. Assessing fractal dimension methods as feature extractors for emg signal classification. *Engineering Applications of Artificial Intelligence*, 36:81–98, 2014.

[Cle79] William S Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836, 1979.

[CMS12] J Orestes Cerdeira, M João Martins, and Pedro C Silva. A combinatorial approach to assess the separability of clusters. *Journal of classification*, 29(1):7–22, 2012.

[CWB11] Minmin Chen, Kilian Q Weinberger, and John Blitzer. Co-training for domain adaptation. In *Advances in neural information processing systems*, pages 2456–2464, 2011.

[CWH+05] James P Cleary, Dominic M Walsh, Jacki J Hofmeister, Ganesh M Shankar, Michael A Kuskowski, Dennis J Selkoe, and Karen H Ashe. Natural oligomers of the amyloid-β protein specifically disrupt cognitive function. *Nature neuroscience*, 8(1):79–84, 2005.

[DCM03] Murray Dworetzky, Sheldon Cohen, and David Mullin. Prometheus in gloucestershire: Edward jenner, 1749-1823. *Journal of allergy and clinical immunology*, 112(4):810–814, 2003.

[DRR+00] Christopher L Drake, Timothy A Roehrs, Heather Royer, Gale Koshorek, Ronald B Turner, and Thomas Roth. Effects of an experimentally induced rhinovirus cold on sleep, performance, and daytime alertness. *Physiology & behavior*, 71(1-2):75–81, 2000.

[DRS18] Jessilyn Dunn, Ryan Runge, and Michael Snyder. Wearables and the medical revolution. *Personalized medicine*, 15(5):429–448, 2018.

[EGZ+12] Rachel S Edgar, Edward W Green, Yuwei Zhao, Gerben Van Ooijen, Maria Olmedo, Ximing Qin, Yao Xu, Min Pan, Utham K Valekunja, Kevin A Feeney, et al. Peroxiredoxins are conserved markers of circadian rhythms. *Nature*, 485(7399):459–464, 2012.

[EMSC09] Kristin Eckel-Mahan and Paolo Sassone-Corsi. Metabolism control by the circadian clock and vice versa. *Nature structural & molecular biology*, 16(5):462, 2009.

[EMSC13] Kristin Eckel-Mahan and Paolo Sassone-Corsi. Metabolism and the circadian clock converge. *Physiological reviews*, 93(1):107–135, 2013.

[FDSR02] Arthur Flexerand, Georg Dorffner, Peter Sykacekand, and Iead Rezek. An automatic, continuous and probabilistic sleep stager based on a hidden markov model. *Applied Artificial Intelligence*, 16(3):199–207, 2002.

[FHT01] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Springer series in statistics New York, NY, 2001.

[Fis36] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[Gaz00] Michael S Gazzaniga. *The new cognitive neurosciences*. MIT press, 2000.

[GPPK+20] Krystal J Godri Pollitt, Jordan Peccia, Albert I Ko, Naftali Kaminski, Charles S Dela Cruz, Daniel W Nebert, Juergen KV Reichardt, David C Thompson, and Vasilis Vasiliou. Covid-19 vulnerability: the potential impact of genetic susceptibility and airborne transmission. *Human genomics*, 14:1–7, 2020.

[GR10] Diego A Golombek and Ruth E Rosenstein. Physiology of circadian entrainment. *Physiological reviews*, 90(3):1063–1102, 2010.

[Gre01] John Greene. Feature subset selection using Thornton's separability index and its applicability to a number of sparse proximity-based classifiers. In *Proceedings of Annual Symposium of the Pattern Recognition Association of South Africa*, 2001.

[GS11] Xin Gao and Peter X-K Song. Composite likelihood em algorithm with applications to multivariate hidden markov model. *Statistica Sinica*, pages 165–185, 2011.

[GV07] Monica Gallego and David M Virshup. Post-translational modifications regulate the ticking of the circadian clock. *Nature reviews Molecular cell biology*, 8(2):139–148, 2007.

[GŽB+14] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys*, 46(4):44, 2014.

[HAB+20] Jeffrey A Haspel, Ron Anafi, Marishka K Brown, Nicolas Cermakian, Christopher Depner, Paula Desplats, Andrew E Gelman, Monika Haack, Sanja Jelic, Brian S Kim, et al. Perfect timing: circadian rhythms, sleep, and immunity—an nih workshop summary. *JCI insight*, 5(1), 2020.

[HBB17] Becky I Haynes, Sarah Bauermeister, and David Bunce. A systematic review of longitudinal associations between reaction time intraindividual variability and age-related cognitive decline or impairment, dementia, and mortality. *Journal of the International Neuropsychological Society*, 23(5):431–445, 2017.

[HCL+18] Heonjong Han, Jae-Won Cho, Sangyoung Lee, Ayoung Yun, Hyojin Kim, Dasom Bae, Sunmo Yang, Chan Yeong Kim, Muyoung Lee, Eunbeen Kim, et al. Trrust v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research*, 46(D1):D380–D386, 2018.

[HCS+14] Jeffrey A Haspel, Sukrutha Chettimada, Rahamthulla S Shaik, Jen-Hwa Chu, Benjamin A Raby, Manuela Cernadas, Vincent Carey, G Matthew Hunninghake, Emeka Ifedigbo, James A Lederer, et al. Circadian rhythm reprogramming during lung inflammation. *Nature communications*, 5(1):1–15, 2014.

[HHK10] Michael E Hughes, John B Hogenesch, and Karl Kornacker. Jtk_cycle: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of biological rhythms*, 25(5):372–380, 2010.

[HHMH03] Robert B Henderson, Josie AR Hobbs, Meg Mathies, and Nancy Hogg. Rapid recruitment of inflammatory monocytes is independent of neutrophil migration. *Blood*, 102(1):328–335, 2003.

[HIT86] David C Hoaglin, Boris Iglewicz, and John W Tukey. Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396):991–999, 1986.

[HL94] David P Helmbold and Philip M Long. Tracking drifting concepts by minimizing disagreements. *Machine learning*, 14(1):27–45, 1994.

[HLW+20] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.

[Hru85] WJ Hrushesky. Circadian timing of cancer chemotherapy. *Science*, 228(4695):73–75, 1985.

[Hug17] Jacob J Hughey. Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. *Genome medicine*, 9(1):1–11, 2017.

[HvdAS01] Bas T Hemker, L Andries van der Ark, and Klaas Sijtsma. On measurement properties of continuation ratio models. *Psychometrika*, 66(4):487–506, 2001.

[HWK09] Dirk H Hellhammer, Stefan Wüst, and Brigitte M Kudielka. Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, 34(2):163–171, 2009.

[HZR+11] Yongsheng Huang, Aimee K Zaas, Arvind Rao, Nicolas Dobigeon, Peter J Woolf, Timothy Veldman, N Christine Øien, Micah T McClain, Jay B Varkey, Bradley Nicholson, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS genetics*, 7(8):e1002234, 2011.

[Irw15] Michael R Irwin. Why sleep is important for health: a psychoneuroimmunology perspective. *Annual review of psychology*, 66, 2015.

[JCB07] Francine O James, Nicolas Cermakian, and Diane B Boivin. Circadian rhythms of melatonin, cortisol, and clock gene expression during simulated night shift work. *Sleep*, 30(11):1427–1436, 2007.

[JHU] Johns hopkins university covid-19 dashboard. `https://coronavirus.jhu.edu/map.html`. Accessed: 2021-06-30.

[JWL+21] Huajie Jin, Haiyin Wang, Xiao Li, Weiwei Zheng, Shanke Ye, Sheng Zhang, Jiahui Zhou, and Mark Pennington. Economic burden of covid-19, china, january–march, 2020: a cost-of-illness study. *Bulletin of the World Health Organization*, 99(2):112, 2021.

[KDM+18] Dae Y Kang, Pamela N DeYoung, Atul Malhotra, Robert L Owens, and Todd P Coleman. A state space and density estimation framework for sleep staging in obstructive sleep apnea. *IEEE Transactions on Biomedical Engineering*, 65(6):1201–1212, 2018.

[KF11] Walter Karlen and Dario Floreano. Adaptive sleep–wake discrimination for wearable devices. *IEEE Transactions on Biomedical Engineering*, 58(4):920–926, 2011.

[KGD12] Jonathan Kipnis, Sachin Gadani, and Noël C Derecki. Pro-cognitive properties of t cells. *Nature Reviews Immunology*, 12(9):663–669, 2012.

[KGK+04] Minoru Kanehisa, Susumu Goto, Shuichi Kawashima, Yasushi Okuno, and Masahiro Hattori. The kegg resource for deciphering the genome. *Nucleic acids research*, 32(suppl_1):D277–D280, 2004.

[KHH+17] Katherine A Kaplan, Jason Hirshman, Beatriz Hernandez, Marcia L Stefanick, Andrew R Hoffman, Susan Redline, Sonia Ancoli-Israel, Katie Stone, Leah Friedman, Jamie M Zeitzer, et al. When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography. *Biological psychology*, 123:37–46, 2017.

[KJ00] Ralf Klinkenberg and Thorsten Joachims. Detecting concept drift with support vector machines. In *Proceedings of the 17th International Conference on Machine Learning*, pages 487–494, 2000.

[KKB+18] Jungyoon Kim, Boncho Ku, Jang-Han Bae, Gyu-Cheol Han, and Jaeuk U Kim. Contrast in the circadian behaviors of an electrodermal activity and bioimpedance spectroscopy. *Chronobiology international*, 35(10):1413–1422, 2018.

[KLMNS89] Lauren B Krupp, Nicholas G LaRocca, Joanne Muir-Nash, and Alfred D Steinberg. The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of neurology*, 46(10):1121–1123, 1989.

[KMB+10] Juha M Kortelainen, Martin O Mendez, Anna Maria Bianchi, Matteo Matteucci, and Sergio Cerutti. Sleep staging based on signals acquired through bed sensor. *IEEE Transactions on Information Technology in Biomedicine*, 14(3):776–785, 2010.

[KMF08] Walter Karlen, Claudio Mattiussi, and Dario Floreano. Improving actigraph sleep/wake classification with cardio-respiratory signals. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 5262–5265. IEEE, 2008.

[KML20] Ananya Kumar, Tengyu Ma, and Percy Liang. Understanding self-training for gradual domain adaptation. *arXiv preprint arXiv:2002.11361*, 2020.

[KŽ09] Ludmila I Kuncheva and Indrė Žliobaitė. On the window size for classification in changing environments. *Intelligent Data Analysis*, 13(6):861–872, 2009.

[LCJ+21] Ming Li, Zhencong Chen, Tian Jiang, Xiaodong Yang, Yajing Du, Jiaqi Liang, Lin Wang, Junjie Xi, Miao Lin, and Mingxiang Feng. Circadian rhythm-associated clinical relevance and tumor microenvironment of non-small cell lung cancer. *Journal of Cancer*, 12(9):2582, 2021.

[LH02] Theodore C Lystig and James P Hughes. Exact computation of the observed information matrix for hidden markov models. *Journal of Computational and Graphical Statistics*, 11(3):678–689, 2002.

[LHK+16] Ji-Won Lee, Bill X Huang, HeungSun Kwon, Md Abdur Rashid, Giorgi Kharebava, Abhishek Desai, Samarjit Patnaik, Juan Marugan, and Hee-Yong Kim. Orphan gpr110 (adgrf1) targeted by n-docosahexaenoylethanolamine in development of neurons and cognitive function. *Nature communications*, 7(1):1–16, 2016.

[LMLP+17] Emma E Laing, Carla S Möller-Levet, Norman Poh, Nayantara Santhi, Simon N Archer, and Derk-Jan Dijk. Blood transcriptome based biomarkers for human circadian phase. *Elife*, 6:e20214, 2017.

[Lup13] Deborah Lupton. The digitally engaged patient: Self-monitoring and self-care in the digital health era. *Social Theory & Health*, 11(3):256–270, 2013.

[LWNM+18] Rob Lambkin-Williams, Nicolas Noulin, Alex Mann, Andrew Catchpole, and Anthony S Gilbert. The human viral challenge model: accelerating the evaluation of respiratory antivirals, vaccines and novel diagnostics. *Respiratory research*, 19(1):1–15, 2018.

[MBCG05] Luciano Marpegán, Tristán A Bekinschtein, Monica A Costas, and Diego A Golombek. Circadian responses to endotoxin treatment in mice. *Journal of neuroimmunology*, 160(1-2):102–109, 2005.

[MDB+17] Bertalan Meskó, Zsófia Drobni, Éva Bényei, Bence Gergely, and Zsuzsanna Győrffy. Digital health is a cultural transformation of traditional healthcare. *Mhealth*, 3, 2017.

[MDKK+02] Gerald Mundigler, Georg Delle-Karth, Maria Koreny, Manfred Zehetgruber, Petra Steindl-Munda, Wolfgang Marktl, Lisa Ferti, and Peter Siostrzonek. Impaired circadian rhythm of melatonin secretion in sedated critically ill patients with severe sepsis. *Critical care medicine*, 30(3):536–540, 2002.

[Mil15] Michelle A Miller. The role of sleep and sleep disorders in the development, diagnosis, and management of neurocognitive disorders. *Frontiers in neurology*, 6:224, 2015.

[ML97] Christopher JL Murray and Alan D Lopez. Global mortality, disability, and the contribution of risk factors: Global burden of disease study. *The lancet*, 349(9063):1436–1442, 1997.

[MMW+00] Martial M Massin, Krystel Maeyns, Nadia Withofs, Françoise Ravet, and Paul Gérard. Circadian rhythm of heart rate and heart rate variability. *Archives of disease in childhood*, 83(2):179–182, 2000.

[MPR+98] Mika J Makela, Tuomo Puhakka, Olli Ruuskanen, Maija Leinonen, Pekka Saikku, Marko Kimpimäki, Soile Blomqvist, Timo Hyypiä, and Pertti Arstila. Viruses and bacteria in the etiology of the common cold. *Journal of clinical microbiology*, 36(2):539–542, 1998.

[MR10] Amit K Mishra and Shantanu Raghav. Local fractal dimension based ecg arrhythmia classification. *Biomedical Signal Processing and Control*, 5(2):114–123, 2010.

[MSNH15] Glenn E Morrison, Christa M Simone, Nicole F Ng, and Joseph L Hardy. Reliability and validity of the neurocognitive performance test, a web-based neuropsychological assessment. *Frontiers in psychology*, 6:1652, 2015.

[MTRAR+12] Jose G Moreno-Torres, Troy Raeder, RocíO Alaiz-RodríGuez, Nitesh V Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern Recognition*, 45(1):521–530, 2012.

[MVDC+10] Gianluigi Mazzoccoli, Gianluigi Vendemiale, Angelo De Cata, Stefano Carughi, and Roberto Tarquini. Altered time structure of neuro-endocrine-immune system function in lung cancer patients. *BMC cancer*, 10(1):1–10, 2010.

[MYK+13] Atul Malhotra, Magdy Younes, Samuel T Kuna, Ruth Benca, Clete A Kushida, James Walsh, Alexandra Hanlon, Bethany Staley, Allan I Pack, and Grace W Pien. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep*, 36(4):573–582, 2013.

[NCB16] Amy H Newton, Amber Cardani, and Thomas J Braciale. The host immune response in respiratory virus infection: balancing virus clearance and immunopathology. In *Seminars in immunopathology*, volume 38, pages 471–482. Springer, 2016.

[OCH13] Koichi Ogami, Rihe Cho, and Shin-ichi Hoshino. Molecular cloning and characterization of a novel isoform of the non-canonical poly (a) polymerase papd7. *Biochemical and biophysical research communications*, 432(1):135–140, 2013.

[OYD+08] Kazuyuki Okada, Masahiko Yano, Yuichiro Doki, Takashi Azama, Hiroshi Iwanaga, Hirofumi Miki, Mitsuo Nakayama, Hiroshi Miyata, Shuji Takiguchi, Yoshiyuki Fujiwara, et al. Injection of lps causes transient suppression of biological clock genes in rats. *Journal of Surgical Research*, 145(1):5–12, 2008.

[PB08] Maria Palomba and Marina Bentivoglio. Chronic inflammation affects the photic response of the suprachiasmatic nucleus. *Journal of neuroimmunology*, 193(1-2):24–27, 2008.

[PCC+03] Joseph Sriyal Malik Peiris, Chung-Ming Chu, Vincent Chi-Chung Cheng, KS Chan, IFN Hung, Leo LM Poon, Kin-Ip Law, BSF Tang, TYW Hon, CS Chan, et al. Clinical progression and viral load in a community outbreak of coronavirus-associated sars pneumonia: a prospective study. *The Lancet*, 361(9371):1767–1772, 2003.

[PJDHC15] Aric A Prather, Denise Janicki-Deverts, Martica H Hall, and Sheldon Cohen. Behaviorally assessed sleep and susceptibility to the common cold. *Sleep*, 38(9):1353–1359, 2015.

[PL16] Aric A Prather and Cindy W Leung. Association of insufficient sleep with respiratory infection among adults in the united states. *JAMA internal medicine*, 176(6):850–852, 2016.

[PSSH11] An Pan, Eva S Schernhammer, Qi Sun, and Frank B Hu. Rotating night shift work and risk of type 2 diabetes: two prospective cohort studies in women. *PLoS Med*, 8(12):e1001141, 2011.

[PY+10] Sinno Jialin Pan, Qiang Yang, et al. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[Ran09] Christoph Randler. Validation of the full and reduced composite scale of morningness. *Biological Rhythm Research*, 40(5):413–423, 2009.

[RCW+20] Li Ran, Xuyu Chen, Ying Wang, Wenwen Wu, Ling Zhang, and Xiaodong Tan. Risk factors of healthcare workers with coronavirus disease 2019: a retrospective cohort study in a designated hospital of wuhan in china. *Clinical Infectious Diseases*, 71(16):2218–2221, 2020.

[RK] Allan Rechtschaffen and Anthony Kales. A manual of standardized terminology, techniques and scoring system of sleep stages in human subjects. Brain Information Service, University of California, Los Angeles.

[RWM15] Luca Rossi, James Walker, and Mirco Musolesi. Spatio-temporal techniques for user identification by means of gps mobility data. *EPJ Data Science*, 4(1):1–16, 2015.

[Sal07] Timothy A Salthouse. Implications of within-person variability in cognitive and neuropsychological functioning for the interpretation of change. *Neurpsychology*, 21(4):401–411, 2007.

[SAS+09] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2009.

[SBG+07] Kurt Straif, Robert Baan, Yann Grosse, Béatrice Secretan, Fatiha El Ghissassi, Véronique Bouvard, Andrea Altieri, Lamia Benbrahim-Tallaa, Vincent Cogliano, WHO International Agency for Research on Cancer Monograph Working Group, et al. Carcinogenicity of shift-work, painting, and fire-fighting, 2007.

[Sch99] Henry Scheffe. *The analysis of variance*, volume 72. John Wiley & Sons, 1999.

[SCS+17] Sohrab Saeb, Thaddeus R Cybulski, Stephen M Schueller, Konrad P Kording, and David C Mohr. Scalable passive sleep monitoring using mobile phones: opportunities and obstacles. *Journal of medical Internet research*, 19(4), 2017.

[SCVO13] Diane Slats, Jurgen AHR Claassen, Marcel M Verbeek, and Sebastiaan Overeem. Reciprocal interactions between sleep, circadian rhythms and alzheimer's disease: focus on the role of hypocretin and melatonin. *Ageing research reviews*, 12(1):188–200, 2013.

[SDTD06] Beverly A Shipley, Geoff Der, Michelle D Taylor, and Ian J Deary. Cognition and all-cause mortality across the entire adult age range: health and lifestyle survey. *Psychosomatic medicine*, 68(1):17–24, 2006.

[SGS+15] Ueli Schibler, Ivana Gotic, Camille Saini, Pascal Gos, Thomas Curie, Yann Emmenegger, Flore Sinturel, Pauline Gosselin, Alan Gerber, Fabienne Fleury-Olela, et al. Clock-talk: interactions between central and peripheral circadian oscillators in mammals. In *Cold Spring Harbor symposia on quantitative biology*, volume 80, pages 223–232. Cold Spring Harbor Laboratory Press, 2015.

[SHMS09] Frank AJL Scheer, Michael F Hilton, Christos S Mantzoros, and Steven A Shea. Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proceedings of the National Academy of Sciences*, 106(11):4453–4458, 2009.

[SJ08] Areejit Samal and Sanjay Jain. The regulatory network of e. coli metabolism as a boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC systems biology*, 2(1):1–18, 2008.

[SL03] Ke Shuai and Bin Liu. Regulation of jak–stat signalling in the immune system. *Nature Reviews Immunology*, 3(11):900–911, 2003.

[SLBG+15] Aziz Sancar, Laura A Lindsey-Boltz, Shobhan Gaddameedhi, Christopher P Selby, Rui Ye, Yi-Ying Chiou, Michael G Kemp, Jinchuan Hu, Jin Hyup Lee, and Nuri Ozturk. Circadian clock, cancer, and chemotherapy. *Biochemistry*, 54(2):110–123, 2015.

[SLL+02] Kai-Florian Storch, Ovidiu Lipan, Igor Leykin, N Viswanathan, Fred C Davis, Wing H Wong, and Charles J Weitz. Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417(6884):78–83, 2002.

[SLS+01] Eva S Schernhammer, Francine Laden, Frank E Speizer, Walter C Willett, David J Hunter, Ichiro Kawachi, and Graham A Colditz. Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. *Journal of the national cancer institute*, 93(20):1563–1568, 2001.

[Smi12] Andrew P Smith. Effects of the common cold on mood, psychomotor performance, the encoding of new information, speed of working memory and semantic processing. *Brain, behavior, and immunity*, 26(7):1072–1076, 2012.

[SMRS+20] Cláudia Serre-Miranda, Susana Roque, Nadine Correia Santos, Patrício Costa, Nuno Sousa, Joana Almeida Palha, and Margarida Correia-Neves. Cognition is associated with peripheral immune molecules in healthy older adults: a cross-sectional study. *Frontiers in immunology*, 11:2045, 2020.

[SOG+10] Dee Unglaub Silverthorn, William C Ober, Claire W Garrison, Andrew C Silverthorn, and Bruce R Johnson. *Human physiology: an integrated approach*. Pearson/Benjamin Cummings San Francisco, 2010.

[SR14] Frederik C Schadd and Nico Roos. A feature selection approach for anchor evaluation in ontology mapping. In *International Conference on Knowledge Engineering and the Semantic Web*, pages 160–174. Springer, 2014.

[SRM89] Carlla S Smith, Christopher Reilly, and Karen Midkiff. Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *Journal of Applied psychology*, 74(5):728, 1989.

[SS+12] Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

[SWE+20] Yang Shen, Wei Wang, Mehari Endale, Lauren J Francey, Rachel L Harold, David W Hammers, Zhiguang Huo, Carrie L Partch, John B Hogenesch, Zhao-Hui Wu, et al. Nf-κb modifies the mammalian circadian clock through interaction with the core clock protein bmal1. *bioRxiv*, 2020.

[SZH+20] Xichen She, Yaya Zhai, Ricardo Henao, Christopher Woods, Christopher Chiu, Geoffrey S. Ginsburg, Peter X. K. Song, and Alfred O. Hero. Adaptive multichannel event segmentation and feature extraction for monitoring health outcomes. *IEEE Transactions on Biomedical Engineering*, ?(?):1–1, 2020.

[THKM08] Joseph S Takahashi, Hee-Kyung Hong, Caroline H Ko, and Erin L McDearmon. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nature reviews genetics*, 9(10):764–775, 2008.

[Tho98] Chris Thornton. Separability is a learner's best friend. In *4th Neural Computation and Psychology Workshop, London, 9–11 April 1997*, pages 40–46. Springer, 1998.

[TMC+13] Dorra Trabelsi, Samer Mohammed, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. An unsupervised approach for automatic activity recognition based on hidden markov model regression. *IEEE Transactions on automation science and engineering*, 10(3):829–835, 2013.

[TMN10] Elio Tuci, Gianluca Massera, and Stefano Nolfi. Active categorical perception of object shapes in a simulated anthropomorphic robotic arm. *IEEE transactions on evolutionary computation*, 14(6):885–899, 2010.

[Tuk77] John W Tukey. *Exploratory data analysis*. Addison-Wesley, Reading, MA, 1977.

[TYH+01] Satomi Takahashi, Shin-ichi Yokota, Reiko Hara, Tomoko Kobayashi, Masashi Akiyama, Takahiro Moriya, and Shigenobu Shibata. Physical and inflammatory stressors elevate circadian clock gene mper1 mrna levels in the paraventricular nucleus of the mouse. *Endocrinology*, 142(11):4910–4917, 2001.

[VKO17] Alejandro V Villarino, Yuka Kanno, and John J O'Shea. Mechanisms and consequences of jak–stat signaling in the immune system. *Nature immunology*, 18(4):374–384, 2017.

[VLM+11] Saul A Villeda, Jian Luo, Kira I Mosher, Bende Zou, Markus Britschgi, Gregor Bieri, Trisha M Stan, Nina Fainberg, Zhaoqing Ding, Alexander Eggel, et al. The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature*, 477(7362):90–94, 2011.

[VS+10] Ingmar Visser, Maarten Speekenbrink, et al. depmixs4: an r package for hidden markov models. *Journal of Statistical Software*, 36(7):1–21, 2010.

[WK96] Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. *Machine learning*, 23(1):69–101, 1996.

[WKL12] Günter P Wagner, Koryu Kin, and Vincent J Lynch. Measurement of mrna abundance using rna-seq data: Rpkm measure is inconsistent among samples. *Theory in biosciences*, 131(4):281–285, 2012.

[XOF+16] Cheng Xu, Hiroki Ochi, Toru Fukuda, Shingo Sato, Satoko Sunamura, Takeshi Takarada, Eiichi Hinoi, Atsushi Okawa, and Shu Takeda. Circadian clock regulates bone resorption in mice. *Journal of Bone and Mineral Research*, 31(7):1344–1355, 2016.

[Yu12] Jianbo Yu. Health condition monitoring of machines based on hidden markov model and contribution analysis. *IEEE Transactions on Instrumentation and Measurement*, 61(8):2200–2211, 2012.

[YYL+04] Seung-Hee Yoo, Shin Yamazaki, Phillip L Lowrey, Kazuhiro Shimomura, Caroline H Ko, Ethan D Buhr, Sandra M Siepka, Hee-Kyung Hong, Won Jun Oh, Ook Joon Yoo, et al. Period2:: Luciferase real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proceedings of the National Academy of Sciences*, 101(15):5339–5346, 2004.

[ZCV+09a] Aimee K. Zaas, Minhua Chen, Jay Varkey, Timothy Veldman, Alfred O. Hero, Joseph Lucas, Yongsheng Huang, Ronald Turner, Anthony Gilbert, Robert Lambkin-Williams, N. Christine Øien, Bradly Nicholson, Stephen Kingsmore, Lawrence Carin, Christopher W. Woods, and Geoffrey S. Ginsburg. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host and Microbe*, 6(3):207–217, 2009.

[ZCV+09b] Aimee K Zaas, Minhua Chen, Jay Varkey, Timothy Veldman, Alfred O Hero III, Joseph Lucas, Yongsheng Huang, Ronald Turner, Anthony Gilbert, Robert Lambkin-Williams, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell host & microbe*, 6(3):207–217, 2009.

[ZFS91] Theodore P Zahn, Christopher D Frith, and Stuart R Steinhauer. Autonomic functioning in schizophrenia: Electrodermal activity, heart rate, pupillography. 1991.

[ZHW+19] Zhenguang Zhang, Louise Hunter, Gang Wu, Robert Maidstone, Yasutaka Mizoro, Ryan Vonslow, Mark Fife, Thomas Hopwood, Nicola Begley, Ben Saer, et al. Genome-wide effect of pulmonary airway epithelial cell–specific bmal1 deletion. *The FASEB Journal*, 33(5):6226–6238, 2019.

[ZLB+14] Ray Zhang, Nicholas F Lahens, Heather I Ballance, Michael E Hughes, and John B Hogenesch. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proceedings of the National Academy of Sciences*, 111(45):16219–16224, 2014.

[ZML17] Walter Zucchini, Iain L MacDonald, and Roland Langrock. *Hidden Markov models for time series: an introduction using R*. CRC press, 2017.

[ZYL+19] Xue Zhou, Youjun Yu, Anais Lopez, Josephine Felber, Helene Gylling, Gitte Friis, Soren Ottosen, and Henrik Mueller. Liver-targeted inhibition of papd5 and papd7 leads to sustainable hbsag reduction in the aav-hbv mouse model. In *HEPATOLOGY*, volume 70, pages 439A–440A. WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2019.